



ulm university universität
uulm

Stochastik I

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

ULM
2012

Vorwort

Dieses Skript entstand aus dem Zyklus der Vorlesungen über Statistik, die ich in den Jahren 2005–2012 an der Universität Ulm gehalten habe. Dabei handelt es sich um die erste Einführung in die Statistik, die durch die aufbauende Vorlesung Stochastik III ergänzt wird.

Dieses Skript gibt eine Übersicht über die typischen Fragestellungen und Methoden der mathematischen Statistik. Es stellt einen Versuch dar, einen Mittelweg zwischen praktisch orientierten (aber mathematisch oft sehr dürftigen) Statistik-Monographien einerseits und trockenen Büchern über die mathematische Statistik andererseits einzuschlagen. Ob es mir gelungen ist, soll der Leser beurteilen.

Ich möchte gerne meinen Kollegen aus dem Institut für Stochastik, Herrn Prof. Volker Schmidt und Herrn Dipl.-Math. Malte Spiess, für ihre Unterstützung und anregenden Diskussionen während der Entstehung des Skriptes danken. Herr Tobias Brosch hat eine hervorragende Arbeit beim Tippen des Skriptes und bei der Erstellung zahlreicher Abbildungen, die den Text begleiten, geleistet. Dafür gilt ihm mein herzlicher Dank.

Ulm, den 19.04.2012

Evgeny Spodarev

Inhaltsverzeichnis

1	Einführung	1
1.1	Typische Fragestellungen, Aufgaben und Ziele der Statistik	1
1.2	Statistische Merkmale und ihre Typen	4
1.3	Statistische Daten und Stichproben	5
1.4	Stichprobenfunktionen	6
2	Beschreibende Statistik	7
2.1	Verteilungen und ihre Darstellungen	7
2.1.1	Häufigkeiten und Diagramme	7
2.1.2	Empirische Verteilungsfunktion	10
2.2	Beschreibung von Verteilungen	12
2.2.1	Lagemaße	12
2.2.2	Streuungsmaße	16
2.2.3	Konzentrationsmaße	17
2.2.4	Maße für Schiefe und Wölbung	20
2.3	Quantilplots (Quantil-Grafiken)	22
2.4	Dichteschätzung	25
2.5	Beschreibung und Exploration von bivariaten Datensätzen	27
2.5.1	Grafische Darstellung von bivariaten Datensätzen	27
2.5.2	Zusammenhangsmaße	29
2.5.3	Einfache lineare Regression	33
3	Punktschätzer	41
3.1	Parametrisches Modell	41
3.2	Parametrische Familien von statistischen Prüfverteilungen	42
3.2.1	Gamma-Verteilung	42
3.2.2	Student-Verteilung (t-Verteilung)	45
3.2.3	Fisher-Snedecor-Verteilung (F-Verteilung)	48
3.3	Punktschätzer und ihre Grundeigenschaften	50
3.3.1	Eigenschaften von Punktschätzern	50
3.3.2	Schätzer des Erwartungswertes und empirische Momente	52
3.3.3	Schätzer der Varianz	54
3.3.4	Eigenschaften der Ordnungsstatistiken	62
3.3.5	Empirische Verteilungsfunktion	64
3.4	Methoden zur Gewinnung von Punktschätzern	72
3.4.1	Momentenschätzer	72
3.4.2	Maximum-Likelihood-Schätzer	74
3.4.3	Bayes-Schätzer	83
3.4.4	Resampling-Methoden zur Gewinnung von Punktschätzern	86

3.5	Weitere Güteeigenschaften von Punktschätzern	89
3.5.1	Ungleichung von Cramér-Rao	89
3.5.2	Bedingte Erwartung	94
3.5.3	Suffizienz	97
3.5.4	Vollständigkeit	101
3.5.5	Bester erwartungstreuer Schätzer	103
4	Konfidenzintervalle	106
4.1	Einführung	106
4.2	Ein-Stichproben-Probleme	108
4.2.1	Normalverteilung	108
4.2.2	Konfidenzintervalle aus stochastischen Ungleichungen	110
4.2.3	Asymptotische Konfidenzintervalle	111
4.3	Zwei-Stichproben-Probleme	113
4.3.1	Normalverteilte Stichproben	113
4.3.2	Poissonverteilte Stichproben	115
	Literatur	118
	Index	120

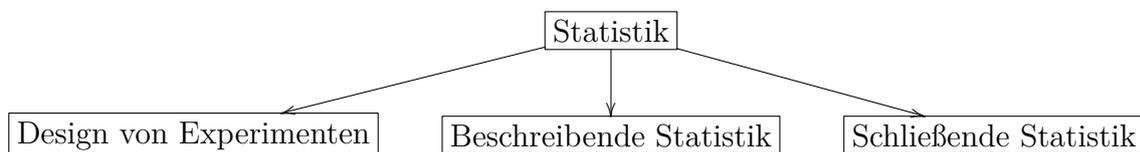
1 Einführung

1.1 Typische Fragestellungen, Aufgaben und Ziele der Statistik

Im alltäglichen Sprachgebrauch versteht man unter „Statistik“ eine Darstellung von Ergebnissen des Zusammenzählens von Daten und Fakten jeglicher Art, wie z.B. ökonomischen Kenngrößen, politischen Umfragen, Daten der Marktforschung, klinischen Studien in der Biologie und Medizin, usw.

Die *mathematische Statistik* jedoch kann viel mehr. Sie arbeitet mit *Daten-Stichproben*, die nach einem bestimmten Zufallsmechanismus aus der *Grundgesamtheit* aller Daten, die in Folge von Beobachtung, Experimenten (reale Daten) oder Computersimulation (synthetische Daten) erhoben wurden. Dabei beschäftigt sich die mathematische Statistik mit folgenden Fragestellungen:

1. Wie sollen die Daten gewonnen werden? (Design von Experimenten)
2. Wie sollen (insbesondere riesengroße) Datensätze beschrieben werden, um die Gesetzmäßigkeiten und Strukturen in ihnen entdecken zu können? (Beschreibende (deskriptive) und explorative Statistik)
3. Welche Schlüsse kann man aus den Daten ziehen? (Schließende oder induktive Statistik)



In dieser einführenden Vorlesung werden wir Teile der beschreibenden und schließenden Statistik kennenlernen, wobei die Datenerhebung aus Platzgründen ausgelassen wird. Die *Arbeitsweise eines Statistikers* sieht folgendermaßen aus:

1. *Datenerhebung*
2. *Visualisierung und beschreibende Datenanalyse*
3. *Datenbereinigung* (z.B. Erkennung fehlerhafter Messungen, Ausreißern, usw.)
4. *Explorative Datenanalyse* (Suche nach Gesetzmäßigkeiten)
5. *Modellierung der Daten* mit Methoden der Stochastik
6. *Modellanpassung* (Schätzung der Modellparameter)
7. *Modellvalidierung* (wie gut war die Modellanpassung?)

Pflanze	1	2	3	4	5	6	7	8	9	10
rund	45	27	24	19	32	26	88	22	28	25
kantig	12	8	7	10	11	6	24	10	6	7
Verhältnis ... : 1	3,8	3,4	3,4	1,9	2,9	4,3	3,7	2,2	4,7	3,6

Tab. 1.1: Ergebnisse für die 10 Pflanzen des ersten Versuchs von Mendel

8. *Schließende Datenanalyse:*

- Konstruktion von *Vertrauensintervallen* (Konfidenzintervallen) für Modellparameter und deren Funktionen,
- Tests statistischer Hypothesen,
- Vorhersage von Zielgrößen (z.B. auf Basis modellbezogener Computersimulation).

Uns werden in diesem Vorlesungsskript vor allem die Arbeitspunkte 2), 4)–6) und 8) beschäftigen.

Beispiel 1.1.1

Nachfolgend geben wir einige typische Fragestellungen der Statistik an Beispielen von Datensätzen:

1. *Statistische Herleitung von Grundsätzen der biologischen Evolution (Mendel, 1865):*

Es wurden Nachkommen von zwei Erbsensorten, die sich in der Samenform unterscheiden, gezüchtet: die erste Sorte hat runde, die zweite kantige Erbsen. Johann Gregor Mendel hat festgestellt, dass sich runde Samen dominant vererben. Dabei werden bei einer Bestäubung von Pflanzen der einen Sorte mit Pollen der anderen alle Nachkommen runde Samen zeigen, die genetisch heterozygot sind, d.h., beide Allele aufweisen. Kreuzt man diese hybriden Pflanzen, so zeigen sie runde und kantige Samen im Verhältnis 3 : 1 (Spaltungs- und Dominanzregeln von Mendel). Bei der statistischen Überprüfung seiner Vermutungen erhielt Mendel 5475 runde und 1850 kantige Samen, die somit im Verhältnis 2,96 : 1 stehen. In der Tabelle 1.1 sind Ergebnisse für die ersten 10 Pflanzen gezeigt. Man sieht, dass das oben genannte Verhältnis zufällig um 3 : 1 schwankt. Durch die Bildung des Mittels über das Gesamtkollektiv der Daten wird die Gesetzmäßigkeit 3 : 1 gefunden (explorative Statistik).

2. *Kreditwürdigkeit bei Kreditvergabe*

Die Banken sind offensichtlich daran interessiert, Bankkredite an Kunden zu vergeben, die in der Zukunft solvent bleiben, also die Kreditraten regelmäßig zurückzahlen können. Um die Kreditwürdigkeit zu überprüfen, werden Umfragen gemacht, wobei die Antworten unter anderem in folgenden Variablen kodiert werden:

- X_1 Laufendes Konto bei der Bank (1 = nein, 2 = ja und durchschnittlich geführt, 3 = ja und gut geführt)
- X_2 Laufzeit des Kredits in Monaten
- X_3 Kredithöhe in €
- X_4 Rückzahlung früherer Kredite (gut/ schlecht)

X_1 : laufendes Konto	Y	
	1	0
nein	45,0	19,9
gut	15,3	49,7
mittel	39,7	30,2
X_3 : Kredithöhe in €	1	0
$0 < \dots \leq 500$	1,00	2,14
$500 < \dots \leq 1000$	11,33	9,14
$1000 < \dots \leq 1500$	17,00	19,86
$1500 < \dots \leq 2500$	19,67	24,57
$2500 < \dots \leq 5000$	25,00	28,57
$5000 < \dots \leq 7500$	11,33	9,71
$7500 < \dots \leq 10000$	6,67	3,71
$10000 < \dots \leq 15000$	7,00	2,00
$15000 < \dots \leq 20000$	1,00	0,29
X_4 : Frühere Kredite	1	0
gut	82,33	94,85
schlecht	17,66	5,15
X_5 : Verwendungszweck	1	0
privat	57,53	69,29
beruflich	42,47	30,71

Tab. 1.2: Lernstichprobe zur Vergabe von Krediten

- X_5 Verwendungszweck (privat / geschäftlich)
- X_6 Geschlecht (weiblich / männlich)

Um an Hand eines ausgefüllten Fragebogens wie diesem eine Entscheidung über die Vergabe des Kredits treffen zu können, werden *Lernstichproben* herangezogen, bei denen das Ergebnis Y der erfolgten Kreditvergabe bekannt ist. Dabei bedeutet $Y = 0$ gut und $Y = 1$ schlecht. Betrachten wir eine solche Stichprobe einer süddeutschen Bank, die 1000 Umfragebögen umfasst. Dabei sind 700 kreditwürdig und 300 davon nicht kreditwürdig gewesen. Die Tabelle 1.2 zeigt Prozentzahlen dieses Datensatzes für ausgewählte Merkmale X_i . Dabei ist es möglich, mit Hilfe statistischer Methoden (Regression) eine Kreditentscheidung bei einem Kunden an Hand dieser Lernprobe automatisch treffen zu können. Dieser Vorgang wird manchmal auch „statistisches Lernen“ genannt. Fragestellungen wie diese werden erst in Stochastik III (verallgemeinerte lineare Modelle) behandelt.

3. Korrosion von Legierungen

In diesem Beispiel wurde der Korrosionsgrad einer Kupfer-Nickel-Legierung in Abhängigkeit ihres Eisengehalts untersucht. Dazu wurden 13 verschiedene Räder mit dieser Legierung beschichtet und 60 Tage lang in Meerwasser gedreht. Danach wurde der Gewichtsverlust in mg pro dm^2 und Tag bestimmt. Aus dem Bild 1.1 ist zu sehen, dass die Korrosion in Abhängigkeit vom Eisengehalt linear abnimmt. Mit statistischen Methoden (einfache lineare Regression) kann die Geschwindigkeit dieser Abnahme geschätzt werden.

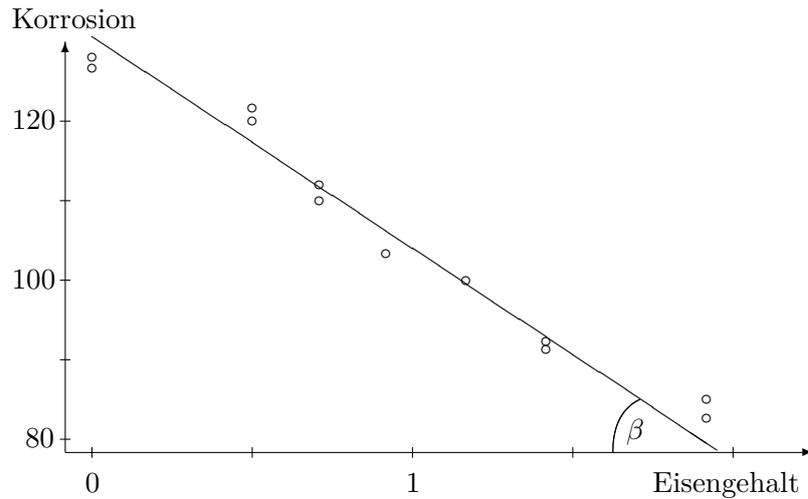
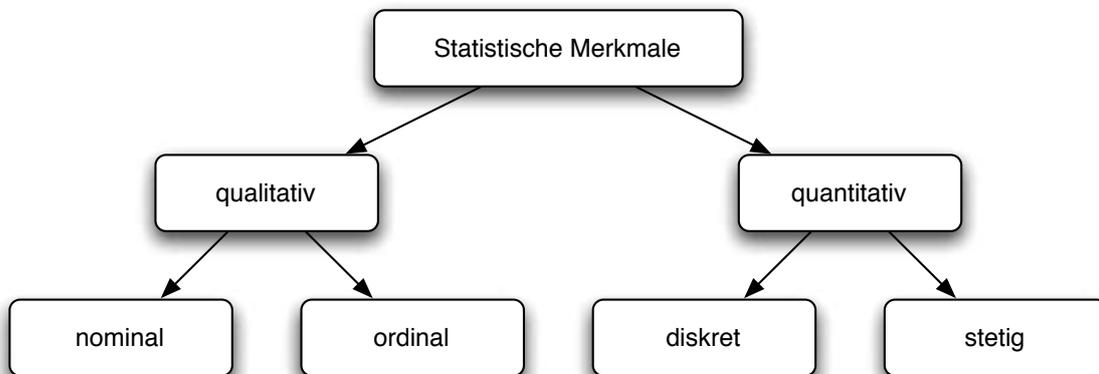


Abb. 1.1: Korrosion von Kupfer-Nickel-Legierung

1.2 Statistische Merkmale und ihre Typen

Die Daten, die zur statistischen Analyse vorliegen, können eine oder mehrere interessierende Größen (die auch *Variablen* oder *Merkmale* genannt werden) umfassen. Ihre Werte werden *Merkmalsausprägungen* genannt. In dem nachfolgenden Diagramm werden mögliche Typen der statistischen Merkmale gegeben.



Diese Typen entstehen in Folge der Klassifikation von Wertebereichen (Skalen) der Merkmale. Dennoch ist diese Einteilung nicht vollständig und kann bei Bedarf erweitert werden. Man unterscheidet *qualitative* und *quantitative* Merkmale. *Quantitative Merkmale* lassen sich inhaltlich gut durch Zahlen darstellen (z.B. Kredithöhe in €, Körpergewicht und Körpergröße, Blutdruck usw.). Sie können *diskrete* oder *stetige* Wertebereiche haben, wobei diskrete Merkmale isolierte Werte annehmen können (z.B. Anzahl der Schäden eines Versicherers pro Jahr). Stetige Wertebereiche hingegen sind überabzählbar. Dennoch liegen in der Praxis stetige Merkmale in gerundeter Form vor (z.B. Körpergröße auf cm gerundet, Geldbeträge auf € gerundet usw.).

Im Gegensatz zu den quantitativen Merkmalen sind die Inhalte der *qualitativen Merkmale*, wie z.B. Blutgruppe (0, A, B und AB) oder Familienstand (ledig, verheiratet, verwitwet),

nicht sinnvoll durch Zahlen darzustellen. Sie können zwar formell mit Zahlen kodiert werden (z.B. bei Blutgruppen $0 = 0$, $A = 1$, $B = 2$, $AB = 3$), aber solche Kodierungen stellen keinen inhaltlichen Zusammenhang zwischen Ausprägungen und Zahlen-Codes dar sondern dienen lediglich der besseren Identifikation der Merkmale auf einem Rechner. Es ist insbesondere unsinnig, Mittelwerte und ähnliches von solchen Codes zu bilden.

Ein qualitatives Merkmal mit nur 2 Ausprägungen (z.B. männlich / weiblich, Raucher / Nichtraucher) heißt *alternativ*. Ein qualitatives Merkmal kann *ordinal* (wenn sich eine natürliche lineare Ordnung in den Merkmalsausprägungen finden lässt, wie z.B. gut / mittel / schlecht bei Qualitätsbewertung in Umfragen oder sehr gut / gut / befriedigend / ausreichend / mangelhaft / ungenügend bei Schulnoten) oder *nominal* (wenn eine solche Ordnung nicht vorhanden ist) sein. Beispiele von nominalen Merkmalen sind Fahrzeugmarken in der KFZ-Versicherung (z.B. BMW, Peugeot, Volvo, usw.) oder Führerscheinklassen (A, B, C, \dots). Datenmerkmale können auch mehrdimensionale Ausprägungen haben. In dieser Vorlesung behandeln wir jedoch hauptsächlich eindimensionale Merkmale.

1.3 Statistische Daten und Stichproben

Aus den obigen Beispielen wird klar, dass ein Statistiker mit Datensätzen der Form (x_1, \dots, x_n) arbeitet, wobei die Einzeleinträge x_i aus einer Grundgesamtheit $G \subset \mathbb{R}^k$ stammen, die hypothetisch unendlich groß ist. Der vorliegende Datensatz (x_1, \dots, x_n) wird auch (*konkrete*) *Stichprobe* von Umfang n genannt. Die Menge B aller potentiell möglichen Stichproben bezeichnen wir als *Stichprobenraum* und setzen zur Vereinfachung der Notation $B = \mathbb{R}^{kn}$. In diesem Skript werden wir meistens die univariate statistische Analyse (also $k = 1$, ein eindimensionales Merkmal) betreiben. In der beschreibenden Statistik arbeitet man mit Stichproben (x_1, \dots, x_n) und ihren Funktionen, um diese Daten visualisieren zu können. Für die Aufgabe der schließenden Statistik jedoch reicht diese Datenebene nicht mehr aus. Daher wird die zweite Ebene der Betrachtung eingeführt, die sogenannte *Modellebene*. Dabei wird angenommen, dass die konkrete Stichprobe (x_1, \dots, x_n) eine *Realisierung* eines stochastischen Modells (X_1, \dots, X_n) darstellt, wobei X_1, \dots, X_n (meistens unabhängige identisch verteilte) Zufallsvariablen auf einem (nicht näher spezifiziertem) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ sind. Diese Zufallsvariablen X_i , $i = 1, \dots, n$ können als konsequente Beobachtungen eines Merkmals interpretiert werden. In Bsp. 1.1.1, 1) z.B. die Erbsenform mit

$$X_i = \begin{cases} 0, & \text{falls Erbse } i \text{ rund,} \\ 1, & \text{falls Erbse } i \text{ eckig,} \end{cases} \quad i = 1, \dots, n.$$

Der Vektor (X_1, \dots, X_n) wird dabei *Zufallsstichprobe* genannt. Man setzt weiter voraus, dass $EX_i^2 < \infty \forall i = 1, \dots, n$, damit man von der Varianz $\text{Var } X_i$ der Einzeleinträge sprechen kann. Es wird außerdem angenommen, dass ein $\omega \in \Omega$ existiert, sodass $X_i(\omega) = x_i \quad \forall i = 1, \dots, n$. Sei F die Verteilungsfunktion der Zufallsvariablen X_i . Eine der wichtigsten Aufgaben der Statistik ist die Bestimmung von F (man sagt, „Schätzung von F “) aus den konkreten Daten (x_1, \dots, x_n) . Dabei können auch Momente von F und ihre Funktionen (Erwartungswert, Varianz, Schiefe, usw.) von Interesse sein.

1.4 Stichprobenfunktionen

Um die obigen Aufgaben erfüllen zu können, braucht man gewisse Funktionen $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m \in \mathbb{N}$ auf dem Stichprobenraum, die diese Stichprobe bewerten.

Definition 1.4.1

Eine Borel-messbare Abbildung $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt *Stichprobenfunktion*. Wenn man auf der Modellebene mit einer Zufallsstichprobe (X_1, \dots, X_n) arbeitet, so heißt die Zufallsvariable

$$\varphi(X_1, \dots, X_n)$$

eine *Statistik*. In der Schätztheorie spricht man dabei von *Schätzern* und bei statistischen Tests wird $\varphi(X_1, \dots, X_n)$ *Teststatistik* genannt.

Beispiele für Stichprobenfunktionen sind unter anderen das *Stichprobenmittel*

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i,$$

die *Stichprobenvarianz*

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

und die *Ordnungsstatistiken*

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

die entstehen, wenn man eine Stichprobe, die aus quantitativen Merkmalen besteht, linear ordnet ($x_{(1)} = \min_{i=1, \dots, n} x_i, \dots, x_{(n)} = \max_{i=1 \dots n} x_i$). Weitere Beispiele und ihre Charakteristiken werden in Kapitel 2 gegeben.

2 Beschreibende Statistik

Sei eine konkrete Stichprobe (x_1, \dots, x_n) , $x_i \in \mathbb{R}$ gegeben, wobei die x_i als Realisierungen der Zufallsvariablen $X_i \stackrel{d}{=} X$ mit Verteilungsfunktion F interpretiert werden können.

2.1 Verteilungen und ihre Darstellungen

In diesem Abschnitt werden wir Methoden zur statistischen Beschreibung und grafischen Darstellung der (unbekannten) Verteilung F betrachten.

2.1.1 Häufigkeiten und Diagramme

Falls das quantitative Merkmal X eine endliche Anzahl von Ausprägungen $\{a_1, \dots, a_k\}$, $a_1 < a_2 < \dots < a_k$, besitzt, also

$$\mathbb{P}(X \in \{a_1, \dots, a_k\}) = 1,$$

dann kann eine Schätzung der Zähldichte $p_i = P(X = a_i)$ von X aus den Daten (x_1, \dots, x_n) grafisch dargestellt werden. Ähnliche Darstellungen sind für die Dichte $f(x)$ von absolut stetigen Merkmalen X möglich, wobei ihr Wertebereich C sich in k Klassen aufteilen lässt: $(c_{i-1}, c_i]$, $i = 1, \dots, k$, wobei $c_0 = -\infty$, $c_1 < \dots < c_{k-1}$, $c_k = \infty$ ist. Dann kann die Zähldichte $p_i = \mathbb{P}(X \in (c_{i-1}, c_i])$ gegeben durch

$$p_i = \int_{c_{i-1}}^{c_i} f(x) dx, \quad i = 0, \dots, k$$

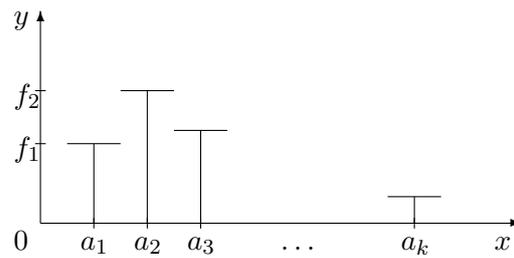
betrachtet werden.

Definition 2.1.1

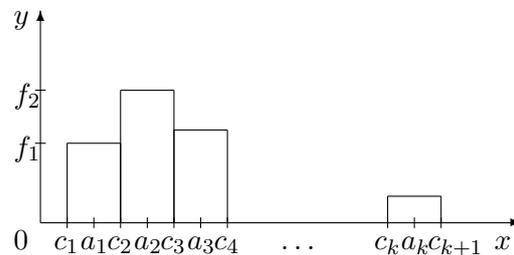
1. Die *absolute Häufigkeit* von Merkmalsausprägung a_i bzw. Klasse $(c_{i-1}, c_i]$, $i = 1, \dots, k$ ist $n_i = \#\{x_j, j = 1, \dots, n : x_j = a_i\}$ bzw. $n_i = \#\{x_j, j = 1, \dots, n : x_j \in (c_{i-1}, c_i]\}$.
2. Die *relative Häufigkeit* von Merkmalsausprägung a_i bzw. Klasse $(c_{i-1}, c_i]$ ist $f_i = n_i/n$, $i = 1, \dots, k$.

Es gilt offensichtlich $n = \sum_{i=1}^k n_i$, $0 \leq f_i \leq 1$, $\sum_{i=1}^k f_i = 1$. Die absoluten und relativen Häufigkeiten werden oft in Häufigkeitstabellen zusammengefasst. Zu ihrer Visualisierung dienen so genannte *Diagramme*. Es wird grundsätzlich zwischen *Histogrammen* und *Kreisdiagrammen* unterschieden.

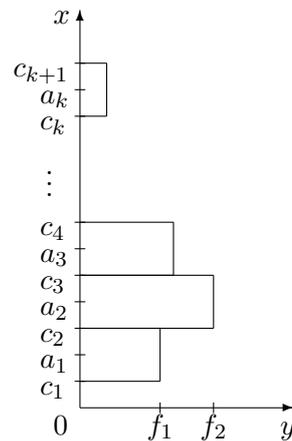
1. *Histogramme* werden gebildet, indem man die Paare (a_i, f_i) (bzw. $(1/2(c_1 + x_{(1)}), f_1)$, $(1/2(c_{i-1} + c_i), f_i)$, $i = 2, \dots, k-1$, $(1/2(c_{k-1} + x_{(n)}), f_k)$ im absolut stetigen Fall, wobei hier die Bezeichnung $a_i = 1/2(c_{i-1} + c_i)$ verwendet wird und $x_{(1)} < c_1$, $x_{(n)} > c_{k-1}$ angenommen wird.) auf der Koordinatenebene (x, y) folgendermaßen aufträgt:
 - *Stabdiagramm*: f_i wird als Höhe des senkrechten Strichs über a_i dargestellt:



- *Säulendiagramm*: genauso wie ein Stabdiagramm, nur werden Striche durch Säulen der Form $(c_{i-1}, c_i] \times f_i$ ersetzt, wobei im diskreten Fall die Aufteilung der reellen Achse $-\infty = c_0 < c_1 < c_2 < \dots < c_{k-1} < c_k = \infty$ in Intervalle beliebig vorgenommen werden kann.

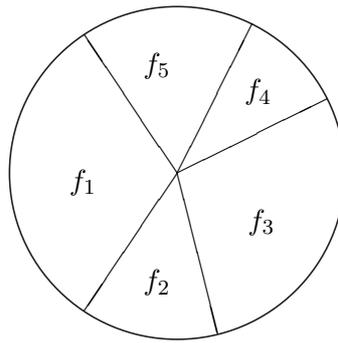


- *Balkendiagramm*: genauso wie Säulendiagramm, nur mit vertikalen statt horizontaler x -Achse.



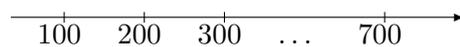
2. Kreisdiagramme (Tortendiagramme):

Ein Kreis wird in Segmente mit Öffnungswinkel α_i eingeteilt, die proportional zu f_i sind:
 $\alpha_i = 2\pi f_i, \quad i = 1, \dots, n.$

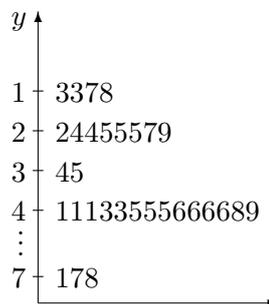


3. Stamm-Blatt-Diagramme (stem-leaf display):

Diese werden heutzutage relativ selten und nur für kleine Datensätze verwendet. Dabei arbeitet man mit Stichprobenwerten, die auf ganze Zahlen gerundet sind. Sei (x_1, \dots, x_n) eine Stichprobe von solchen Werten, die Ausprägungen eines quantitativen Merkmals sind. Zunächst teilt man den Wertebereich $[x_{(1)}, x_{(n)}]$ in Klassen gleicher Breite 10^d , $d \in \mathbb{N}$, wobei jede Klasse mit den ersten Ziffern der dazugehörigen Beobachtungen markiert wird. Zum Beispiel, wenn die Klasseneinteilung so aussieht



werden die Klassen $[100(i-1), 100i)$ mit den Zahlen i markiert und auf der y -Achse wie folgt aufgetragen:



Auf diese Weise wird der Stamm des Baumes festgelegt. In jeder Klasse ordnet man Beobachtungen ihrer Größe nach und rundet sie auf die Stelle, die nach der gewählten Genauigkeit des Stammes folgt. Als Beispiel erhält man aus $127 \rightarrow 130$, aus $652 \rightarrow 650$ usw. und trägt diese Beobachtungen als Blätter des Baums horizontal ihrer Reihenfolge nach als 3 in Klasse 1 und 5 in Klasse 6 auf. Dabei darf man nicht vergessen, die Einheit zu notieren: $1/3 = 130$, um sich das Rückrechnen zu ermöglichen. Bei der Wahl der Klassenanzahl m hält man sich an die Faustregel $m \approx 10 \log_{10} n$, um einerseits den Dateiverlust durch das unnötige Runden zu minimieren und andererseits das Diagramm so übersichtlich wie möglich zu halten.

Bemerkung 2.1.1

Die in Abschnitt 2.1.1 betrachteten Methoden dienen der Visualisierung von (Zähl-) Dichten der Verteilung eines beobachteten Merkmals X . Aus dem Histogramm kann z.B. die Interpretation der Form der Dichte abgelesen werden:

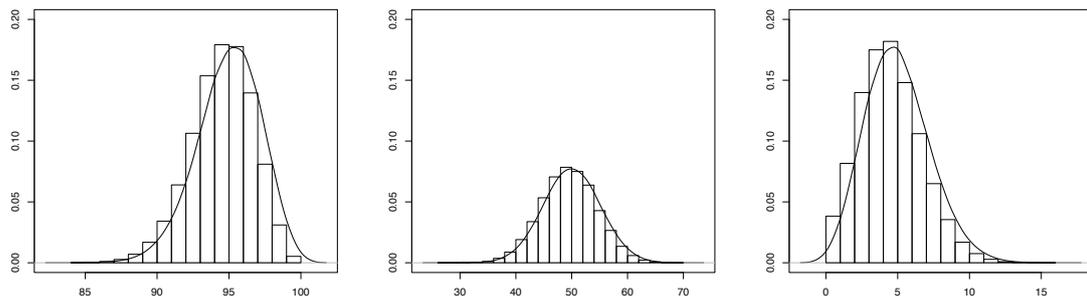


Abb. 2.1: Das Histogramm der Daten mit einer rechtssteilen (linksschiefen), symmetrischen und linkssteilen (rechtsschiefen) Verteilung und ihre Dichte.

Ist die zugrundeliegende Verteilung F_X symmetrisch bzw. linkssteil (rechtsschief) oder rechtssteil (linksschief) (vgl. Abb. 2.1) oder ist sie unimodal (d.h. eingipflig), bimodal (d.h. mit 2 Gipfeln) oder multimodal (also mit mehreren Gipfeln) (vgl. Abb. 2.2).

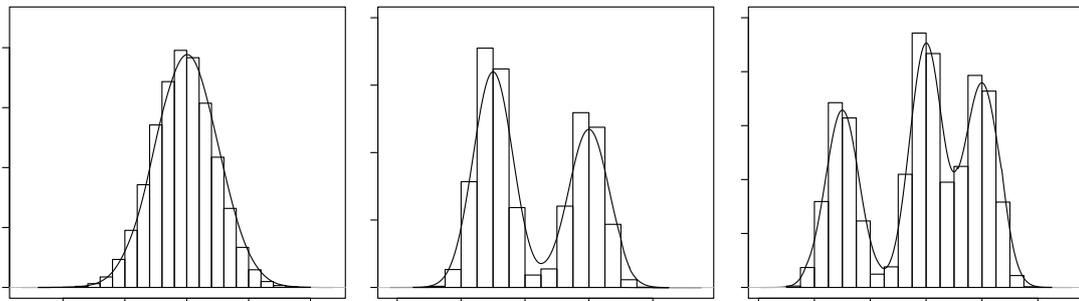


Abb. 2.2: Histogramm der Daten mit der Dichte einer unimodalen, bimodalen und multimodalen Verteilung

2.1.2 Empirische Verteilungsfunktion

Es sei eine konkrete Stichprobe (x_1, \dots, x_n) gegeben, die eine Realisierung des statistischen Modells (X_1, \dots, X_n) ist, wobei X_1, \dots, X_n unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F_X : X_i \stackrel{d}{=} X \sim F_X$ sind. Wie kann die unbekannte Verteilungsfunktion F_X aus den Daten (x_1, \dots, x_n) rekonstruiert (die Statistiker sagen „geschätzt“) werden? Dies ist mit Hilfe der sogenannten empirischen Verteilungsfunktion möglich:

Definition 2.1.2

1. Die Funktion $\hat{F}_n(x) = \#\{x_i : x_i \leq x, i = 1, \dots, n\}/n, \quad \forall x \in \mathbb{R}$ heißt *empirische Verteilungsfunktion der konkreten Stichprobe* (x_1, \dots, x_n) . Dabei gilt $\hat{F}_n : \mathbb{R}^{n+1} \rightarrow [0, 1]$, weil $\hat{F}_n(x) = \varphi(x_1, \dots, x_n, x)$.
2. Die mit $x \in \mathbb{R}$ indizierte Zufallsvariable $\hat{F}_n : \Omega \times \mathbb{R} \rightarrow [0, 1]$ heißt *empirische Verteilungsfunktion der Zufallsstichprobe* (X_1, \dots, X_n) , wenn

$$\hat{F}_n(x, \omega) = \hat{F}_n(x) = \frac{1}{n} \#\{X_i, i = 1, \dots, n : X_i(\omega) \leq x\}, \quad x \in \mathbb{R}.$$

Äquivalent zur Definition 2.1.2 kann man

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \leq x), \quad x \in \mathbb{R}$$

schreiben, wobei

$$\mathbb{I}(x \in A) = \begin{cases} 1, & x \in A \\ 0, & \text{sonst.} \end{cases}$$

Es gilt

$$\hat{F}_n(x) = \begin{cases} 1, & x \geq x_{(n)}, \\ \frac{i}{n}, & x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, \dots, n-1, \\ 0, & x < x_{(1)}. \end{cases}$$

für $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

Dabei ist die Höhe des Sprungs an Stelle $x_{(i)}$ gleich der relativen Häufigkeit f_i des Wertes $x_{(i)}$. Falls $x_{(i)} = x_{(i+1)}$ für ein $i \in \{1, \dots, n\}$, so tritt der Wert i/n nicht auf. In Abbildung 2.3 sieht man, dass $\hat{F}_n(x)$ eine rechtsstetige monoton nichtfallende Treppenfunktion ist, für die

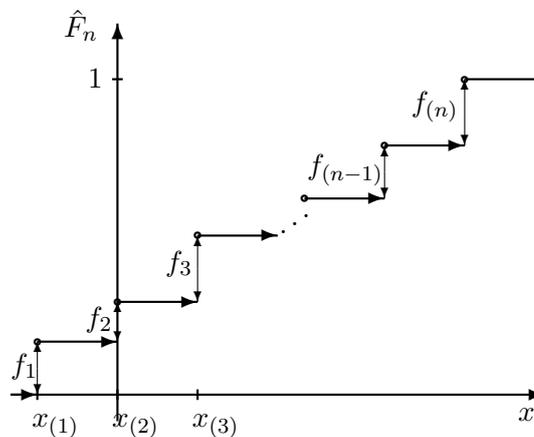


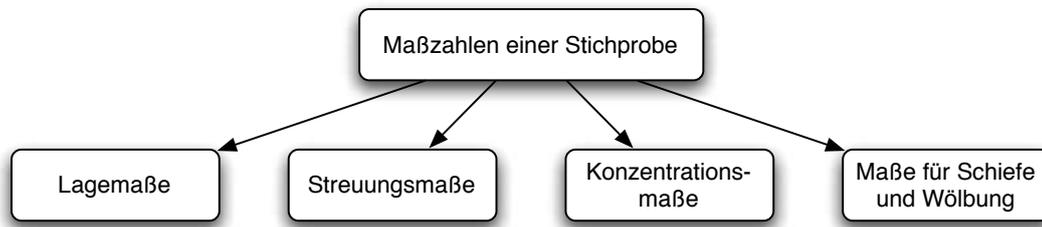
Abb. 2.3: Eine typische empirische Verteilungsfunktion

$$\hat{F}_n(x) \xrightarrow{x \rightarrow -\infty} 0, \quad \hat{F}_n(x) \xrightarrow{x \rightarrow \infty} 1 \text{ gilt.}$$

Übungsaufgabe 2.1.1

Zeigen Sie, dass $\hat{F}_n(x)$ eine Verteilungsfunktion ist.

2.2 Beschreibung von Verteilungen



Es sei eine konkrete Stichprobe (x_1, \dots, x_n) gegeben. Im Folgenden werden Kennzahlen (die sogenannten Maße) dieser Stichprobe betrachtet, welche die wesentlichen Aspekte der der Stichprobe zugrundeliegenden Verteilung wiedergeben:

1. Wo liegen die Werte x_i (Mittel, Ordnungsstatistiken, Quantile)? \implies Lagemaße
2. Wie stark streuen die Werte x_i (Varianz) \implies Streuungsmaße
3. Wie stark sind die Werte x_i in gewissen Bereichen von \mathbb{R} konzentriert \implies Konzentrationsmaße
4. Wie schief bzw. gewölbt ist die Verteilung von X \implies Maße für Schiefe und Wölbung

2.2.1 Lagemaße

Man unterscheidet folgende wichtige Lagemaße:

- Mittelwerte: Stichprobenmittel (arithmetisch), geometrisches und harmonisches Mittel, gewichtetes Mittel, getrimmtes Mittel
- Ordnungsstatistiken und Quantile, insbesondere Median und Quartile
- Modus

Betrachten wir sie der Reihe nach:

1. *Mittelwertbildung*: Seit der Antike kennt man mindestens 3 Arten der *Mittelberechnung* von n Zahlen (x_1, \dots, x_n) :

- *arithmetisch*: $\bar{x}_n = 1/n \sum_{i=1}^n x_i$, $\forall x_1, \dots, x_n \in \mathbb{R}$,
- *geometrisch*: $x_n^g = \sqrt[n]{x_1 \cdot \dots \cdot x_n}$, $x_1, \dots, x_n > 0$,
- *harmonisch*: $x_n^h = \left(1/n \sum_{i=1}^n x_i^{-1}\right)^{-1}$, $x_1, \dots, x_n \neq 0$.

- a) Das *arithmetische Mittel* wird in der Statistik am meisten benutzt, weil es keine Voraussetzungen über den Wertebereich von x_1, \dots, x_n braucht. Es wird auch *Stichprobenmittel* genannt. Offensichtlich ist \bar{x}_n ein Spezialfall des sogenannten gewichteten Mittels $x_n^w = \sum_{i=1}^n w_i x_i$, wobei für die Gewichte $w_i \geq 0 \quad \forall i = 1, \dots, n$ und $\sum_{i=1}^n w_i = 1$ gilt. Als eine natürliche Gewichtewahl kommt $w_i = 1/n$, $\forall i = 1, \dots, n$ bei einer konkreten Stichprobe (x_1, \dots, x_n) in Frage. Die Summe aller Abweichungen von \bar{x}_n ist Null, denn $\sum_{i=1}^n (x_i - \bar{x}_n) = n\bar{x}_n - n\bar{x}_n = 0$, d.h. \bar{x}_n stellt geometrisch

den Schwerpunkt der Werte x_i dar, falls jedem Punkt eine Einheitsmasse zugeordnet wird. Wenn es in der Stichprobe große Ausreißer gibt, so beeinflussen sie das Stichprobenmittel entscheidend und erschweren so die objektive Datenanalyse. Deshalb verwendet man oft die robuste Version des arithmetischen Mittels, das sogenannte *getrimmte Mittel*:

$$\tilde{x}_n^{(k)} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_{(i)},$$

bei dessen Berechnung die k kleinsten und k größten Ausreißer ausgelassen werden, wobei $k \ll n/2$.

- b) Das *geometrische Mittel* wird hauptsächlich bei der Beobachtung von Wachstums- und Zinsfaktoren verwendet. Sei $x_i = B_i/B_{i-1}$, $i = 1, \dots, n$ der Wachstumsfaktor des Merkmals B_i , das in den Jahren $i = 1, \dots, n$ beobachtet wurde (z.B. Inflationsfaktor). Dann ist $B_n = B_0 \cdot x_1 \cdot \dots \cdot x_n$ und somit wäre der Zins im Jahre n

$$B_n^g = B_0 \cdot x_1 \cdot \dots \cdot x_n = B_0 \cdot (x_n^g)^n.$$

Für das geometrische Mittel gilt

$$\log x_n^g = \frac{1}{n} \sum_{i=1}^n \log x_i \leq \log \left(\frac{1}{n} \sum_{i=1}^n x_i \right)$$

wegen der Konkavität des Logarithmus, d.h. $\log x_n^g = \overline{\log x_n} \leq \log \bar{x}_n$ und somit $x_n^g \leq \bar{x}_n$, wobei $x_n^g = \bar{x}_n$ genau dann, wenn $x_1 = \dots = x_n$.

- c) Das *harmonische Mittel* wird bei der Ermittlung von z.B. durchschnittlicher Geschwindigkeiten gebraucht.

Beispiel 2.2.1

Seien x_i Geschwindigkeiten mit denen Bauteile eine Produktionslinie der Länge l durchlaufen. Die gesamte Bearbeitungszeit ist $l/x_1 + \dots + l/x_n$ und die Durchschnittslaufgeschwindigkeit

$$\frac{l + \dots + l}{l/x_1 + \dots + l/x_n} = x_n^h.$$

Es gilt $x_{(1)} \leq x_n^h \leq x_n^g \leq \bar{x}_n \leq x_{(n)}$ und $x_n^h = (x_n^g)^2 / \bar{x}_n$ für $x_i > 0$, $i = 1, \dots, n$.

Übungsaufgabe 2.2.1

Beweisen Sie diese Relation per Induktion bzgl. n .

2. Ordnungsstatistiken und Quantile

Definition 2.2.1

Die *Ordnungsstatistiken* $x_{(i)}$, $i = 1, \dots, n$ der Stichprobe (x_1, \dots, x_n) sind durch die messbare Permutation $\varphi(x_1, \dots, x_n)$ gegeben, so dass

$$x_{(i)} = \min \{x_j : \#\{k : x_k \leq x_j\} \geq i\}, \quad \forall i = 1, \dots, n.$$

Somit gilt $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Dieselbe Definition kann auch auf der Modellebene gegeben werden.

Definition 2.2.2

- a) Sei nun X die Zufallsvariable, die das Merkmal modelliert. Sei F_X ihre Verteilungsfunktion. Die verallgemeinerte Inverse von F_X , definiert durch

$$F_X^{-1}(y) = \inf \{x : F_X(x) \geq y\}, \quad y \in [0, 1],$$

heißt *Quantilfunktion* von F_X bzw. X . Es gilt $F_X^{-1} : [0, 1] \rightarrow \mathbb{R} \cup \{\pm\infty\}$. Die Zahl $F_X^{-1}(\alpha)$, $\alpha \in [0, 1]$ wird α -*Quantil* von F_X genannt.

- b) • $F_X^{-1}(0,25)$ heißt *unteres Quartil*,
 • $F_X^{-1}(0,75)$ heißt *oberes Quartil*,
 • $F_X^{-1}(0,5)$ heißt der *Median* der Verteilung von X .

Zwischen Ordnungsstatistiken und Quantilen besteht ein enger Zusammenhang. So bedeutet $F_X^{-1}(\alpha)$, $\alpha \in (0, 1)$, dass ca. $\alpha \cdot 100\%$ aller Merkmalsausprägungen in der Stichprobe (x_1, \dots, x_n) unter $F_X^{-1}(\alpha)$ und ca. $(1 - \alpha) \cdot 100\%$ über $F_X^{-1}(\alpha)$ liegen (im absolut stetigen Fall). Insbesondere gilt $F_X^{-1}(\alpha) \approx x_{(\lfloor n\alpha \rfloor)}$, deshalb werden Ordnungsstatistiken auch *empirische Quantile* genannt. Dabei ist x_α definiert als

$$x_\alpha = \begin{cases} x_{(\lfloor n\alpha \rfloor + 1)}, & n\alpha \notin \mathbb{N} \\ 1/2(x_{(\lfloor n\alpha \rfloor)} + x_{(\lfloor n\alpha \rfloor + 1)}), & n\alpha \in \mathbb{N} \end{cases}.$$

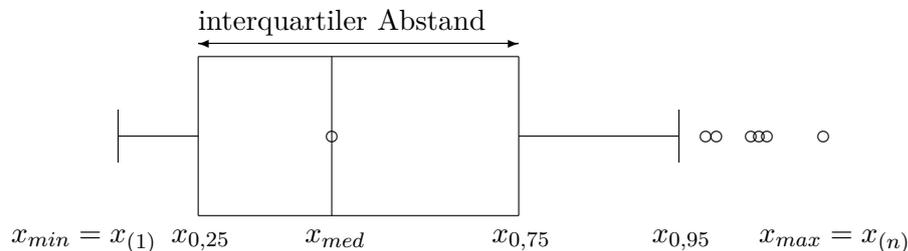
Dies ist die allgemeine Definition des *empirischen α -Quantils*.

Der *empirische Median* ist

$$x_{med} = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ ungerade} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right), & n \text{ gerade.} \end{cases}$$

Somit sind mindestens 50% aller Stichprobenwerte kleiner gleich und 50% größer gleich x_{med} . Der Median ist ein Lagemaß, das ein robuster Ersatz für den Mittelwert darstellt, denn er ist bzgl. Ausreißern in der Stichprobe nicht sensibel.

Die oben genannten Statistiken werden in einem *Box-Plot* zusammengefasst und grafisch dargestellt:



Manchmal werden $x_{(1)}$ und $x_{(n)}$ durch $x_{0,05}$ und $x_{0,95}$ ersetzt. Die restlichen Werte werden darüber hinaus als Einzelpunkte auf der x -Achse abgebildet. Dann liegt ein sogenannter *modifizierter Box-Plot* vor.

3. *Modus*: Sei (x_1, \dots, x_n) eine Stichprobe, die aus n unabhängigen Realisierungen des Merkmals X besteht. Sei $(p(x)) f(x)$ die (Zähl-) Dichte von X , wobei die Verteilung von X unimodal ist.

Definition 2.2.3

- a) Der Wert $x_{mod} = \arg \max f(x)$ ($\arg \max p(x)$) wird der *Modus der Verteilung von X* genannt (vgl. Abb. 2.4).
- b) Empirisch wird \hat{x}_{mod} als $\frac{c_{m-1} + c_m}{2}$ für $m = \arg \max f_i$ definiert, also als die Mitte des Intervalls mit der größten Häufigkeit des Vorkommens in der Stichprobe, falls dieser eindeutig bestimmbar ist.

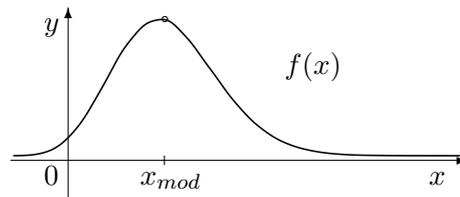


Abb. 2.4: Veranschaulichung des Modus

Den Mittelwert \bar{x}_n , Median x_{med} und Modus x_{mod} kann man auch wie folgt definieren:

$$\bar{x}_n = \arg \min_{x \in \mathbb{R}} \sum_{i=1}^n (x_i - x)^2$$

$$x_{med} = \arg \min_{x \in \mathbb{R}} \sum_{i=1}^n |x_i - x|$$

$$\hat{x}_{mod} = \frac{c_{m-1} + c_m}{2}, \quad \text{wobei } m = \arg \min_{j=1, \dots, n} \sum_{i=1}^n \mathbb{I}(x_i \notin (c_{j-1}, c_j])$$

Übungsaufgabe 2.2.2

Zeigen Sie die Äquivalenz der oben genannten Definitionen des Mittelwerts \bar{x}_n , Medians x_{med} und des Modus x_{mod} zu den bekannten Definitionen.

Die Größen \bar{x}_n , x_{med} und \hat{x}_{mod} können auch zur Beschreibung der Symmetrie einer unimodalen Verteilung F_X von Daten (x_1, \dots, x_n) verwendet werden, da

- bei symmetrischen Verteilung F_X gilt $\bar{x}_n \approx x_{med} \approx \hat{x}_{mod}$
- bei linkssteilen Verteilung F_X gilt $\hat{x}_{mod} < x_{med} < \bar{x}_n$
- bei rechtssteilen Verteilung F_X gilt $\bar{x}_n < x_{med} < \hat{x}_{mod}$.

2.2.2 Streuungsmaße

Bekannte Streuungsmaße einer konkreten Stichprobe (x_1, \dots, x_n) sind die folgenden Größen:

- *Spannweite* $x_{(n)} - x_{(1)}$,
- *empirische Varianz* $\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$,
- *Stichprobenvarianz* $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n}{n-1} \bar{s}_n^2$,
- *empirische Standardabweichungen* $\bar{s}_n = \sqrt{\bar{s}_n^2}$, $s_n = \sqrt{s_n^2}$,
- *empirischer Variationskoeffizient* $\gamma_n = s_n/\bar{x}_n$, falls $\bar{x}_n > 0$.

Die Spannweite zeigt die *maximale Streuung* in den Daten, wobei sich die empirische Varianz mit der *mittleren quadratischen Abweichung* vom Stichprobenmittel auseinandersetzt. Hier sind einige Eigenschaften von \bar{s}_n^2 (bzw. s_n^2 , da sie sich nur durch einen Faktor unterscheiden):

Lemma 2.2.1

1. Für jedes $b \in \mathbb{R}$ gilt

$$\sum_{i=1}^n (x_i - b)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - b)^2$$

und somit für $b = 0$

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - \bar{x}_n^2) \quad \text{bzw.} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^2 - \bar{x}_n^2).$$

2. *Transformationsregel:*

Falls die Daten (x_1, \dots, x_n) linear transformiert werden, d.h. $y_i = ax_i + b$, $a \neq 0$, $b \in \mathbb{R}$, dann gilt

$$\bar{s}_{n,y}^2 = a^2 \bar{s}_{n,x}^2 \quad \text{bzw.} \quad \bar{s}_{n,y} = |a| \bar{s}_{n,x},$$

wobei

$$\bar{s}_{n,y}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2, \quad \bar{s}_{n,x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

Beweis 1. Es gilt:

$$\begin{aligned} \sum_{i=1}^n (x_i - b)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - b)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + 2 \sum_{i=1}^n (x_i - \bar{x}_n) \cdot (\bar{x}_n - b) + \sum_{i=1}^n (\bar{x}_n - b)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + 2(\bar{x}_n - b) \cdot \underbrace{\sum_{i=1}^n (x_i - \bar{x}_n)}_{=0} + n(\bar{x}_n - b)^2, \quad \forall b \in \mathbb{R}. \end{aligned}$$

2. Es gilt:

$$\bar{s}_{n,y}^2 = \frac{1}{n} \sum_{i=1}^n (ax_i + b - a\bar{x}_n - b)^2 = \frac{a^2}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = a^2 \bar{s}_{n,x}^2.$$

□

Der Skalierungsunterschied zwischen \bar{s}_n^2 und s_n^2 ist den Eigenschaften der *Erwartungstreue* von s_n^2 zu verdanken, die später im Laufe dieser Vorlesung behandelt wird, und besagt, dass für eine Zufallsstichprobe (X_1, \dots, X_n) mit X_i unabhängig identisch verteilt, $X_i \sim X$, $\text{Var } X = \sigma^2 \in (0, \infty)$ gilt $\mathbb{E}s_n^2 = \sigma^2$, wobei $\mathbb{E}\bar{s}_n^2 = \frac{n}{n-1}\sigma^2 \xrightarrow{n \rightarrow \infty} \sigma^2$. Das heißt, während bei der Verwendung von s_n^2 zur Schätzung von σ^2 kein Fehler „im Mittel“ gemacht wird, ist diese Aussage für \bar{s}_n^2 nur asymptotisch (für große Datenmengen n) richtig.

Aufgrund von $\sum_{i=1}^n (x_i - \bar{x}_n) = 0$ ist z.B. $x_n - \bar{x}_n$ durch $x_i - \bar{x}_n$, $i = 1, \dots, n-1$ bestimmt. Somit verringert sich die *Anzahl der Freiheitsgrade* in der Summe $\sum_{i=1}^n (x_i - \bar{x}_n)^2$ um 1 und somit scheint die Normierung $\frac{1}{n-1}$ plausibel zu sein.

Die *Standardabweichungen* \bar{s}_n und s_n werden verwendet, damit man die selben Einheiten (und nicht ihre Quadrate, also z.B. Euro und nicht Euro²) erhält. Für normalverteilte Stichproben ($X \sim N(\mu, \sigma^2)$) liefert \bar{s}_n auch die „*k*-Sigma-Regel“ (vgl. Vorlesung WR), die besagt, dass in den Intervallen

$$\begin{aligned} [\bar{x}_n - \bar{s}_n, \bar{x}_n + \bar{s}_n] & \quad \text{ca.} \quad 68\%, \\ [\bar{x}_n - 2\bar{s}_n, \bar{x}_n + 2\bar{s}_n] & \quad \text{ca.} \quad 95\%, \\ [\bar{x}_n - 3\bar{s}_n, \bar{x}_n + 3\bar{s}_n] & \quad \text{ca.} \quad 99\% \end{aligned}$$

aller Daten liegen.

Der Vorteil vom *empirischen Variationskoeffizienten* ist, dass er *maßstabsunabhängig* ist und somit den Vergleich von Streuungseigenschaften unterschiedlicher Stichproben zulässt.

2.2.3 Konzentrationsmaße

Insbesondere in den Wirtschaftswissenschaften interessiert man sich oft für die Konzentration von Merkmalsausprägungen in der Stichprobe, z.B. wie sich das Familieneinkommen einer demographischen Einheit auf unterschiedliche Einkommensbereiche (Vielverdiener, Mittelstand, Wenigverdiener) aufteilt, oder wie sich der Markt auf Marktanbieter aufteilt (Marktkonzentration). Dabei ist es wünschenswert, diese Relation mit Hilfe weniger Zahlen oder einer Grafik zum Ausdruck zu bringen. Dies ist mit Hilfe folgender Stichprobenfunktionen möglich:

- *Lorenzkurve* L ,
- *Gini-Koeffizient* G ,
- *Konzentrationsrate* CR_g ,
- *Herfindahl-Index* H .

1. Die Lorenzkurve wurde von M. Lorenz am Anfang des XX. Jahrhunderts für die Charakterisierung der Vermögenskonzentration benutzt. Sei (x_1, \dots, x_n) eine Stichprobe, die in aufsteigender Reihenfolge geordnet werden muss: $(x_{(1)}, \dots, x_{(n)})$. Die *Lorenzkurve* verbindet Punkte

$$(0, 0), (u_1, v_1), \dots, (u_n, v_n), (1, 1)$$

durch Liniensegmente, wobei $u_j = j/n$ der Anteil der j kleinsten Merkmalsträger und $v_j = \sum_{i=1}^j x_{(i)} / \sum_{i=1}^n x_i$ die kumulierte relative Merkmalssumme ist. Der Grundgedanke

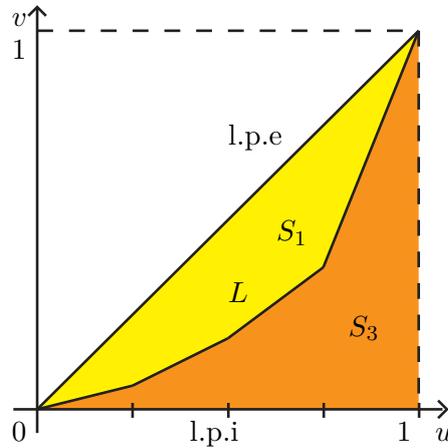


Abb. 2.5: Abbildung einer typischen Lorenzkurve

ist darzustellen, welcher Anteil des Merkmalsträgers auf welchen Anteil der Gesamtmerkmalssumme entfällt. Zum Beispiel lassen sich dadurch Aussagen wie etwa „Auf 20% aller Haushalte im Land entfällt 78% des Gesamteinkommens“ machen. Eine Interpretation der Lorenzkurve L ist nur an den Knoten (u_j, v_j) möglich: „Auf $u_j \cdot 100\%$ der kleinsten Merkmalsträger konzentrieren sich $v_j \cdot 100\%$ der Merkmalssumme“. Dabei liegt L auf $[0, 1]^2$ immer zwischen der „line of perfect equality“ (l.p.e.) $v_i = u_i \quad \forall i$ (Einkommen ist absolut gleichmäßig—also „gerecht“—verteilt) und „line of perfect inequality“ (l.p.i.) $v = 0, u \in [0, 1)$ und $(1, 1)$ (das Gesamteinkommen besitzt nur die reichste Familie) und ist immer monoton und konvex. Auf Modellebene gibt es ein Analogon der Lorenzkurve. Dieses ist

$$L = \left\{ (u, v) \in [0, 1]^2 : v = \frac{\int_0^u F_X^{-1}(t) dt}{\int_0^1 F_X^{-1}(t) dt}, \quad u \in [0, 1] \right\},$$

wobei

$$\mathbb{E}X = \int_0^1 F_X^{-1}(t) dt$$

(vgl. WR Satz 4.3.2). Dementsprechend können die Knoten (u_i, v_j) der oben eingeführten empirischen Lorenzkurve als

$$v_j = \frac{\sum_{i=1}^j \frac{x_{(i)}}{n}}{\bar{x}_n}$$

interpretiert werden.

- Der *Gini-Koeffizient* G ist gegeben durch $G = S_1/S_2$, wobei S_1 die Fläche zwischen der Lorenzkurve L und der Diagonalen $v = u$, S_2 die Fläche zwischen der Diagonalen und der u -Achse ($= 1/2|[0, 1]^2| = 1/2$) ist.

Satz 2.2.1 (Darstellung des Gini-Koeffizienten):

Es gilt

$$G = 2S_1 = \frac{2 \sum_{i=1}^n ix_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n}.$$

Beweis Beginnen wir damit, die Darstellung $G = (n+1)/n - 2\bar{v}_n$ zu zeigen. Nach Definition ist

$$G = \frac{S_1}{S_2} = \frac{S_2 - S_3}{S_2} = 1 - \frac{S_3}{S_2} = 1 - 2S_3,$$

wobei S_3 die Fläche zwischen der Lorenzkurve und der x -Achse ist (vgl. Abb. 2.5). Berechnen wir S_3 :

$S_3 = \sum_{j=1}^n F_j$, wobei $F_j = 1/n \cdot v_{j-1} + \frac{1}{2} \frac{1}{n} \cdot (v_j - v_{j-1}) = \frac{1}{2n}(v_j + v_{j-1})$ die Fläche unter einem Liniensegment der Lorenzkurve ist (vgl. Abb. 2.6). Es gilt

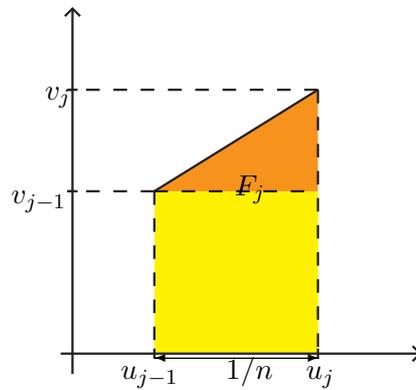


Abb. 2.6: Liniensegment der Lorenzkurve

$$S_3 = \frac{1}{2n} \sum_{j=1}^n (v_j + v_{j-1}) = \frac{1}{2n} \left(2 \sum_{j=1}^n v_j - 1 \right) = \bar{v}_n - \frac{1}{2n},$$

somit

$$G = 1 - 2\bar{v}_n + \frac{1}{n} = \frac{n+1}{n} - 2\bar{v}_n.$$

Beweisen wir jetzt, dass

$$G = \frac{2 \sum_{i=1}^n ix_{(i)}}{n \sum_{i=1}^n x_i} - \frac{n+1}{n}$$

ist. Sei $w = \sum_{i=1}^n ix_{(i)}$. Aufgrund der Definition von v_j gilt $s_j = \sum_{i=1}^j x_{(i)} = s_n \cdot v_j$, $\forall j = 1, \dots, n$ und $x_{(i)} = s_i - s_{i-1}$, $s_0 = 0$. Daher erhalten wir

$$\begin{aligned} w &= \sum_{i=1}^n i(s_i - s_{i-1}) = \sum_{i=1}^n is_i - \sum_{i=0}^{n-1} (i+1)s_i = ns_n - \sum_{i=0}^{n-1} s_i \\ &= (n+1)s_n - \sum_{i=1}^n s_i = (n+1)s_n - s_n \cdot \sum_{i=1}^n v_i = (n+1)s_n - s_n \cdot n\bar{v}_n \end{aligned}$$

und somit

$$\frac{2w}{ns_n} - \frac{n+1}{n} = \frac{2w - (n+1)s_n}{ns_n} = \frac{2(n+1)s_n - 2s_n n \bar{v}_n - (n+1)s_n}{ns_n} = \frac{n+1}{n} - 2\bar{v}_n = G.$$

□

Es gilt $G \in [0, (n-1)/n]$, wobei

$$\begin{aligned} G_{min} &= 0 & \text{bei } x_1 = x_2 = \dots = x_n & & \text{„perfect equality“}, \\ G_{max} &= \frac{n-1}{n} & \text{bei } x_1 = \dots = x_{n-1} = 0, x_n \neq 0 & & \text{„perfect inequality“}. \end{aligned}$$

Somit hängt G_{max} vom Datenumfang ab. Um dies zu vermeiden, betrachtet man oft den normierten Gini-Koeffizienten

$$G^* = \frac{G}{G_{max}} = \frac{n}{n-1}G \in [0, 1]$$

(Lorenz-Münzner-Koeffizient).

3. Konzentrationsrate CR_g :

In den Punkten 1) und 2) betrachteten wir die *relative Konzentration*, wie etwa bei der Fragestellung „Wieviel % der Familien teilen sich wieviel % des Gesamteinkommens?“. Dabei beantwortet die Konzentrationsrate die Frage „Wieviele Familien haben wieviel Prozent des Gesamteinkommens?“ für die g reichsten Familien, somit wird auch die absolute Anzahl aller Familien berücksichtigt.

Sei $g \in \{1, \dots, n\}$ und seien $x_{(1)} \leq \dots \leq x_{(n)}$ die Ordnungsstatistiken der Stichprobe (x_1, \dots, x_n) . Für $i \in \{1, \dots, n\}$ sei

$$p_i = \frac{x_{(i)}}{\sum_{j=1}^n x_j} = \frac{x_{(i)}}{n\bar{x}_n} \quad (2.2.1)$$

der Merkmalsanteil der i -ten Einheit.

Dann gibt die *Konzentrationsrate* $CR_g = \sum_{i=n-g+1}^n p_i$ wieder, welcher Anteil des Gesamteinkommens von g reichsten Familien gehalten wird.

4. Der *Herfindahl-Index* ist definiert durch $M = \sum_{i=1}^n p_i^2$, wobei der Merkmalsanteil p_i nach (2.2.1) definiert ist. Bei der gleichen Verteilung des Einkommens ($x_1 = x_2 = \dots = x_n$) gilt $H_{min} = 1/n$, bei völlig ungerechter Verteilung ($x_1 = \dots = x_{n-1} = 0, x_n \neq 0$) $H_{max} = 1$. Sonst gilt $H \in [H_{min}, H_{max}]$, also $1/n \leq H \leq 1$. H ist umso kleiner, je gerechter das Gesamteinkommen verteilt ist.

2.2.4 Maße für Schiefe und Wölbung

Im Vorlesungsskript WR, Abschnitt 4.5 S. 99 wurden folgende Maße für Schiefe bzw. Wölbung der Verteilung einer Zufallsvariable X eingeführt:

Schiefe oder Symmetriekoeffizient:

$$\gamma_1 = \frac{\mu_3'}{\sigma^3} = E(\tilde{X}^3),$$

wobei

$$\mu'_k = \mathbb{E}(X - \mathbb{E}X)^k, \quad \sigma^2 = \mu'_2 = \text{Var } X, \quad \tilde{X} = \frac{X - \mathbb{E}X}{\sigma}.$$

Wölbung (*Exzess*):

$$\gamma_2 = \frac{\mu'_4}{\sigma^4} - 3 = \mathbb{E}(\tilde{X}^4) - 3,$$

vorausgesetzt, dass $\mathbb{E}(X^4) < \infty$. Für ihre Bedeutung und Interpretation siehe die oben genannten Seiten des WR-Vorlesungsskriptes. Falls nun das Merkmal X statistisch in einer Stichprobe (x_1, \dots, x_n) beobachtet wird, wie können γ_1 und γ_2 aus diesen Daten geschätzt und interpretiert werden?

Als Schätzer für das k -te zentrierte Moment $\mu'_k = \mathbb{E}(X - \mathbb{E}X)^k$, $k \in \mathbb{N}$ schlagen wir

$$\hat{\mu}'_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^k$$

vor, die Varianz σ^2 wird durch

$$\hat{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

geschätzt. Somit bekommt man den Momentenkoeffizient der Schiefe (engl. „skewness“)

$$\hat{\gamma}_1 = \frac{\hat{\mu}'_3}{\hat{s}_n^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)^{3/2}}.$$

Falls die Verteilung von X links schief ist, überwiegen positive Abweichungen im Zähler und somit gilt $\hat{\gamma}_1 > 0$ für linksschiefe Verteilungen. Analog gilt $\hat{\gamma}_1 \approx 0$ für symmetrische und $\hat{\gamma}_1 < 0$ für rechtsschiefe Verteilungen.

Das *Wölbungsmaß von Fisher* (engl. „kurtosis“) ist gegeben durch

$$\hat{\gamma}_2 = \frac{\hat{\mu}'_4}{\hat{s}_n^4} - 3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)^2} - 3.$$

Falls $\hat{\gamma}_2 > 0$ so ist die Verteilung von X steilgipflig, für $\hat{\gamma}_2 < 0$ ist sie flachgipflig. Falls $X \sim N(\mu, \sigma^2)$, so gilt $\hat{\gamma}_2 \approx 0$. Die Ursache dafür ist, dass die steilgipfligen Verteilungen schwerere Tails haben als die flachgipfligen. Als Maß dient dabei die Normalverteilung, für die $\gamma_1 = \gamma_2 = 0$ und somit $\hat{\gamma}_1 \approx 0$, $\hat{\gamma}_2 \approx 0$. So definiert, sind $\hat{\gamma}_1$ und $\hat{\gamma}_2$ nicht resistent gegenüber Ausreißern. Eine robuste Variante von $\hat{\gamma}_1$ ist beispielsweise durch den sogenannten *Quantilkoeffizienten der Schiefe*

$$\hat{\gamma}_q(\alpha) = \frac{(x_{1-\alpha} - x_{med}) - (x_{med} - x_\alpha)}{x_{1-\alpha} - x_\alpha}, \quad \alpha \in (0, 1/2)$$

gegeben.

Für $\alpha = 0,25$ erhält man den Quartilkoeffizienten. $\hat{\gamma}_q(\alpha)$ misst den Unterschied zwischen der Entfernung des α - und $(1 - \alpha)$ -Quantils zum Median. Bei linkssteilen (bzw. rechtssteilen) Verteilungen liegt das (untere) x_α -Quantil näher an (bzw. weiter entfernt von) dem Median. Somit gilt

- $\hat{\gamma}_q(\alpha) > 0$ für linkssteile Verteilungen,

- $\hat{\gamma}_q(\alpha) < 0$ für rechtssteile Verteilungen,
- $\hat{\gamma}_q(\alpha) = 0$ für symmetrische Verteilungen.

Durch das zusätzliche Normieren (Nenner) gilt $-1 \leq \hat{\gamma}_q(\alpha) \leq 1$.

2.3 Quantilplots (Quantil-Grafiken)

Nach der ersten beschreibenden Analyse eines Datensatzes (x_1, \dots, x_n) soll überlegt werden, mit welcher Verteilung diese Stichprobe modelliert werden kann. Hier sind die sogenannten *Quantilplots* behilflich, da sie grafisch zeigen, wie gut die Daten (x_1, \dots, x_n) mit dem Verteilungsgesetz G übereinstimmen, wobei G die Verteilungsfunktion einer hypothetischen Verteilung ist.

Sei X eine Zufallsvariable mit (unbekannter) Verteilungsfunktion F_X . Auf Basis der Daten (X_1, \dots, X_n) , X_i unabhängig identisch verteilt und $X_i \stackrel{d}{=} X$ möchte man prüfen, ob $F_X = G$ für eine bekannte Verteilungsfunktion G gilt. Die Methode der *Quantil-Grafiken* besteht darin, dass man die entsprechenden Quantil-Funktionen \hat{F}_n^{-1} und G^{-1} von \hat{F}_n und G grafisch vergleicht. Hierzu

- plote man $G^{-1}(k/n)$ gegen $\hat{F}_n^{-1}(k/n) = X_{(k)}$, $k = 1, \dots, n$.
- Falls die Punktwolke

$$\left\{ \left(G^{-1}(k/n), X_{(k)} \right), \quad k = 1, \dots, n \right\}$$

näherungsweise auf einer Geraden $y = ax + b$ liegt, so sagt man, dass $F_X(x) \approx G\left(\frac{x-a}{b}\right)$, $x \in \mathbb{R}$.

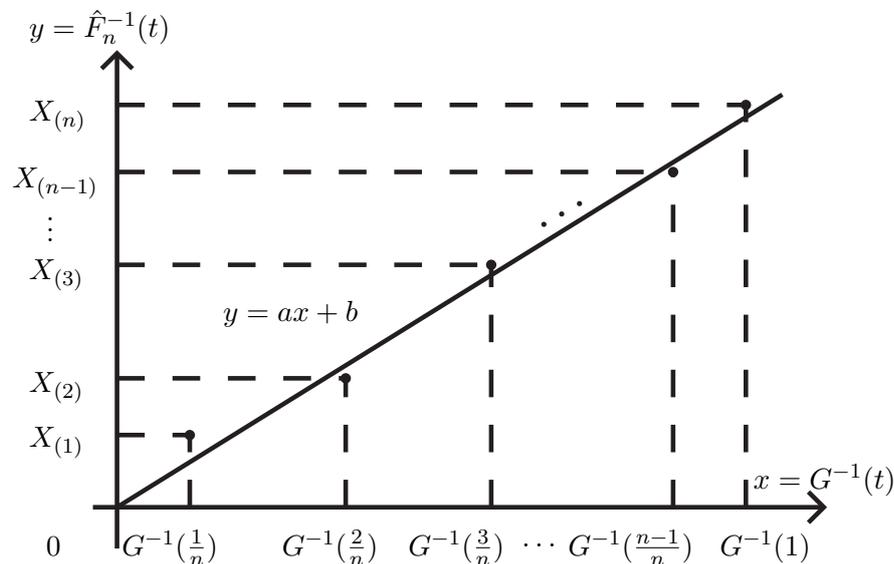


Abb. 2.7: Quantil-Grafik

Diese empirische Vergleichsmethode beruht auf folgenden Überlegungen:

- Man ersetzt die unbekannt Funktion F_X durch die aus den Daten berechenbare Funktion \hat{F}_n . Dabei macht man einen Fehler, der allerdings asymptotisch (für $n \rightarrow \infty$) klein ist. Dies folgt aus dem Satz 3.3.9 von Gliwenko-Cantelli, der besagt, dass

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_X(x) \right| \xrightarrow{n \rightarrow \infty} 0.$$

Der Vergleich der entsprechenden Quantil-Funktionen wird durch folgendes Ergebnis bestärkt: Falls $\mathbb{E}X < \infty$, dann gilt

$$\sup_{t \in [0,1]} \left| \int_0^t (\hat{F}_n^{-1}(y) - F_X^{-1}(y)) dy \right| \xrightarrow[n \rightarrow \infty]{\text{f.s.}} 0.$$

Somit setzt man bei der Verwendung der Quantil-Grafiken voraus, dass der Stichprobenumfang n ausreichend groß ist, um $\hat{F}_n^{-1} \approx F_X^{-1}$ zu gewährleisten.

- Man setzt zusätzlich voraus, dass die Gleichungen

$$\begin{aligned} y &= ax + b, \\ y &= F_X^{-1}(t), \\ x &= G^{-1}(t) \end{aligned}$$

für alle t (und nicht nur näherungsweise für $t = k/n$, $k = 1, \dots, n$) gelten. Daraus folgt, dass $G(x) = t = F_X(y) = F_X(ax + b)$ für alle x , oder $F_X(y) = G\left(\frac{y-b}{a}\right)$ für alle y , weil $x = \frac{y-a}{b}$ ist.

Aus praktischer Sicht ist es besser, Paare $\left(G^{-1}\left(\frac{k}{n+1}\right), X_{(k)}\right)$, $k = 1, \dots, n$ zu plotten. Dadurch wird vermieden, dass $G^{-1}(n/n) = G^{-1}(1) = \infty$ vorkommt, wie es zum Beispiel im Falle einer Verteilung G der Fall ist, bei der $F(x) < 1$ gilt für alle $x \in \mathbb{R}$. Tatsächlich gilt für $k = n$, dass $\frac{n}{n+1} < 1$ und somit $G^{-1}\left(\frac{n}{n+1}\right) < \infty$.

Beispiel 2.3.1 (Exponential-Verteilung, $G(x) = (1 - e^{-\lambda x}) \cdot \mathbb{I}(x \geq 0)$):

Es gilt $G^{-1}(y) = -1/\lambda \log(1 - y)$, $y \in (0, 1)$. So wird man beim Quantil-Plot Paare

$$\left(-\frac{1}{\lambda} \log\left(1 - \frac{k}{n+1}\right), X_{(k)}\right), \quad k = 1, \dots, n$$

zeichnen, wobei der Faktor $1/\lambda$ für die Linearität unwesentlich ist und weggelassen werden kann.

Beispiel 2.3.2 (Normalverteilung, $G(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$, $x \in \mathbb{R}$):

Leider ist die analytische Berechnung von Φ^{-1} mit einer geschlossenen Formel nicht möglich. Aus diesem Grund wird $\Phi^{-1}\left(\frac{k}{n+1}\right)$ numerisch berechnet und in Tabellen oder statistischen Software-Paketen (wie z.B. R) abgelegt. Um die empirische Verteilung der Daten mit der Normalverteilung zu vergleichen, trägt man Punkte mit Koordinaten

$$\left(\Phi^{-1}\left(\frac{k}{n+1}\right), X_{(k)}\right), \quad k = 1, \dots, n$$

auf der Ebene auf und prüft, ob sie eine Gerade bilden (vgl. Abb. 2.8).

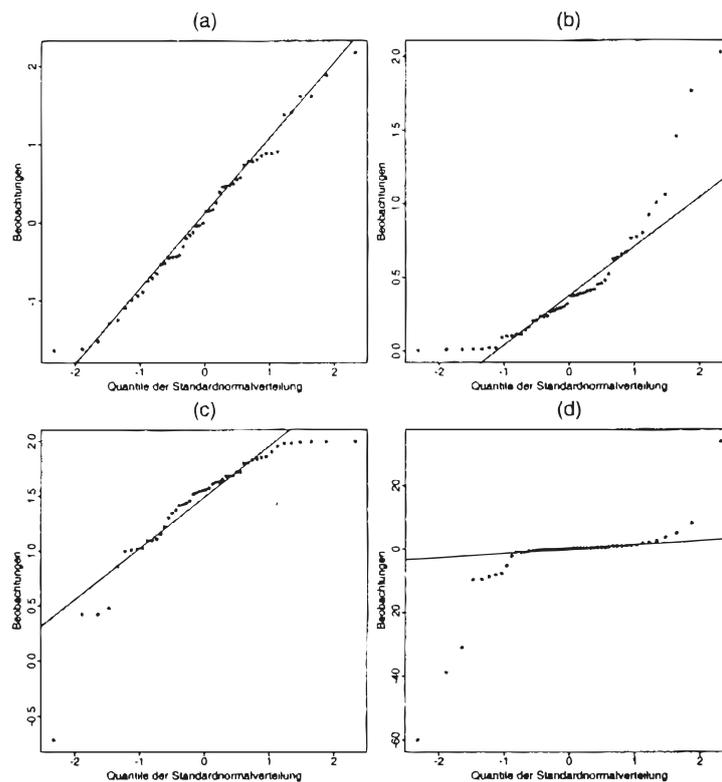


Abb. 2.8: QQ-Plot einer Normalverteilung (a), einer linkssteilen Verteilung (b), einer rechtssteilen Verteilung (c) und einer symmetrischen, aber stark gekrümmten Verteilung (d)

Übungsaufgabe 2.3.1

Entwerfen Sie die Quantil-Grafiken für den Vergleich der empirischen Verteilung mit der Lognormal und der Weibull-Verteilung.

Bemerkung 2.3.1

Falls $\bar{x}_n = 0$ und die Verteilung F_X linkssteil ist, so sind die Quantile von F_X kleiner als die von Φ . Somit ist der Normal-Quantilplot konvex. Falls $\bar{x}_n = 0$ und F_X rechtssteil ist, so wird der Normal-Quantilplot konkav sein.

Beispiel 2.3.3 (Haftpflichtversicherung (Belgien, 1992)):

In Abbildung 2.9 sind Ordnungsstatistiken der Stichprobe von $n = 227$ Schadenhöhen der Industrie-Unfälle in Belgien im Jahr 1992 (Haftpflichtversicherung) gegen Quantile von Exponential-, Pareto-, Standardnormal- und Weibull-Verteilungen geplottet. Im Bereich von Kleinschäden zeigen die Exponential- und Pareto-Verteilungen eine gute Übereinstimmung mit den Daten. Die Verteilung von mittelgroßen Schäden kann am besten durch die Normal- und Weibull-Verteilungen modelliert werden. Für Großschäden erweist sich die Weibull-Verteilung als geeignet.

Beispiel 2.3.4 (Rendite der BMW-Aktie):

In Abbildung 2.10 ist der Quantilplot für Renditen der BMW-Aktie beispielhaft zu sehen.

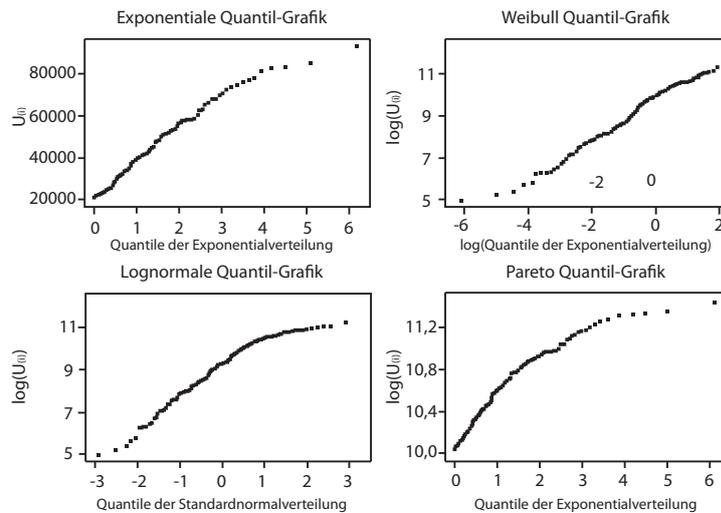


Abb. 2.9: Ordnungsstatistiken einer Stichprobe von Schadenhöhen der Industrie-Unfälle in Belgien im Jahr 1992

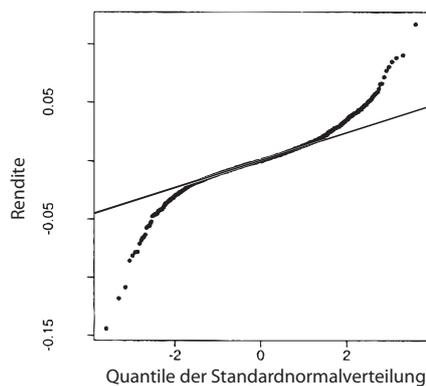


Abb. 2.10: Quantilplot der Rendite der BMW-Aktie

2.4 Dichteschätzung

Sei eine Stichprobe (x_1, \dots, x_n) von unabhängigen Realisierungen eines absolut stetig verteilten Merkmals X mit Dichte f_X gegeben. Mit Hilfe der in Abschnitt 2.1.1 eingeführten Histogramme lässt sich f_X grafisch durch eine Treppenfunktion \hat{f}_X darstellen. Dabei gibt es zwei entscheidende Nachteile der Histogrammdarstellung:

1. Willkür in der Wahl der Klasseneinteilung $[c_{i-1}, c_i]$,
2. Eine (möglicherweise) stetige Funktion f_X wird durch eine Treppenfunktion \hat{f}_X ersetzt.

In diesem Abschnitt werden wir versuchen, diese Nachteile zu beseitigen, indem wir eine Klasse von Kerndichteschätzern einführen, die (je nach Wahl des Kerns) auch zu stetigen Schätzern \hat{f}_X führen.

Definition 2.4.1

Der Kern $K(x)$ wird definiert als eine nicht-negative messbare Funktion auf \mathbb{R} mit der Eigenschaft $\int_{\mathbb{R}} K(x) dx = 1$.

Definition 2.4.2

Der *Kerndichteschätzer* der Dichte f_X aus den Daten (x_1, \dots, x_n) mit Kernfunktion $K(x)$ ist gegeben durch

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R},$$

wobei $h > 0$ die sogenannte *Bandbreite* ist.

Beispiele für Kerne:

1. *Rechteckskern:*

$$K(x) = 1/2 \cdot \mathbb{I}(x \in [-1, 1)).$$

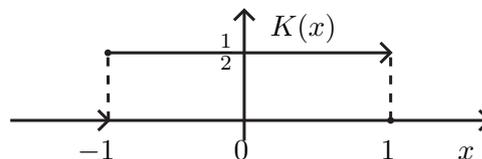
Dabei ist

$$\frac{1}{h} K\left(\frac{x - x_i}{h}\right) = \begin{cases} 1/(2h), & x_i - h \leq x < x_i + h, \\ 0, & \text{sonst,} \end{cases}$$

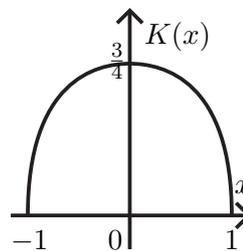
und somit

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^k K\left(\frac{x - x_i}{h}\right) = \frac{\#\{x_i \in [x - h, x + h)\}}{2nh},$$

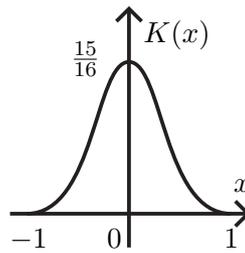
das auch *gleitendes Histogramm* genannt wird. Dieser Dichteschätzer ist (noch) nicht stetig, was durch die (besonders einfache rechteckige unstetige) Form des Kerns erklärt wird.

2. *Epanechnikov-Kern:*

$$K(x) = \begin{cases} 3/4(1 - x^2), & x \in [-1, 1) \\ 0, & \text{sonst.} \end{cases}$$

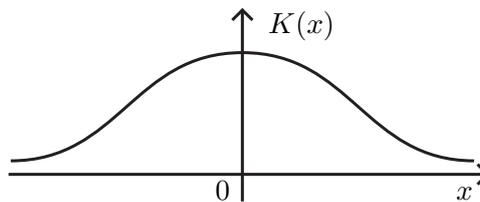
3. *Bisquare-Kern:*

$$K(x) = \frac{15}{16} \left((1 - x^2)^2 \cdot \mathbb{I}(x \in [-1, 1)) \right).$$



4. Gauss-Kern:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$



Dabei ist die Wahl der Bandbreite h entscheidend für die Qualität der Schätzung. Je größer $h > 0$, desto glatter wird \hat{f}_X sein und desto mehr „Details“ werden „herausgemittelt“. Für kleinere h wird \hat{f}_X rauer. Dabei können aber auch Details auftreten, die rein stochastischer Natur sind und keine Gesetzmäßigkeiten zeigen. Mit der adäquaten Wahl von h beschäftigen sich viele wissenschaftliche Arbeiten, die empirische Faustregeln, aber auch kompliziertere Optimierungsmethoden dafür vorschlagen. Insgesamt ist das Problem der optimalen Dichteschätzung in der Statistik immer noch offen.

2.5 Beschreibung und Exploration von bivariaten Datensätzen

Im Gegensatz zu der Datenlage in den Abschnitten 2.1 bis 2.4 betrachten wir im Folgenden Datensätze bestehend aus 2 Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) , die als Realisierungen von stochastischen Stichproben (X_1, \dots, X_n) und (Y_1, \dots, Y_n) aufgefasst werden, wobei X_1, \dots, X_n unabhängige identisch verteilte Zufallsvariablen mit $X_i \stackrel{d}{=} X \sim F_X$, Y_1, \dots, Y_n unabhängige identisch verteilte Zufallsvariablen mit $Y_i \stackrel{d}{=} Y \sim F_Y$ sind. Wir betrachten hier ausschließlich quantitative Merkmale X und Y . Es wird ein Zusammenhang zwischen X und Y vermutet, der an Hand von (konkreten) Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) näher untersucht werden soll. Mit anderen Worten, wir interessieren uns für die Eigenschaften der bivariaten Verteilung $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ des Zufallsvektors $(X, Y)^T$.

2.5.1 Grafische Darstellung von bivariaten Datensätzen

Um die Verteilung von (x_1, \dots, x_n) und (y_1, \dots, y_n) zu visualisieren, betrachten wir drei Möglichkeiten:

1. Streudiagramme
2. Zweidimensionale Histogramme

3. *Kerndichteschätzer* (im Falle eines absolut stetig verteilten Zufallsvektors $(X, Y)^T$)
1. *Streudiagramme* sind die erste sehr einfache und intuitive Visualisierungsmöglichkeit von bivariaten Daten. Um ein Streudiagramm zu erstellen, plottet man die „Punktwolke“ $(x_i, y_i)_{i=1, \dots, n}$ auf einer Koordinatenebene im \mathbb{R}^2 . Dabei zeigt die Form der Punktwolke, ob ein linearer ($y = ax + b$) bzw. polynomialer ($y = P_d(x)$) Zusammenhang in den Daten zu erwarten ist. Später werden solche Zusammenhänge im Rahmen der Regressionstheorie untersucht (vgl. Abschnitt 2.5.3 für die einfache lineare Regression).

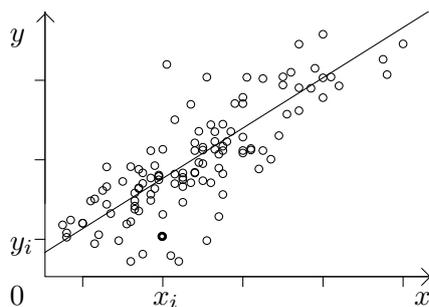


Abb. 2.11: Punktwolke

2. *Zweidimensionale Histogramme* dienen der Darstellung der bivariaten Zähldichte $p(x, y)$ des Zufallsvektors (X, Y) , falls er diskret verteilt ist, bzw. seiner Dichte $f(x, y)$ im Falle einer absolut stetigen Verteilung von (X, Y) aus den Daten (x_1, \dots, x_n) und (y_1, \dots, y_n) . Dabei teilt man den Wertebereich von X in Intervalle

$$[c_{i-1}, c_i), \quad i = 1, \dots, k, \quad -\infty = c_0 < c_1 < \dots < c_k = +\infty$$

und den Wertebereich von Y in Intervalle

$$[e_{i-1}, e_i), \quad i = 1, \dots, m, \quad -\infty = e_0 < e_1 < \dots < e_m = +\infty.$$

Bezeichnen wir

$$h_{ij} = \#\{(x_k, y_k), k = 1, \dots, n : x_k \in [c_{i-1}, c_i), y_k \in [e_{j-1}, e_j)\}$$

als die absolute Häufigkeit von (X, Y) in $[c_{i-1}, c_i) \times [e_{j-1}, e_j)$, $f_{ij} = h_{ij}/n$ als die relative Häufigkeit. Das zweidimensionale Histogramm setzt sich aus den Säulen mit Grundriss $[c_{i-1}, c_i) \times [e_{j-1}, e_j)$ und Höhe

$$\frac{h_{ij}}{(c_i - c_{i-1})(e_j - e_{j-1})}$$

für das Histogramm absoluter Häufigkeiten bzw.

$$\frac{f_{ij}}{(c_i - c_{i-1})(e_j - e_{j-1})}$$

für das Histogramm relativer Häufigkeiten zusammen, damit das Volumen dieser Säulen h_{ij} bzw. f_{ij} ist. Dabei hat solch ein Histogramm dieselben Vor- bzw. Nachteile wie ein ein-dimensionales, wenn es um die grafische Darstellung einer bivariaten Dichte $f(x, y)$ geht. Deshalb benutzt man oft Kerndichteschätzer, um eine glatte Darstellung zu bekommen.

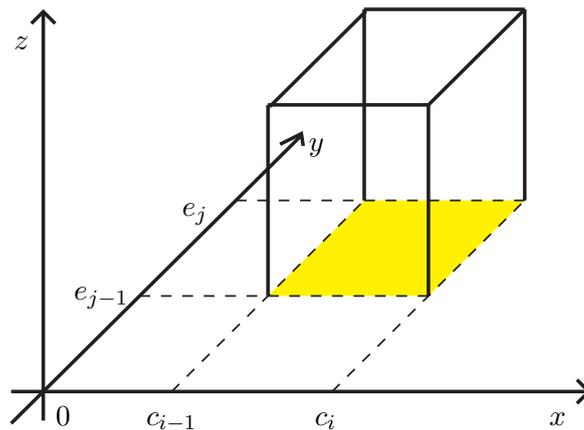


Abb. 2.12: Zweidimensionales Histogramm

3. *Zweidimensionale Kerndichteschätzer* haben die Form

$$\hat{f}(x, y) = \frac{1}{nh_1h_2} \sum_{i=1}^n K\left(\frac{x-x_i}{h_1}\right) K\left(\frac{y-y_i}{h_2}\right)$$

für die Bandbreiten $h_1, h_2 > 0$, die Glättungsparameter sind. Dabei ist $K(\cdot)$ eine Kernfunktion (vgl. Abschnitt 2.4). Seine Eigenschaften übertragen sich aus dem eindimensionalen Fall.

2.5.2 Zusammenhangsmaße

Jetzt wird uns die Frage beschäftigen, in welchem Maße die Merkmale X und Y voneinander abhängig sind. Um die $\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$ aus den Daten zu schätzen, setzt man die sogenannte *empirische Kovarianz*

$$S_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)$$

ein. Dabei ist S_{xy}^2 jedoch von den Skalen von X und Y abhängig.

1. Um ein skaleninvariantes Zusammenhangsmaß zu bekommen, betrachtet man die empirische Variante des Korrelationskoeffizienten

$$\varrho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \cdot \sqrt{\text{Var } Y}},$$

den sogenannten *Bravais-Pearson-Korrelationskoeffizienten*

$$\varrho_{xy} = \frac{S_{xy}^2}{\sqrt{S_{xx}^2 \cdot S_{yy}^2}},$$

wobei

$$S_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2, \quad S_{yy}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_n)^2$$

die Stichprobenvarianzen der Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) sind. Dabei erbt ϱ_{xy} alle Eigenschaften des Korrelationskoeffizienten $\varrho(X, Y)$:

- $|\varrho_{xy}| \leq 1$
- $\varrho_{xy} = \pm 1$, falls ein linearer Zusammenhang in den Daten $(x_i, y_i)_{i=1, \dots, n}$ vorliegt, d.h. alle Punkte (x_i, y_i) , $i = 1, \dots, n$ liegen auf einer Gerade mit positivem (bei $\varrho_{xy} = 1$) bzw. negativem (bei $\varrho_{xy} = -1$) Anstieg.
- Wenn $|\varrho_{xy}|$ klein ist ($\varrho_{xy} \approx 0$), so sind die Datensätze unkorreliert. Dabei wird oft folgende grobe Einteilung vorgenommen:

Merkmale X und Y sind

- „*schwach korreliert*“, falls $|\varrho_{xy}| < 0.5$,
- „*stark korreliert*“, falls $|\varrho_{xy}| \geq 0.8$.

Ansonsten liegt ein mittlerer Zusammenhang zwischen X und Y vor.

Lemma 2.5.1

Für ϱ_{xy} gilt die alternative rechengünstige Darstellung

$$\varrho_{xy} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}_n^2) (\sum_{i=1}^n y_i^2 - n \bar{y}_n^2)}}. \quad (2.5.1)$$

Beweis Man muss lediglich zeigen, dass

$$\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n.$$

Alles andere folgt daraus für $x_i = y_i$, $i = 1, \dots, n$. Es gilt

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) &= \sum_{i=1}^n x_i y_i - \bar{x}_n \sum_{i=1}^n y_i - \bar{y}_n \sum_{i=1}^n x_i + n \bar{x}_n \bar{y}_n \\ &= \sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n - n \bar{y}_n \bar{x}_n + n \bar{x}_n \bar{y}_n = \sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n \end{aligned}$$

□

Falls die vorliegenden Daten (x_1, \dots, x_n) und (y_1, \dots, y_n) nur 2 Ausprägungen zeigen und somit *binär* kodiert werden können, d.h. $x_i, y_i \in \{0, 1\}$, dann gilt

$$\varrho_{xy} = \frac{h_{00}h_{11} - h_{01}h_{10}}{\sqrt{h_{0\cdot} \cdot h_{1\cdot} \cdot h_{\cdot 0} \cdot h_{\cdot 1}}} = \varphi$$

(der sogenannte Phi-Koeffizient), wobei

$$\begin{aligned} h_{00} &= \#\{(x_i, y_i) : x_i = y_i = 0\} \\ h_{11} &= \#\{(x_i, y_i) : x_i = y_i = 1\} \\ h_{01} &= \#\{(x_i, y_i) : x_i = 0, y_i = 1\} \\ h_{10} &= \#\{(x_i, y_i) : x_i = 1, y_i = 0\} \\ h_{0.} &= h_{11} + h_{12} \\ h_{.0} &= h_{11} + h_{21} \\ h_{1.} &= h_{22} + h_{21} \\ h_{.1} &= h_{22} + h_{12} \end{aligned}$$

Übungsaufgabe 2.5.1

Zeigen Sie diese Darstellungsform!

2. Spearmans Korrelationskoeffizient

Einen alternativen Korrelationskoeffizienten erhält man, wenn man die Stichprobenwerte x_i bzw. y_i in ϱ_{xy} durch ihre *Ränge* $\text{rg}(x_i)$ bzw. $\text{rg}(y_i)$ ersetzt, die als Position dieser Werte in den ansteigend geordneten Stichproben zu verstehen sind:

$\text{rg}(x_i) = j$, falls $x_i = x_{(j)}$ für ein $j \in \{1, \dots, n\}$, $\forall i = 1, \dots, n$. Es bedeutet, dass $\text{rg}(x_{(i)}) = i \forall i = 1, \dots, n$, falls $x_i \neq x_j$ für $i \neq j$.

Falls die Stichprobe (x_1, \dots, x_n) k identische Werte x_i (die sogenannten *Bindungen*) enthält, so wird diesen Werten der sogenannte Durchschnittsrang $\text{rg}(x_i)$ zugewiesen, der als arithmetisches Mittel der k in Frage kommenden Ränge errechnet wird. Zum Beispiel findet folgende Zuordnung statt:

$$\frac{x_i \mid (3, 1, 7, 5, 3, 3)}{\text{rg}(x_i) \mid (a, 1, 6, 5, a, a)}$$

wobei der Durchschnittsrang a von Stichprobeneintrag 3 gleich $a = 1/3(2 + 3 + 4) = 3$ ist.

Somit wird der sogenannte *Spearmans Korrelationskoeffizient* (Rangkorrelationskoeffizient) der Stichproben

$$(x_1, \dots, x_n) \quad \text{und} \quad (y_1, \dots, y_n)$$

als der *Bravais-Pearson-Koeffizient* der Stichproben ihrer Ränge

$$(\text{rg}(x_1), \dots, \text{rg}(x_n)) \quad \text{und} \quad (\text{rg}(y_1), \dots, \text{rg}(y_n))$$

definiert:

$$\varrho_{sp} = \frac{\sum_{i=1}^n (\text{rg}(x_i) - \overline{\text{rg}}_x)(\text{rg}(y_i) - \overline{\text{rg}}_y)}{\sqrt{\sum_{i=1}^n (\text{rg}(x_i) - \overline{\text{rg}}_x)^2 \sum_{i=1}^n (\text{rg}(y_i) - \overline{\text{rg}}_y)^2}},$$

wobei

$$\begin{aligned} \overline{\text{rg}}_x &= \frac{1}{n} \sum_{i=1}^n \text{rg}(x_i) = \frac{1}{n} \sum_{i=1}^n \text{rg}(x_{(i)}) = \frac{1}{n} \sum_{i=1}^n i = \frac{n(n+1)}{2n} = \frac{n+1}{2}, \\ \overline{\text{rg}}_y &= \frac{1}{n} \sum_{i=1}^n \text{rg}(y_i) = \frac{n+1}{2}. \end{aligned}$$

Dieselbe Darstellung $\overline{\text{rg}}_y$ gilt auch, wenn Bindungen vorhanden sind.

Dieser Koeffizient misst monotone Zusammenhänge in den Daten. Aus den Eigenschaften der Bravais-Pearson-Koeffizienten folgt $|\varrho_{sp}| \leq 1$. Betrachten wir die Fälle $\varrho_{sp} = \pm 1$ gesondert:

- $\varrho_{sp} = 1$ bedeutet, dass die Punkte $(\text{rg}(x_i), \text{rg}(y_i))$, $i = 1, \dots, n$ auf einer Geraden mit positiver Steigung liegen. Da aber $\text{rg}(x_i), \text{rg}(y_i) \in \mathbb{N}$, kann diese Steigung nur 1 sein. Es bedeutet, dass dem kleinsten Wert in der Stichprobe (x_1, \dots, x_n) der kleinste Wert in (y_1, \dots, y_n) entspricht, usw., d.h., für wachsende x_i wachsen auch die y_i streng monoton: $x_i < x_j \implies y_i < y_j \quad \forall i \neq j$.
- Analog gilt dann für $\varrho_{sp} = -1$, dass $x_i < x_j \implies y_i > y_j \quad \forall i \neq j$.

Dies kann folgendermaßen zusammengefaßt werden:

- $\varrho_{sp} > 0$: gleichsinniger monotoner Zusammenhang (x_i groß \iff y_i groß)
- $\varrho_{sp} < 0$: gegensinniger monotoner Zusammenhang (x_i groß \iff y_i klein)
- $\varrho_{sp} \approx 0$: kein monotoner Zusammenhang.

Da der Spearmans Korrelationskoeffizient nur Ränge von x_i und y_i betrachtet, eignet er sich auch für ordinale (und nicht nur quantitative) Daten.

Lemma 2.5.2

Falls die Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) keine Bindung enthalten ($x_i \neq x_j, y_i \neq y_j \quad \forall i \neq j$), dann gilt

$$\varrho_{sp} = 1 - \frac{6}{(n^2 - 1)n} \sum_{i=1}^n d_i^2,$$

wobei $d_i = \text{rg}(x_i) - \text{rg}(y_i) \quad \forall i = 1, \dots, n$.

Beweis Als Übungsaufgabe. □

Satz 2.5.1 (Invarianzeigenschaften):

1. Wenn die Merkmale X und Y linear transformiert werden:

$$\begin{aligned} f(X) &= a_x X + b_x, & a_x &\neq 0, b_x \in \mathbb{R}, \\ g(Y) &= a_y Y + b_y, & a_y &\neq 0, b_y \in \mathbb{R}, \end{aligned}$$

dann gilt $\varrho_{f(x)g(y)} = \text{sgn}(a_x a_y) \cdot \varrho_{xy}$.

2. Falls Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ und $g: \mathbb{R} \rightarrow \mathbb{R}$ beide monoton wachsend oder beide monoton fallend sind, dann gilt

$$\varrho_{sp}(f(x), g(y)) = \varrho_{sp}(x, y).$$

Falls f monoton wachsend und g monoton fallend (oder umgekehrt) sind, dann gilt $\varrho_{sp}(f(x), g(y)) = -\varrho_{sp}(x, y)$.

Beweis Beweisen wir nur 1), weil 2) offensichtlich ist.

1.

$$\begin{aligned} \varrho_{f(x)g(y)} &= \frac{\sum_{i=1}^n ((a_x x_i + b_x) - (a_x \bar{x}_n + b_x))((a_y y_i + b_y) - (a_y \bar{y}_n + b_y))}{\sqrt{a_x^2 \sum_{i=1}^n (x_i - \bar{x}_n)^2 a_y^2 \sum_{i=1}^n (y_i - \bar{y}_n)^2}} \\ &= \frac{a_x a_y}{|a_x| |a_y|} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2 \sum_{i=1}^n (y_i - \bar{y}_n)^2}} = \operatorname{sgn}(a_x a_y) \cdot \varrho_{xy}. \end{aligned}$$

□

Bemerkung 2.5.1

1. Da lineare Transformationen monoton sind, gilt Aussage 1) auch für Spearmans Korrelationskoeffizienten ϱ_{sp} .
2. Der Koeffizient ϱ_{xy} erfasst lineare Zusammenhänge, während ϱ_{sp} monotone Zusammenhänge aufspürt.

2.5.3 Einfache lineare Regression

Wenn man den Zusammenhang von Merkmalen X und Y mit Hilfe von Streudiagrammen visualisiert, wird oft ein linearer Trend erkennbar, obwohl der Bravais-Pearson-Korrelationskoeffizient einen Wert kleiner als 1 liefert, z.B. $\varrho_{xy} \approx 0,6$ (vgl. Abb. 2.13). Dies ist der Fall, weil die Da-

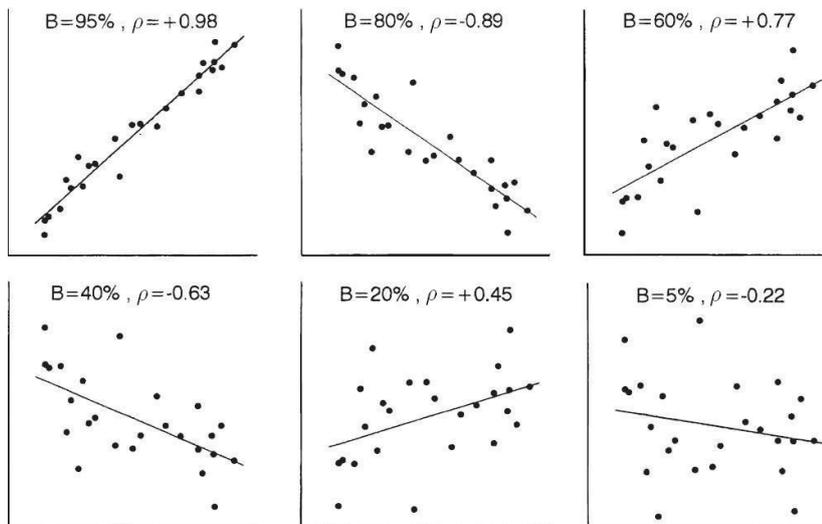


Abb. 2.13: Vergleich verschiedenwertiger Bestimmtheitsmaße. Es sind Regressionsgerade, Bestimmtheitsmaß B und Korrelationskoeffizient ρ verschiedener (fiktiver) Punktwolken vom Umfang $n = 25$ dargestellt. Die Beschriftung der Achsen ist weggelassen, weil sie hier ohne Bedeutung ist.

tenpunkte (x_i, y_i) , $i = 1, \dots, n$ oft um eine Gerade streuen und nicht exakt auf einer Geraden

X	Y
Geschwindigkeit	Länge des Bremswegs
Körpergröße des Vaters	Körpergröße des Sohnes
Produktionsfaktor	Qualität des Produktes
Spraydosen-Verbrauch	Ozongehalt der Atmosphäre
Noten im Bachelor-Studium	Noten im Master-Studium

Tab. 2.1: Beispiele möglicher Ausgangs- und Zielgrößen

liegen. Um solche Situationen stochastisch modellieren zu können, nimmt man den Zusammenhang der Form

$$Y = f(X) + \varepsilon$$

an, wobei ε die sogenannte Störgröße ist, die auf mehrere Ursachen wie z.B. Beobachtungsfehler (Messfehler, Berechnungsfehler, usw.) zurückzuführen sein kann. Dabei nennt man die Zufallsvariable Y *Zielgröße* oder *Regressand*, die Zufallsvariable X *Einflussfaktor*, *Regressor* oder *Ausgangsvariable*. Der Zusammenhang $Y = f(X) + \varepsilon$ wird *Regression* genannt, wobei man oft über ε voraussetzt, dass $\mathbb{E}\varepsilon = 0$ (kein systematischer Beobachtungsfehler). Wenn $f(x) = \alpha + \beta x$ eine lineare Funktion ist, so spricht man von der *einfachen linearen Regression*. Es sind aber durchaus andere Arten der Zusammenhänge denkbar, wie z.B.

$$f(x) = \sum_{i=0}^n \alpha_i x^i$$

(*polynomiale Regression*), usw. Beispiele für mögliche Ausgangs- bzw. Zielgrößen sind in Tabelle 2.1 zusammengefasst, einige Beispiele in Abbildung 2.14.

Auf Modellebene ist damit folgende Fragestellung gegeben: Es gebe Zufallsstichproben von Ziel- bzw. Ausgangsvariablen (Y_1, \dots, Y_n) und (X_1, \dots, X_n) , zwischen denen ein verrauschter linearer Zusammenhang $Y_i = \alpha + \beta X_i + \varepsilon_i$ besteht, wobei ε_i Störgrößen sind, die nicht direkt beobachtbar und uns somit unbekannt sind. Meistens nimmt man an, dass $\mathbb{E}\varepsilon_i = 0 \quad \forall i = 1, \dots, n$ und $\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{ij}$, d.h. $\varepsilon_1 \dots \varepsilon_n$ sind unkorreliert mit $\text{Var}\varepsilon_i = \sigma^2$. Wenn wir über die Eigenschaften der Schätzer für α , β und σ^2 reden, gehen wir davon aus, dass die X -Werte nicht zufällig sind, also $X_i = x_i \quad \forall i = 1, \dots, n$. Wenn man von einer konkreten Stichprobe (y_1, \dots, y_n) für (Y_1, \dots, Y_n) ausgeht, so sollen anhand von den Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) Regressionsparameter α (*Regressionskonstante*) und β (*Regressionskoeffizient*) sowie *Regressionsvarianz* σ^2 geschätzt werden. Dabei verwendet man die sogenannte *Methode der kleinsten Quadrate*, die den mittleren quadratischen Fehler von den Datenpunkten $(x_i, y_i)_{i=1, \dots, n}$ des Streudiagramms zur *Regressionsgeraden* $y = \alpha + \beta x$ minimiert:

$$(\alpha, \beta) = \arg \min_{\alpha, \beta \in \mathbb{R}} e(\alpha, \beta) \quad \text{mit} \quad e(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

Da die Darstellung $y_i = \alpha + \beta x_i + \varepsilon_i$ gilt, kann man $e(\alpha, \beta) = 1/n \sum_{i=1}^n \varepsilon_i^2$ schreiben. Es ist der vertikale mittlere quadratische Abstand von den Datenpunkten (x_i, y_i) zur Geraden $y = \alpha + \beta x$ (vgl. Abb. 2.15). Das Minimierungsproblem $e(\alpha, \beta) \mapsto \min$ löst man durch das

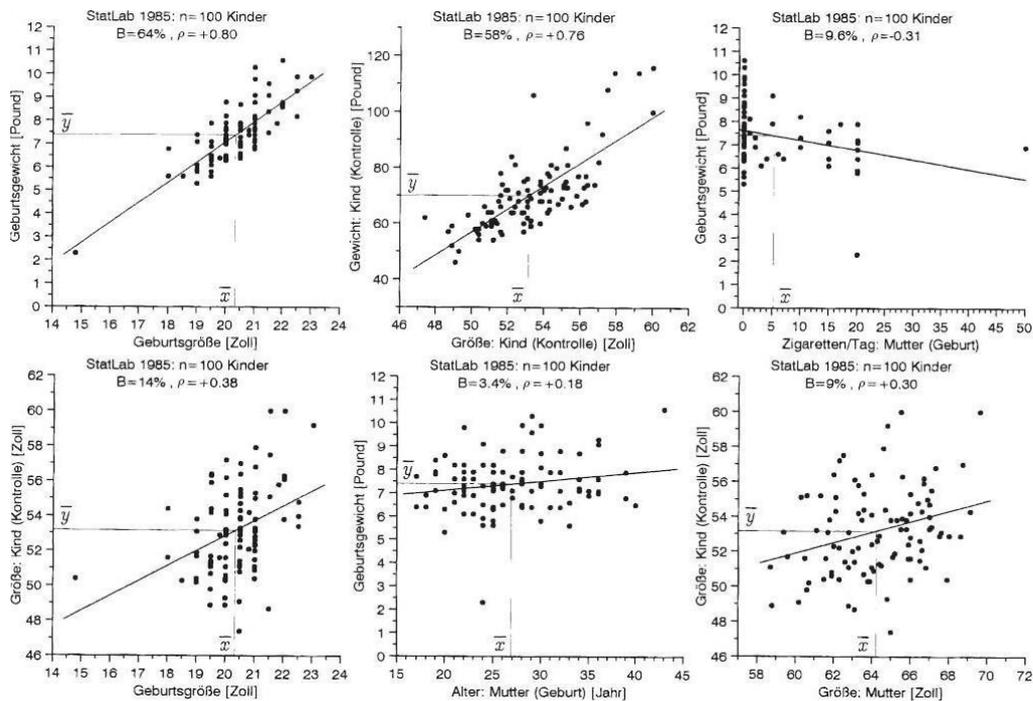


Abb. 2.14: Punktwolken verschiedener Merkmale der StatLab-Auswahl 1985 mit Regressionsgerade, Bestimmtheitsmaß B und Korrelationskoeffizient ρ .

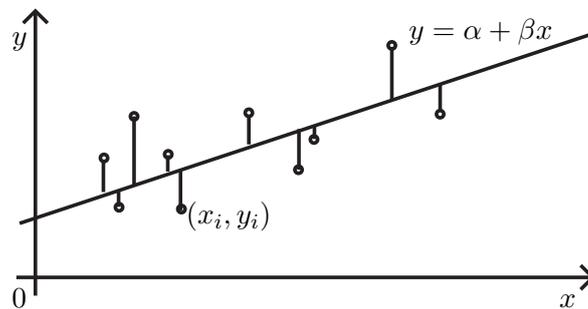


Abb. 2.15: Methode kleinster Quadrate

zweifache Differenzieren von $e(\alpha, \beta)$. Somit erhält man $\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n$, wobei

$$\hat{\beta} = \frac{S_{xy}^2}{S_{xx}^2}, \quad \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$S_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n), \quad S_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Übungsaufgabe 2.5.2

Leiten Sie die Schätzer $\hat{\alpha}$ und $\hat{\beta}$ selbstständig her.

Die Varianz σ^2 schätzt man durch $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2$, wobei $\hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta}x_i$, $i = 1, \dots, n$

Kind i	1	2	3	4	5	6	7	8	9
Fernsehzeit x_i	0,3	2,2	0,5	0,7	1,0	1,8	3,0	0,2	2,3
Tiefschlafdauer y_i	5,8	4,4	6,5	5,8	5,6	5,0	4,8	6,0	6,1

Tab. 2.2: Daten von Fernsehzeit und korrespondierender Tiefschlafdauer

die sogenannten *Residuen* sind. Die Gründe, warum $\hat{\sigma}^2$ diese Gestalt hat, können an dieser Stelle noch nicht angegeben werden, weil wir noch nicht die Maximum-Likelihood-Methode kennen. Zu gegebener Zeit (in der Vorlesung Stochastik III) wird jedoch klar, dass diese Art der Schätzung sehr natürlich ist.

Bemerkung 2.5.2

Die angegebenen Schätzer für α und β sind nicht symmetrisch bzgl. Variablen x_i und y_i . Wenn man also die *horizontalen* Abstände (statt vertikaler) zur Bildung des mittleren quadratischen Fehlers nimmt (was dem Rollentausch $x \leftrightarrow y$ entspricht), so bekommt man andere Schätzer für α und β , die mit $\hat{\alpha}$ und $\hat{\beta}$ nicht übereinstimmen müssen:

$$d_i = y_i - \alpha - \beta x_i \mapsto d'_i = x_i - \frac{(y_i - \alpha)}{\beta}.$$

Ein Ausweg aus dieser asymmetrischen Situation wäre es, die orthogonalen Abstände o_i von

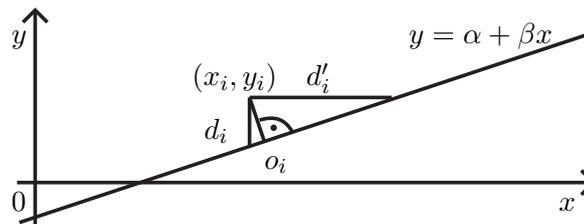


Abb. 2.16: Orthogonale Abstände

(x_i, y_i) zur Geraden $y = \alpha + \beta x$ zu betrachten (vgl. Abb. 2.16). Diese Art der Regression, die „errors-in-variables regression“ genannt wird, hat aber eine Reihe von Eigenschaften, die sie zur Prognose von Zielvariablen y_i durch die Ausgangsvariablen x_i unbrauchbar machen. Sie sollte zum Beispiel nur dann verwendet werden, wenn die Standardabweichungen für X und Y etwa gleich groß sind.

Beispiel 2.5.1

Ein Kinderpsychologe vermutet, dass sich häufiges Fernsehen negativ auf das Schlafverhalten von Kindern auswirkt. Um diese Hypothese zu überprüfen, wurden 9 Kinder im gleichen Alter befragt, wie lange sie pro Tag fernsehen dürfen, und zusätzlich die Dauer ihrer Tiefschlafphase gemessen. So ergibt sich der Datensatz in Tabelle 2.2 und die Regressionsgerade aus Abbildung 2.17.

Es ergibt sich für die oben genannten Stichproben (x_1, \dots, x_9) und (y_1, \dots, y_9)

$$\bar{x}_9 = 1,33, \quad \bar{y}_9 = 5,56, \quad \hat{\beta} = -0,45, \quad \hat{\alpha} = 6,16.$$

Somit ist

$$y = 6,16 - 0,45x$$

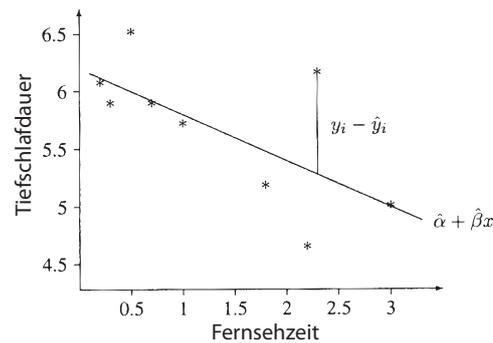


Abb. 2.17: Streudiagramm und Ausgleichsgerade zur Regression der Dauer des Tiefschlafs auf die Fernsehzeit

die Regressionsgerade, die eine negative Steigung hat, was die Vermutung des Kinderpsychologen bestätigt. Außerdem ist es mit Hilfe dieser Geraden möglich, Prognosen für die Dauer des Tiefschlafs für vorgegebene Fernsehzeiten anzugeben. So wäre z.B. für die Fernsehzeit von 1 Stunde der Tiefschlaf von $6,16 - 0,45 \cdot 1 = 5,71$ Stunden plausibel.

Bemerkung 2.5.3 (Eigenschaften der Regressionsgerade):

1. Es gilt $\text{sgn}(\hat{\beta}) = \text{sgn}(\rho_{xy})$, was aus $\hat{\beta} = s_{xy}^2 / s_{xx}^2$ folgt. Dies bedeutet (falls $s_{yy}^2 > 0$):
 - a) Die Regressionsgerade $y = \hat{\alpha} + \hat{\beta}x$ steigt an, falls die Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) positiv korreliert sind.
 - b) Die Regressionsgerade fällt ab, falls sie negativ korreliert sind.
 - c) Die Regressionsgerade ist konstant, falls die Stichproben unkorreliert sind.

Falls $s_{yy}^2 = 0$, dann ist die Regressionsgerade konstant ($y = \bar{y}_n$).

2. Die Regressionsgerade $y = \hat{\alpha} + \hat{\beta}x$ verläuft immer durch den Punkt (\bar{x}_n, \bar{y}_n) : $\hat{\alpha} + \hat{\beta}\bar{x}_n = \bar{y}_n$.
3. Seien $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$, $i = 1, \dots, n$. Dann gilt

$$\bar{\hat{y}}_n = \frac{1}{n} \sum_{i=1}^n \hat{y}_i = \bar{y}_n \quad \text{und somit} \quad \sum_{i=1}^n \underbrace{(y_i - \hat{y}_i)}_{\hat{\varepsilon}_i} = 0.$$

Dabei sind $\hat{\varepsilon}_i$ die schon vorher eingeführten Residuen. Mit ihrer Hilfe ist es möglich, die Güte der Regressionsprognose zu beurteilen.

Residualanalyse und Bestimmtheitsmaß

Definition 2.5.1

Der relative Anteil der Streuungsreduktion an der Gesamtstreuung S_{yy}^2 heißt das *Bestimmtheitsmaß* der Regressionsgeraden:

$$R^2 = \frac{S_{yy}^2 - \frac{1}{n-1} \sum_{i=1}^n \hat{\varepsilon}_i^2}{S_{yy}^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2}.$$

Es ist nur im Fall $S_{xx}^2 > 0$, $S_{yy}^2 > 0$ definiert, d.h., wenn nicht alle Werte x_i bzw. y_i übereinstimmen.

Warum R^2 in dieser Form eingeführt wird, zeigt folgende Überlegung, die *Streuungszerlegung* genannt wird:

Lemma 2.5.3

Die Gesamtstreuung („sum of squares total“) $\text{SQT} = (n-1)S_{yy}^2 = \sum_{i=1}^n (y_i - \bar{y}_n)^2$ lässt sich in die Summe der sogenannten erklärten Streuung „sum of squares explained“ $\text{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2$ und der Residualstreuung „sum of squared residuals“ $\text{SQR} = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ zerlegen:

$$\text{SQT} = \text{SQE} + \text{SQR}$$

bzw.

$$\sum_{i=1}^n (y_i - \bar{y}_n)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Beweis

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^n (y_i - \bar{y}_n)^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y}_n)^2 \\ &= \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{=\text{SQR}} + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}_n) + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2}_{=\text{SQE}} \\ &= \text{SQE} + \text{SQR} + 2 \sum_{i=1}^n \hat{y}_i (y_i - \hat{y}_i) - 2\bar{y}_n \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)}_{=0, \text{vgl. Fig. 3 S. 37}} = \text{SQE} + \text{SQR} + E, \end{aligned}$$

wobei noch zu zeigen ist, dass $E = 2 \sum_{i=1}^n \hat{y}_i (y_i - \hat{y}_i) = 0$, also

$$\begin{aligned} E &= 2 \sum_{i=1}^n (\hat{\alpha} + \hat{\beta}x_i)(y_i - \hat{\alpha} - \hat{\beta}x_i) = 2\hat{\alpha} \underbrace{\sum_{i=1}^n \hat{\varepsilon}_i}_{=0} + 2\hat{\beta} \sum_{i=1}^n x_i (y_i - \hat{\alpha} - \hat{\beta}x_i) \\ &= 2\hat{\beta} \left(\sum_{i=1}^n x_i y_i - \hat{\alpha} \sum_{i=1}^n x_i - \hat{\beta} \sum_{i=1}^n x_i^2 \right) \stackrel{\hat{\alpha} = \bar{y}_n - \bar{x}_n \hat{\beta}}{=} 2\hat{\beta} \left(\underbrace{\sum_{i=1}^n x_i y_i - n\bar{x}_n \bar{y}_n + \hat{\beta} n \bar{x}_n^2}_{=(n-1)S_{xy}^2} - \hat{\beta} \sum_{i=1}^n x_i^2 \right) \\ &= 2\hat{\beta} \left((n-1)S_{xy}^2 - \hat{\beta}(n-1)S_{xx}^2 \right) \stackrel{\hat{\beta} = \frac{S_{xy}^2}{S_{xx}^2}}{=} 2\hat{\beta}(n-1) \left(S_{xy}^2 - \frac{S_{xy}^2}{S_{xx}^2} \cdot S_{xx}^2 \right) = 0. \end{aligned}$$

□

Die erklärte Streuung gibt die Streuung der Regressionsgeradenwerte um \bar{y}_n an. Sie stellt damit die auf den linearen Zusammenhang zwischen X und Y zurückzuführende Variation der y -Werte dar. Das oben eingeführte Bestimmtheitsmaß ist somit der Anteil dieser Streuung an der Gesamtstreuung:

$$R^2 = \frac{\text{SQE}}{\text{SQT}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2} = \frac{\text{SQT} - \text{SQR}}{\text{SQT}} = 1 - \frac{\text{SQR}}{\text{SQT}}.$$

Es folgt aus dieser Darstellung, dass $R^2 \in [0, 1]$ ist.

1. $R^2 = 0$ bedeutet $\text{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 = 0$ und somit $\hat{y}_i = \bar{y}_n \forall i$. Dies weist darauf hin, dass das lineare Modell in diesem Fall schlecht ist, denn aus $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i = \bar{y}_n$ folgt $\hat{\beta} = \frac{S_{xy}^2}{S_{xx}^2} = 0$ und somit $S_{xy}^2 = 0$. Also sind die Merkmale X und Y unkorreliert.
2. $R^2 = 1$ bedingt $\text{SQR} = \sum_{i=1}^n \hat{\varepsilon}_i^2 = 0$. Somit liegen alle (x_i, y_i) perfekt auf der Regressionsgeraden. Dies bedeutet, dass die Daten x_i und y_i , $i = 1, \dots, n$ perfekt linear abhängig sind.

Faustregel zur Beurteilung der Güte der Anpassung eines linearen Modells an Hand von Bestimmtheitsmaß R^2 :

R^2 ist deutlich von Null verschieden (d.h. es besteht noch ein linearer Zusammenhang), falls $R^2 > \frac{4}{n+2}$, wobei n der Stichprobenumfang ist.

Allgemein gilt folgender Zusammenhang zwischen dem Bestimmtheitsmaß R^2 und dem Bravais-Pearson-Korrelationskoeffizienten ϱ_{xy} :

Lemma 2.5.4

$$R^2 = \varrho_{xy}^2$$

Beweis Aus der Eigenschaft 3 S. 37 folgt $\bar{y}_n = \overline{\hat{y}_n}$. Somit gilt

$$\text{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 = \sum_{i=1}^n (\hat{y}_i - \overline{\hat{y}_n})^2 = \sum_{i=1}^n (\hat{\alpha} + \hat{\beta}x_i - \hat{\alpha} - \hat{\beta}\bar{x}_n)^2 = \hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

und damit

$$R^2 = \frac{\text{SQE}}{\text{SQT}} = \frac{\hat{\beta}^2 \sum_{i=1}^n (x_i - \bar{x}_n)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2} = \frac{(S_{xy}^2)^2}{(S_{xx}^2)^2} \cdot \frac{(n-1)S_{xx}^2}{(n-1)S_{yy}^2} = \left(\frac{S_{xy}^2}{S_{yy}S_{xx}} \right)^2 = \varrho_{xy}^2$$

□

Folgerung 2.5.1

1. Der Wert von R^2 ändert sich bei einer Lineartransformation der Daten (x_1, \dots, x_n) und (y_1, \dots, y_n) nicht. Grafisch kann man die Güte der Modellanpassung bei der linearen Regression folgendermaßen überprüfen:

Man zeichnet Punktpaare $(\hat{y}_i, \hat{\varepsilon}_i)_{i=1, \dots, n}$ als Streudiagramm (der sogenannte *Residualplot*). Falls diese Punktwolke gleichmäßig um Null streut, so ist das lineare Modell gut gewählt worden. Falls das Streudiagramm einen erkennbaren Trend aufweist, bedeutet das, dass die Annahme des linearen Modells für diese Daten ungeeignet sei (vgl. Abb. 2.18)

2. Da $R^2 = \varrho_{xy}^2$, ist der Wert von R^2 symmetrisch bzgl. der Stichproben (x_1, \dots, x_n) und (y_1, \dots, y_n) :

$$\varrho_{xy}^2 = R^2 = \varrho_{yx}^2 \quad \text{bzw.} \quad R_{xy}^2 = R_{yx}^2,$$

wobei R_{xy}^2 das Bestimmtheitsmaß bezeichnet, das sich aus der normalen Regression ergibt und R_{yx}^2 das mit vertauschten Achsen.

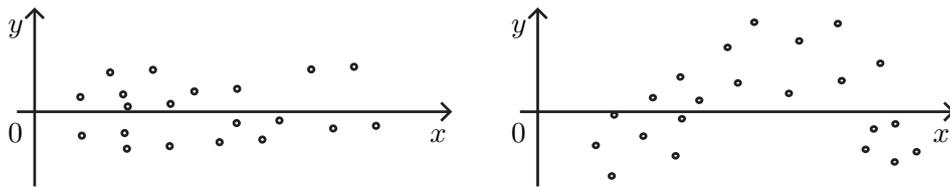


Abb. 2.18: Links: Gute, Rechts: Schlechte Übereinstimmung mit dem linearen Modell

3 Punktschätzer

3.1 Parametrisches Modell

Sei (x_1, \dots, x_n) eine konkrete Stichprobe. Es wird angenommen, dass (x_1, \dots, x_n) eine Realisierung einer Zufallsstichprobe (X_1, \dots, X_n) ist, wobei X_1, \dots, X_n unabhängige identisch verteilte Zufallsvariablen mit der unbekanntem Verteilungsfunktion F sind und F zu einer bekannten parametrischen Familie $\{F_\theta : \theta \in \Theta\}$ gehört. Hier ist $\theta = (\theta_1, \dots, \theta_m) \in \Theta$ der *m-dimensionale Parametervektor* der Verteilung F_θ und $\Theta \subset \mathbb{R}^m$ der sogenannte *Parameterraum* (eine Borel-Teilmenge von \mathbb{R}^m , die die Menge aller zugelassenen Parameterwerte darstellt). Es wird vorausgesetzt, dass die Parametrisierung $\theta \rightarrow F_\theta$ *identifizierbar* ist, indem $F_{\theta_1} \neq F_{\theta_2}$ für $\theta_1 \neq \theta_2$ gilt.

Eine wichtige Aufgabe der Statistik, die wir in diesem Kapitel betrachten werden, besteht in der Schätzung des Parametervektors θ (oder eines Teils von θ) an Hand von der konkreten Stichprobe (x_1, \dots, x_n) . In diesem Fall spricht man von einem *Punktschätzer* $\hat{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, der eine gültige Stichprobenfunktion ist. Meistens wird angenommen, dass

$$\mathbb{P}(\hat{\theta}(X_1, \dots, X_n) \in \Theta) = 1,$$

wobei es zu dieser Regel auch Ausnahmen gibt. Bisher haben wir den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, auf dem unsere Zufallsstichprobe definiert ist, nicht näher spezifiziert. Dies kann man aber leicht tun, indem man den sogenannten *kanonischen Wahrscheinlichkeitsraum* angibt, wobei

$$\Omega = \mathbb{R}^\infty, \quad \mathcal{F} = \mathcal{B}_\mathbb{R}^\infty = \mathcal{B}_\mathbb{R} \times \mathcal{B}_\mathbb{R} \times \dots$$

und das Wahrscheinlichkeitsmaß \mathbb{P} durch

$$\mathbb{P}(\{\omega = (\omega_1, \dots, \omega_n, \dots) \in \mathbb{R}^\infty : \omega_{i_1} \leq x_{i_1}, \dots, \omega_{i_k} \leq x_{i_k}\}) = F_\theta(x_{i_1}) \dots F_\theta(x_{i_k})$$

$\forall k \in \mathbb{N}, 1 \leq i_1 < \dots < i_k$ gegeben sei. Um zu betonen, dass \mathbb{P} vom Parameter θ abhängt, werden wir Bezeichnungen $\mathbb{P}_\theta, \mathbb{E}_\theta$ und Var_θ für das Maß \mathbb{P} , den Erwartungswert und die Varianz bzgl. \mathbb{P} verwenden.

Auf dem kanonischen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ gilt $X_i(\omega) = \omega_i$ (Projektion auf die Koordinate i), $i = 1, \dots, n$,

$$\mathbb{P}_\theta(X_i \leq x_i) = \mathbb{P}_\theta(\{\omega \in \Omega : \omega_i \leq x_i\}) = F_\theta(x_i), \quad i = 1, \dots, n, \quad x_i \in \mathbb{R}.$$

Beispiel 3.1.1

1. Sei X die Dauer des fehlerfreien Arbeitszyklus eines technischen Systems. Oft wird $X \sim \text{Exp}(\lambda)$ angenommen. Dann stellt $\{F_\theta : \theta \in \Theta\}$ mit $m = 1, \theta = \lambda, \Theta = \mathbb{R}_+$ und

$$F_\theta(x) = (1 - e^{-\theta x}) \cdot \mathbb{I}(x \geq 0)$$

ein parametrisches Modell dar, wobei der Parameterraum eindimensional ist. Später wird für λ der (Punkt-) Schätzer $\hat{x}(x_1, \dots, x_n) = 1/\bar{x}_n$ vorgeschlagen.

2. In den Fragestellungen der statistischen Qualitätskontrolle werden n Erzeugnisse auf Mängel untersucht. Falls $p \in (0, 1)$ die unbekannte Wahrscheinlichkeit des Mangels ist, so wird mit $X \sim \text{Bin}(n, p)$ die Gesamtanzahl der mangelhaften Produkte beschrieben. Dabei wird folgendes parametrische Modell unterstellt:

$$\Theta = \{(n, p) : n \in \mathbb{N}, p \in (0, 1)\}, \quad \theta = (n, p), \quad m = 2,$$

$$F_{\theta}(x) = \mathbb{P}_{\theta}(X \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \cdot \mathbb{I}(x \geq 0).$$

Falls n bekannt ist, kann die Wahrscheinlichkeit p des Ausschusses durch den Punktschätzer $\hat{p}(x_1, \dots, x_n) = \bar{x}_n$, $x_i \in \{0, 1\}$ näherungsweise berechnet werden.

3.2 Parametrische Familien von statistischen Prüfverteilungen

In der Vorlesung Wahrscheinlichkeitsrechnung wurden bereits einige parametrische Familien von Verteilungen eingeführt. Hier geben wir weitere Verteilungsfamilien an, die in der Statistik eine besondere Stellung einnehmen, weil sie als Referenzverteilungen in der Schätztheorie, statistischen Tests und Vertrauensintervallen ihre Anwendung finden.

3.2.1 Gamma-Verteilung

Als erstes führen wir zwei spezielle Funktionen aus der Analysis ein:

1. Die *Gamma-Funktion*:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx \quad \text{für } p > 0.$$

Es gelten folgende Eigenschaften:

$$\begin{aligned} \Gamma(1) &= 1, & \Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(p+1) &= p\Gamma(p) \quad \forall p > 0, & \Gamma(n+1) &= n!, \quad \forall n \in \mathbb{N}. \end{aligned}$$

2. Die *Beta-Funktion*:

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt, \quad p, q > 0.$$

Es gelten folgende Eigenschaften:

$$B(p, q) = B(q, p), \quad B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \quad p, q > 0.$$

Definition 3.2.1

Die *Gamma-Verteilung* mit Parametern $\lambda > 0$ und $p > 0$ ist eine absolut stetige Verteilung mit der Dichte

$$f_X(x) = \begin{cases} \frac{\lambda^p x^{p-1}}{\Gamma(p)} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (3.2.1)$$

Dabei verwenden wir die Bezeichnung $X \sim \Gamma(\lambda, p)$ für eine Zufallsvariable X , die Gammaverteilt mit Parametern λ und p ist. Es gilt offensichtlich $X \geq 0$ fast sicher für $X \sim \Gamma(\lambda, p)$.

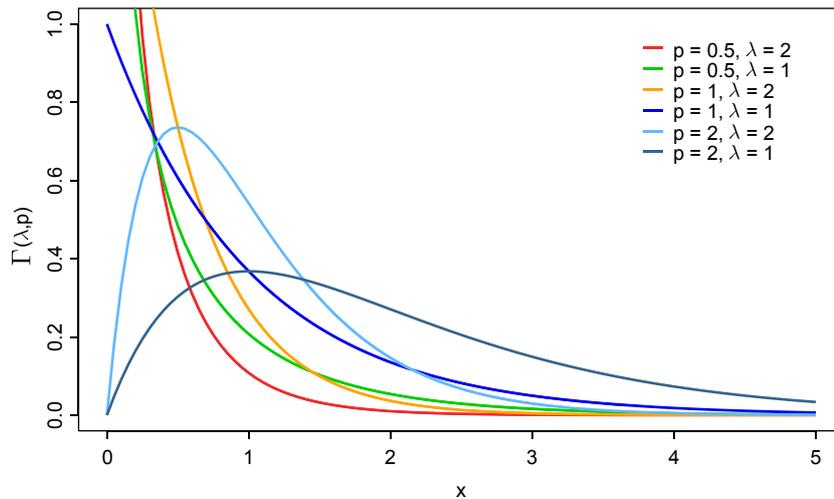


Abb. 3.1: Dichte der Gammaverteilung

Übungsaufgabe 3.2.1

Zeigen Sie, dass (3.2.1) eine Dichte ist.

Beispiel 3.2.1

1. In der Kraftfahrzeugversicherung wird die Gamma-Verteilung oft zur Modellierung des Gesamtschadens verwendet.
2. Falls $p = 1$, dann ist $\Gamma(\lambda, 1) = \text{Exp}(\lambda)$.

Satz 3.2.1 (Momenterzeugende und charakteristische Funktion der Gammaverteilung):

Falls $X \sim \Gamma(\lambda, p)$, dann gilt Folgendes:

1. Die momenterzeugende Funktion der Gammaverteilung $\Psi_X(s)$ ist gegeben durch

$$\Psi_X(s) = \mathbb{E}e^{sX} = \frac{1}{(1 - s/\lambda)^p}, \quad s < \lambda.$$

Die charakteristische Funktion der Gammaverteilung $\varphi_X(s)$ ist gegeben durch

$$\varphi_X(s) = \mathbb{E}e^{isX} = \frac{1}{(1 - is/\lambda)^p}, \quad s \in \mathbb{R}.$$

2. k -te Momente:

$$\mathbb{E}X^k = \frac{p(p+1) \cdot \dots \cdot (p+k-1)}{\lambda^k}, \quad k \in \mathbb{N}.$$

Beweis 1. Betrachte

$$\begin{aligned}\Psi_X(s) &= \int_0^\infty e^{sx} f_X(x) dx = \frac{\lambda^p}{\Gamma(p)} \int_0^\infty x^{p-1} e^{\overbrace{(s-\lambda)x}^{<0}} dx \\ &\stackrel{-(s-\lambda)x=y}{=} \frac{\lambda^p}{\Gamma(p)} \int_0^\infty \frac{y^{p-1}}{-(s-\lambda)^p} e^{-y} dy = \frac{\lambda^p \Gamma(p)}{\Gamma(p)(\lambda-s)^p} \\ &= \left(\frac{\lambda}{\lambda-s} \right)^p = \frac{1}{(1-s/\lambda)^p}, \quad \lambda > s.\end{aligned}$$

Falls $s \in \mathbb{C}$, $\operatorname{Re}(s) < \lambda$, dann ist $\Psi_X(s)$ holomorph auf $D = \{z = x + iy \in \mathbb{C} : x < \lambda\}$. Es gilt

$$\Psi_X(s) = \varphi_X(-is), \quad s = it, t < \lambda$$

Daraus folgt

$$\Psi_X(s) = \varphi_X(-is), \quad s \in D \implies \varphi_X(s) = \frac{1}{(1-is/\lambda)^p}, \quad s \in \mathbb{R}.$$

2.

$$\mathbb{E}X^k = \Psi^{(k)}(0) \implies \mathbb{E}X^k = \frac{p \cdot (p+1) \cdot \dots \cdot (p+k-1)}{\lambda^k}, \quad k \in \mathbb{N}.$$

□

Folgerung 3.2.1 (Faltungstabilität der Γ -Verteilung):

Falls $X \sim \Gamma(\lambda, p_1)$ und $Y \sim \Gamma(\lambda, p_2)$, X, Y unabhängig, dann ist $X + Y \sim \Gamma(\lambda, p_1 + p_2)$.

Beweis Es gilt

$$\varphi_{X+Y}(s) = \varphi_X(s) \cdot \varphi_Y(s) = \frac{1}{(1-is/\lambda)^{p_1}} \cdot \frac{1}{(1-is/\lambda)^{p_2}} = \left(\frac{1}{1-is/\lambda} \right)^{p_1+p_2} = \varphi_{\Gamma(\lambda, p_1+p_2)}(s).$$

Da die charakteristischen Funktionen die Verteilungen eindeutig bestimmen, folgt damit $X + Y \sim \Gamma(\lambda, p_1 + p_2)$. □

Beispiel 3.2.2

Seien $X_1, \dots, X_n \sim \operatorname{Exp}(\lambda)$ unabhängig. Nach der Folgerung 3.2.1 gilt $X = X_1 + \dots + X_n \sim \Gamma(\lambda, \underbrace{1 + \dots + 1}_n) = \Gamma(\lambda, n)$, denn $\operatorname{Exp}(\lambda) = \Gamma(\lambda, 1)$. Dabei heißt X *Erlang-verteilt* mit Parametern λ und n . Man schreibt $X \sim \operatorname{Erl}(\lambda, n)$.

$$\text{Zusammengefasst:} \quad \operatorname{Erl}(\lambda, n) = \Gamma(\lambda, n)$$

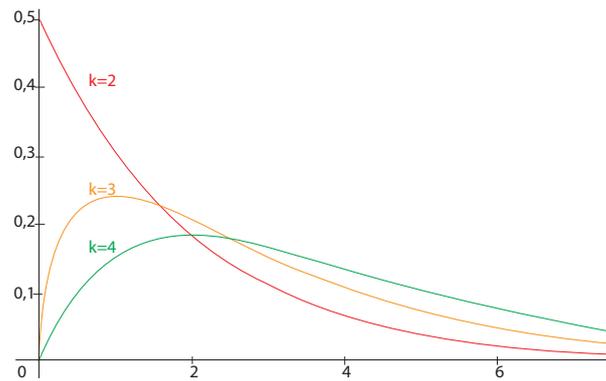
Interpretation: In der Risikotheorie z.B. sind X_i Zwischenankunftszeiten der Einzelschäden. Dann ist $X = \sum_{i=1}^n X_i$ die Ankunftszeit des n -ten Schadens, $X \sim \operatorname{Erl}(\lambda, n)$.

Definition 3.2.2 (χ^2 -Verteilung):

X ist eine χ^2 -verteilte Zufallsvariable mit k Freiheitsgraden ($X \sim \chi_k^2$), falls $X \stackrel{d}{=} X_1^2 + \dots + X_k^2$, wobei $X_1, \dots, X_k \sim N(0, 1)$ unabhängige identisch verteilte Zufallsvariablen sind.

Satz 3.2.2 (χ^2 -Verteilung: Spezialfall der Γ -Verteilung mit $\lambda = 1/2$, $p = k/2$):

Falls $X \sim \chi_k^2$, dann gilt:

Abb. 3.2: Dichte der χ^2 -Verteilung für $k = 2, 3, 4$

1. $X \sim \Gamma(1/2, k/2)$, d.h.

$$f_X(x) = \begin{cases} \frac{x^{k/2-1} e^{-x/2}}{2^{k/2} \Gamma(k/2)}, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (3.2.2)$$

2. Insbesondere ist $\mathbb{E}X = k$, $\text{Var} X = 2k$.

Beweis 1. Sei $X = X_1^2 + \dots + X_k^2$ mit $X_i \sim N(0, 1)$ unabhängigen identisch verteilten Zufallsvariablen. Errechnen wir zunächst die Verteilung der X_i^2 :

$$\begin{aligned} P(X_1^2 \leq x) &= P(X_1 \in [-\sqrt{x}, \sqrt{x}]) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy + \int_{-\sqrt{x}}^0 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &\stackrel{y^2=t}{=} \int_0^x \frac{1}{\sqrt{2\pi}} e^{-t/2} \frac{1}{2\sqrt{t}} dt + \int_x^0 \frac{1}{\sqrt{2\pi}} e^{-t/2} \frac{-1}{2\sqrt{t}} dt \\ &= \int_0^x \frac{(1/2)^{-1/2} t^{1/2-1}}{\Gamma(1/2)} e^{-t/2} dt, \quad x \geq 0. \end{aligned}$$

Somit folgt $X_1^2 \sim \Gamma(1/2, 1/2) \implies X \sim \Gamma(1/2, \underbrace{1/2 + \dots + 1/2}_k) = \Gamma(1/2, k/2)$ und daher gilt der Ausdruck (3.2.2) für die Dichte.

2. Wegen der Additivität des Erwartungswertes und der Unabhängigkeit von X_i gilt

$$\mathbb{E}X = k \cdot \mathbb{E}X_1^2, \quad \text{Var} X = k \text{Var} X_1^2, \quad \mathbb{E}(X_1^2) = \mathbb{E}(\Gamma(1/2, 1/2)).$$

Bitte zeigen Sie selbstständig, dass $\mathbb{E}X_1^2 = 1$, $\text{Var} X_1^2 = 2$.

□

3.2.2 Student-Verteilung (t-Verteilung)

Definition 3.2.3

Seien X, Y unabhängige Zufallsvariablen, wobei $X \sim N(0, 1)$ und $Y \sim \chi_r^2$. Dann heißt die

Zufallsvariable

$$U \stackrel{d}{=} \frac{X}{\sqrt{Y/r}}$$

Student- oder t -verteilt mit r Freiheitsgraden. Wir schreiben $U \sim t_r$.

Satz 3.2.3 (Dichte der t -Verteilung):

Falls $X \sim t_r$, dann gilt:

1.

$$f_X(x) = \frac{1}{\sqrt{r}B\left(\frac{r}{2}, \frac{1}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{r}\right)^{\frac{r+1}{2}}}, \quad x \in \mathbb{R}.$$

2. $\mathbb{E}X = 0$, $\text{Var} X = \frac{r}{r-2}$, $r \geq 3$.

Bemerkung 3.2.1

1. **Grafik von f_r :** Die t_r -Verteilung ist symmetrisch. Insbesondere gilt:

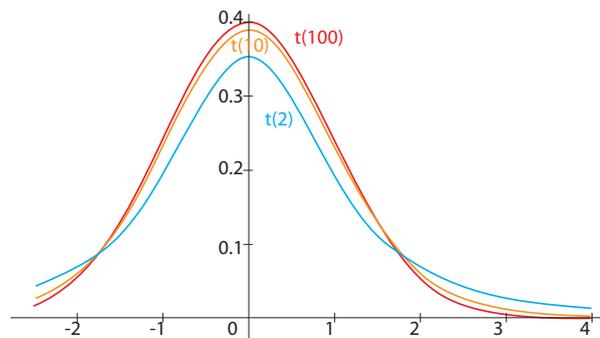


Abb. 3.3: Dichte \hat{f} der t -Verteilung für $r = 2, 10, 100$

$$t_{r,\alpha} = -t_{r,1-\alpha}, \quad \alpha \in (0, 1),$$

wobei $t_{r,\alpha}$ das α -Quantil der Student-Verteilung mit r Freiheitsgraden ist.

2. Falls $r \rightarrow \infty$, dann $f_r(x) \rightarrow \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$. (Übungsaufgabe)

3. Für $r = 1$ gilt: $t_1 = \text{Cauchy}(0, 1)$ mit Dichte $f(x) = \frac{1}{\pi(1+x^2)}$. Der Erwartungswert von t_1 existiert nicht.

Beweis des Satzes 3.2.3:

1. Es gilt $X := \varphi(Y, Z)$, wobei $\varphi(x, y) = \frac{x}{\sqrt{y/r}}$ und $V = (Y, Z)$ ein zweidimensionaler Zufallsvektor ist, $Y \sim N(0, 1)$, $Z \sim \chi_r^2$, Y und Z unabhängig.

Wir wollen den sogenannten *Dichtetransformationssatz für Zufallsvektoren* verwenden, der besagt, dass unter bestimmten Voraussetzungen

$$f_{\varphi(V)}(x) = f_V(\varphi^{-1}(x))|J|$$

gilt, wobei $|J| = |\det J|$, $J = \left(\frac{\partial \varphi_i^{-1}(x)}{\partial x_j} \right)_{i,j=1}^n$, $\varphi = (\varphi_1, \dots, \varphi_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Berechnen wir hier φ^{-1} von $\varphi : (x, y) \mapsto (v, w)$, wobei $v = \frac{x}{\sqrt{y/r}}$, $w = y$:

$$\varphi^{-1} : v = \frac{x}{\sqrt{\frac{y}{r}}} \implies x = v\sqrt{\frac{y}{r}} = v\sqrt{\frac{w}{r}}. \quad \text{Somit } \varphi^{-1} : (v, w) \mapsto \left(v\sqrt{\frac{w}{r}}, w \right)$$

und die Jacobi-Matrix ist gleich

$$J = \begin{pmatrix} \frac{\partial \varphi_1^{-1}}{\partial v} & \frac{\partial \varphi_2^{-1}}{\partial w} \\ \frac{\partial \varphi_2^{-1}}{\partial v} & \frac{\partial \varphi_2^{-1}}{\partial w} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{w}{r}} & \frac{v}{2\sqrt{wr}} \\ 0 & 1 \end{pmatrix}.$$

Falls $V = (Y, Z)$, Y und Z unabhängig, dann

$$f_V(x, y) = f_Y(x) \cdot f_Z(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{y^{r/2-1} e^{-y/2}}{\Gamma(r/2) 2^{r/2}} = \frac{y^{r/2-1} e^{-\frac{y+x^2}{2}}}{2^{\frac{r+1}{2}} \Gamma(1/2) \Gamma(r/2)}, \quad x \in \mathbb{R}, y > 0,$$

und nach dem Dichtetransformationssatz gilt

$$\begin{aligned} f_X(v) &= \int_0^\infty f_{\varphi(V)}(u, w) dw = \int_0^\infty f_V(\varphi^{-1}(v, w)) |J| dw \\ &= \int_0^\infty \frac{e^{-(v^2 \frac{w}{r} + w)/2} w^{r/2-1}}{2^{\frac{r+1}{2}} \Gamma(1/2) \Gamma(r/2)} \sqrt{w/r} dw \\ &= \frac{1}{\sqrt{r} 2^{\frac{r+1}{2}} \Gamma(1/2) \Gamma(r/2)} \cdot \int_0^\infty w^{\frac{r-1}{2}} e^{-\frac{\overbrace{v^2 + 1}^=}{2} \cdot w} dw \\ &\stackrel{w = \frac{2t}{v^2/r + 1}}{=} \frac{1}{\sqrt{r} 2^{\frac{r+1}{2}} \Gamma(1/2) \Gamma(r/2)} \cdot \int_0^\infty \frac{2^{\frac{r-1}{2} + 1} t^{\frac{r-1}{2}}}{(v^2/r + 1)^{\frac{r-1}{2} + 1}} e^{-t} dt \\ &= \frac{2^{\frac{r+1}{2}} \Gamma(\frac{r+1}{2})}{(v^2/r + 1)^{\frac{r+1}{2}} \sqrt{r} 2^{\frac{r+1}{2}} \Gamma(1/2) \Gamma(r/2)} = \frac{1}{\sqrt{r} B(r/2, 1/2) (1 + v^2/r)^{\frac{r+1}{2}}} \end{aligned}$$

2. Übungsaufgabe

□

Da im WR-Skript der Dichtetransformationssatz nur für Zufallsvariablen formuliert wurde, geben wir hier die notwendigen Begriffe und verallgemeinerten Sätze für Zufallsvektoren (ohne Beweis). Hierbei verwenden wir die folgende Notation:

Für Vektoren $x = (x_1, \dots, x_n)^T$ und $y = (y_1, \dots, y_n)^T$ schreiben wir $x \leq y$, falls $x_i \leq y_i$ für $i = 1, \dots, n$. Ferner sei für einen Zufallsvektor $X = (X_1, \dots, X_n)^T$ die Verteilungsfunktion definiert als $F(x) = \mathbb{P}(X \leq x)$ für $x = (x_1, \dots, x_n)^T$.

Definition 3.2.4

Die Zufallsvektoren $X_i : \Omega \rightarrow \mathbb{R}^{m_i}$, $i = 1, \dots, n$ sind *unabhängig*, falls

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i),$$

$x_i \in \mathbb{R}^{m_i}$, $i = 1, \dots, n$.

Satz 3.2.4

Falls X_i absolut stetig verteilte und unabhängige Zufallsvektoren mit Dichten f_{X_i} , $i = 1, \dots, n$, sind, dann ist auch (X_1, \dots, X_n) absolut stetig verteilt mit Dichte

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i), \quad x_i \in \mathbb{R}^{m_i}, \quad i = 1, \dots, n.$$

Satz 3.2.5

Falls $X_i : \Omega \rightarrow \mathbb{R}^{m_i}$, $i = 1, \dots, n$ unabhängige Zufallsvektoren sind, und $\varphi_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$, $\forall i = 1, \dots, n$ Borel-messbare Funktionen, dann sind Zufallsvektoren $\varphi_1(X_1), \dots, \varphi_n(X_n)$ unabhängig.

Satz 3.2.6 (Dichtetransformationssatz für Zufallsvektoren):

Sei $X = (X_1, \dots, X_m)^T : \Omega \rightarrow \mathbb{R}^m$ ein absolut stetig verteilter Zufallsvektor mit Dichte f_X . Sei $\varphi = (\varphi_1, \dots, \varphi_m)^T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ eine Borel-messbare Abbildung, die innerhalb von einem Quader $B \subset \mathbb{R}^m$ stetig differenzierbar ist. Falls $\text{supp} f_X \subset B$ und $\det \left(\frac{\partial \varphi_i}{\partial x_j} \right)_{i,j=1, \dots, m} \neq 0$ auf B , dann $\exists \varphi^{-1} : \varphi(B) \rightarrow B$ stetig differenzierbar und

$$f_{\varphi(X)}(x) = \begin{cases} f_X(\varphi^{-1}(x)) \cdot |J|, & x \in \varphi(B), \\ 0, & x \notin \varphi(B), \end{cases}$$

wobei $J = \det \left(\frac{\partial \varphi_i^{-1}}{\partial x_j} \right)_{i,j=1, \dots, m}$

3.2.3 Fisher-Snedecor-Verteilung (F-Verteilung)**Definition 3.2.5**

Falls $X \stackrel{d}{=} \frac{U_r/r}{U_s/s}$, wobei $U_r \sim \chi_r^2$, $U_s \sim \chi_s^2$, $r, s \in \mathbb{N}$, U_r, U_s unabhängig, dann hat X eine F-Verteilung mit Freiheitsgraden r, s . Bezeichnung: $X \sim F_{r,s}$.

Lemma 3.2.1

Falls $X \sim F_{r,s}$, dann ist X absolut stetig verteilt mit Dichte

$$f_X(x) = \frac{x^{r/2-1}}{B(r/2, s/2)(r/s)^{-r/2}(1 + (r/s) \cdot x)^{\frac{r+s}{2}}} \cdot \mathbb{I}(x > 0).$$

Beweis Da $U_r \sim \chi_r^2$, gilt für ihre Dichte

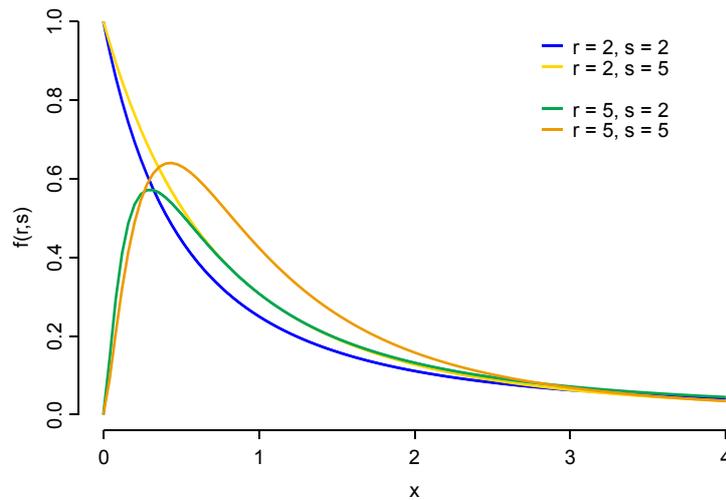
$$f_{U_r}(x) = \frac{x^{r/2-1} e^{-x/2}}{\Gamma(r/2) 2^{r/2}}, \quad x > 0, \quad r \in \mathbb{N}.$$

Somit

$$P(U_r/r \leq x) = P(U_r \leq rx) = F_{U_r}(rx)$$

und deshalb

$$\begin{aligned} f_{U_r/r}(x) &= (F_{U_r}(rx))' = r \cdot f_{U_r}(rx) = \frac{r(rx)^{r/2-1} e^{-rx/2}}{\Gamma(r/2) 2^{r/2}} \cdot \mathbb{I}(x > 0) \\ &= \frac{r^{r/2} x^{r/2-1} e^{-r/2 \cdot x}}{\Gamma(r/2) 2^{r/2}} \cdot \mathbb{I}(x > 0). \end{aligned}$$

Abb. 3.4: Dichte der F-Verteilung für verschiedene Parameter r und s .

Nach dem Dichtetransformationssatz für das Verhältnis von zwei Zufallsvariablen (vgl. Wahrscheinlichkeitsskript Satz 3.15) gilt

$$f_{\frac{U_r/r}{U_s/s}}(x) = \int_0^\infty t f_{U_r/r}(xt) \cdot f_{U_s/s}(t) dt \cdot \mathbb{I}(x > 0).$$

Somit

$$\begin{aligned} f_X(x) &= \int_0^\infty t \frac{r^{r/2} (tx)^{r/2-1} e^{-\frac{rtx}{2}}}{\Gamma(r/2) 2^{r/2}} \cdot \frac{s^{s/2} t^{s/2-1} e^{-st/2}}{\Gamma(s/2) 2^{s/2}} dt \\ &= \frac{r^{r/2} s^{s/2} x^{r/2-1}}{\Gamma(r/2) \Gamma(s/2) 2^{\frac{r+s}{2}}} \cdot \int_0^\infty t^{r/2+s/2-1} e^{-\frac{rx+s}{2}t} dt \\ &= \frac{r^{r/2} s^{s/2} x^{r/2-1}}{\Gamma(r/2) \Gamma(s/2)} \cdot \int_0^\infty \frac{y^{\frac{r+s}{2}-1}}{(rx+s)^{\frac{r+s}{2}}} \cdot e^{-y} dy \\ &\stackrel{t=\frac{y}{rx+s}}{=} \frac{r^{r/2} s^{s/2} x^{r/2-1}}{\Gamma(r/2) \Gamma(s/2) s^{\frac{r+s}{2}} (1 + \frac{r}{s} \cdot x)^{\frac{r+s}{2}}} \cdot \Gamma\left(\frac{r+s}{2}\right) \\ &= \frac{(r/s)^{r/2} x^{r/2-1}}{B(r/2, s/2) (1 + \frac{r}{s}x)^{\frac{r+s}{2}}} \cdot \mathbb{I}(x > 0). \end{aligned}$$

□

Bemerkung 3.2.2

Sei $X \sim F_{r,s}$, $r, s \in \mathbb{N}$ mit Dichte f_X .

1. Einige Graphen der F-Verteilung sind in Abbildung 3.4 dargestellt.
2. Einige Eigenschaften der F-Verteilung:

Lemma 3.2.2

Es gilt:

a)
$$\mathbb{E}X = \frac{s}{s-2}, \quad s \geq 3.$$

b)
$$\text{Var } X = \frac{2s^2(r+s-2)}{r(s-4)(s-2)^2}, \quad s \geq 5.$$

c) Falls $F_{r,s,\alpha}$ das α -Quantil der $F_{r,s}$ -Verteilung ist, dann gilt

$$F_{r,s,\alpha} = \frac{1}{F_{s,r,1-\alpha}}, \quad \alpha \in (0,1).$$

Übungsaufgabe 3.2.2

Beweisen Sie Lemma 3.2.2!

3. Für Quantile $F_{r,s,\alpha}$ gilt folgende Näherungsformel (Abramowitz, Stegun (1972)):

$F_{r,s,\alpha} \approx e^\omega$, wobei

$$\begin{aligned} \omega &= 2 \left(\frac{\alpha(h+a)^{1/2}}{h} - \left(\frac{1}{r-1} - \frac{1}{s-1} \right) \cdot \left(a + \frac{5}{6} - \frac{2}{3h} \right) \right), \\ h &= 2 \left(\frac{1}{r-1} + \frac{1}{s-1} \right)^{-1}, \\ a &= \frac{z_\alpha^2 - 3}{6} \end{aligned}$$

und z_α das α -Quantil der $N(0,1)$ -Verteilung ist.

3.3 Punktschätzer und ihre Grundeigenschaften

Sei (X_1, \dots, X_n) eine Zufallsstichprobe, definiert auf dem kanonischen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$. Seien X_i , $i = 1, \dots, n$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F \in \{F_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$. Finde einen Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für den Parameter θ mit vorgegebenen Eigenschaften.

Unser Ziel im nächsten Abschnitt ist es, zunächst grundlegende Eigenschaften der Schätzer kennenzulernen.

3.3.1 Eigenschaften von Punktschätzern**Definition 3.3.1 (Erwartungstreue):**

Ein Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ heißt *erwartungstreu* oder *unverzerrt*, falls

$$\mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) = \theta, \quad \theta \in \Theta.$$

Dabei wird vorausgesetzt, dass

$$\mathbb{E}_\theta |\hat{\theta}(X_1, \dots, X_n)| < \infty, \quad \theta \in \Theta.$$

Der *Bias* (*Verzerrung*) eines Schätzers $\hat{\theta}(X_1, \dots, X_n)$ ist gegeben durch

$$\text{Bias}(\hat{\theta}) = \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) - \theta.$$

Falls $\hat{\theta}(X_1, \dots, X_n)$ erwartungstreu ist, dann gilt $\text{Bias}(\hat{\theta}) = 0$ (kein systematischer Schätzfehler).

Definition 3.3.2 (Asymptotische Erwartungstreue):

Der Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ heißt *asymptotisch erwartungstreu* (oder *asymptotisch unverzerrt*), falls (für große Datenmengen)

$$\mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) \xrightarrow{n \rightarrow \infty} \theta.$$

Definition 3.3.3 (Konsistenz):

Falls

$$\hat{\theta}(X_1, \dots, X_n) \xrightarrow{n \rightarrow \infty} \theta$$

in L^2 , stochastisch bzw. fast sicher, dann heißt der Schätzer $\hat{\theta}(X_1, \dots, X_n)$ ein *konsistenter Schätzer* für θ im *mittleren quadratischen, schwachen bzw. starken Sinne*.

- $\hat{\theta}$ *L^2 -konsistent*: für $\mathbb{E}_\theta \hat{\theta}^2(X_1, \dots, X_n) < \infty$ gilt

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{L^2} \theta \iff \mathbb{E}_\theta |\hat{\theta}(X_1, \dots, X_n) - \theta|^2 \xrightarrow{n \rightarrow \infty} 0, \quad \theta \in \Theta.$$

- $\hat{\theta}$ *schwach konsistent*:

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta \iff P_\theta(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0, \quad \varepsilon > 0, \quad \theta \in \Theta.$$

- $\hat{\theta}$ *stark konsistent*:

$$\hat{\theta} \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \theta \iff P_\theta \left(\lim_{n \rightarrow \infty} \hat{\theta}(X_1, \dots, X_n) = \theta \right) = 1, \quad \theta \in \Theta.$$

Daraus ergibt sich folgendes Diagramm (vgl. Wahrscheinlichkeitsrechnungsskript, Kapitel 6).



Definition 3.3.4 (Mittlerer quadratischer Fehler (mean squared error)):

Der mittlere quadratische Fehler eines Schätzers $\hat{\theta}(X_1, \dots, X_n)$ für θ ist definiert als

$$MSE(\hat{\theta}) = \mathbb{E}_\theta |\hat{\theta}(X_1, \dots, X_n) - \theta|^2.$$

Lemma 3.3.1

Falls $m = 1$ und $\mathbb{E}_\theta \hat{\theta}^2(X_1, \dots, X_n) < \infty$, $\theta \in \Theta$, dann gilt

$$MSE(\hat{\theta}) = \text{Var}_\theta \hat{\theta} + (\text{Bias}(\hat{\theta}))^2.$$

Beweis

$$\begin{aligned} MSE(\hat{\theta}) &= \mathbb{E}_\theta (\hat{\theta} - \theta)^2 = \mathbb{E}_\theta (\hat{\theta} - \mathbb{E}_\theta \hat{\theta} + \mathbb{E}_\theta \hat{\theta} - \theta)^2 \\ &= \underbrace{\mathbb{E}_\theta (\hat{\theta} - \mathbb{E}_\theta \hat{\theta})^2}_{\text{Var}_\theta \hat{\theta}} + \underbrace{2 \mathbb{E}_\theta (\hat{\theta} - \mathbb{E}_\theta \hat{\theta}) (\mathbb{E}_\theta \hat{\theta} - \theta)}_{=0} + \underbrace{(\mathbb{E}_\theta \hat{\theta} - \theta)^2}_{= \text{Bias}(\hat{\theta})^2} \\ &= \text{Var}_\theta \hat{\theta} + (\text{Bias}(\hat{\theta}))^2. \end{aligned}$$

□

Bemerkung 3.3.1

Falls $\hat{\theta}$ erwartungstreu für θ ist, dann gilt $MSE(\hat{\theta}) = \text{Var}_\theta \hat{\theta}$.

Definition 3.3.5 (Vergleich von Schätzern):

Seien $\hat{\theta}_1(X_1, \dots, X_n)$ und $\hat{\theta}_2(X_1, \dots, X_n)$ zwei Schätzer für θ . Man sagt, dass $\hat{\theta}_1$ *besser* ist als $\hat{\theta}_2$, falls

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2), \quad \theta \in \Theta.$$

Falls $m = 1$ und die Schätzer $\hat{\theta}_1, \hat{\theta}_2$ erwartungstreu sind, so ist $\hat{\theta}_1$ *besser als* $\hat{\theta}_2$, falls $\hat{\theta}_1$ die kleinere Varianz besitzt. Dabei wird stets vorausgesetzt, dass $\mathbb{E}_\theta \hat{\theta}_i^2 < \infty$, $\theta \in \Theta$.

Definition 3.3.6 (Asymptotische Normalverteiltheit):

Sei $\hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ ($m = 1$). Falls $0 < \text{Var}_\theta \hat{\theta}(X_1, \dots, X_n) < \infty$, $\theta \in \Theta$ und

$$\frac{\hat{\theta}(X_1, \dots, X_n) - \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n)}{\sqrt{\text{Var}_\theta \hat{\theta}(X_1, \dots, X_n)}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1),$$

dann ist $\hat{\theta}(X_1, \dots, X_n)$ *asymptotisch normalverteilt*.

Definition 3.3.7 (Bester erwartungstreuer Schätzer):

Der Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ ist der *beste erwartungstreue Schätzer*, falls

$$\mathbb{E}_\theta \hat{\theta}^2(X_1, \dots, X_n) < \infty, \quad \theta \in \Theta, \quad \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) = \theta, \quad \theta \in \Theta,$$

und $\hat{\theta}$ die minimale Varianz in der Klasse aller erwartungstreuen Schätzer für θ besitzt. Das heißt, dass für einen beliebigen erwartungstreuen Schätzer $\tilde{\theta}(X_1, \dots, X_n)$ mit

$$\mathbb{E}_\theta \tilde{\theta}^2(X_1, \dots, X_n) < \infty \quad \text{gilt} \quad \text{Var}_\theta \hat{\theta} \leq \text{Var}_\theta \tilde{\theta}, \quad \theta \in \Theta.$$

3.3.2 Schätzer des Erwartungswertes und empirische Momente

Sei $X \stackrel{d}{=} X_i$, $i = 1, \dots, n$ ein statistisches Merkmal. Sei weiter $\mathbb{E}|X_i|^k < \infty$ für ein $k \in \mathbb{N}$, $m = 1$ und der zu schätzende Parameter $\theta = \mu_k = \mathbb{E}X_i^k$. Insbesondere gilt im Fall $k = 1$, dass $\theta = \mu_1 = \mu$ der Erwartungswert ist.

Definition 3.3.8

Das k -te *empirische Moment* von X wird als

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

definiert. Unter dieser Definition gilt, dass $\hat{\mu}_1 = \bar{X}_n$, also das erste empirische Moment gleich dem Stichprobenmittel ist.

Satz 3.3.1 (Eigenschaften der empirischen Momente):

Unter obigen Voraussetzungen gelten folgende Eigenschaften:

1. $\hat{\mu}_k$ ist erwartungstreu für μ_k (insbesondere \bar{X}_n).
2. $\hat{\mu}_k$ ist stark konsistent.
3. Falls $\mathbb{E}_\theta |X|^{2k} < \infty$, $\forall \theta \in \Theta$, dann ist $\hat{\mu}_k$ asymptotisch normalverteilt.

4. Es gilt $\text{Var } \bar{X}_n = \frac{\sigma^2}{n}$, wobei $\sigma^2 = \text{Var}_\theta X$. Falls $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$ (eine normalverteilte Stichprobe), dann gilt:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Beweis

1.
$$\mathbb{E}_\theta \hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta X_i^k = \frac{1}{n} \sum_{i=1}^n \mu_k = \frac{n\mu_k}{n} = \mu_k.$$

2. Aus dem starken Gesetz der großen Zahlen folgt

$$\frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}_\theta X_i^k = \mu_k.$$

3. Mit dem zentralen Grenzwertsatz gilt

$$\frac{\sum_{i=1}^n X_i^k - n \cdot \mathbb{E} X^k}{\sqrt{n \cdot \text{Var } X^k}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i^k - \mu_k}{\frac{1}{\sqrt{n}} \sqrt{\text{Var } X^k}} = \sqrt{n} \frac{\hat{\mu}_k - \mu_k}{\sqrt{\text{Var } X^k}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

Insbesondere gilt für den Spezialfall $k = 1$

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

- 4.

$$\text{Var } \bar{X}_n = \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \stackrel{X_i \text{ u.i.v.}}{=} \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Falls $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$, dann gilt wegen der Faltungstabilität der Normalverteilung $\bar{X}_n \sim N(\cdot, \cdot)$, weil

$$\frac{1}{n} X_i \sim N\left(\frac{\mu}{n}, \frac{\sigma^2}{n^2}\right), \quad X_i \text{ u.i.v.}$$

Somit folgt aus 1) und 4) $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Damit ist der Satz bewiesen. □

Bemerkung 3.3.2

Aus Satz 3.3.1, 3) folgt

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) &= 1 - \mathbb{P}(-\varepsilon \leq \bar{X}_n - \mu \leq \varepsilon) \\ &= 1 - \mathbb{P}\left(-\frac{\varepsilon\sqrt{n}}{\sigma} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\ &\stackrel{n \rightarrow \infty}{\approx} 1 - \left(\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \\ &\stackrel{\Phi(-x) = 1 - \Phi(x)}{=} 1 - \left(\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 + \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \\ &= 1 - \left(2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1\right), \end{aligned}$$

wobei $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ die Verteilungsfunktion der $N(0, 1)$ -Verteilung ist. Insgesamt gilt also für großes n

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \approx 2 \left(1 - \Phi \left(\frac{\varepsilon \sqrt{n}}{\sigma} \right) \right).$$

3.3.3 Schätzer der Varianz

Seien X_i , $i = 1, \dots, n$ unabhängig identisch verteilt, $X_i \stackrel{d}{=} X$, $\mathbb{E}_\theta X^2 < \infty \quad \forall \theta \in \Theta$, $\theta = (\theta_1, \dots, \theta_m)^T$, $\theta_i = \sigma^2 = \text{Var}_\theta X$ für ein $i \in \{1, \dots, m\}$. Die *Stichprobenvarianz*

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

ist dann ein Schätzer für σ^2 . Falls der Erwartungswert $\mu = \mathbb{E}_\theta X$ der Stichprobenvariablen explizit benannt ist, so kann ein Schätzer für σ^2 auch als

$$\tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

definiert werden.

Wir werden nun die Eigenschaften von S_n^2 und \tilde{S}_n^2 untersuchen und sie miteinander vergleichen.

Satz 3.3.2

1. Die Stichprobenvarianz S_n^2 ist erwartungstreu für σ^2 :

$$\mathbb{E}_\theta S_n^2 = \sigma^2, \quad \theta \in \Theta.$$

2. Wenn $\mathbb{E}_\theta X^4 < \infty$, dann gilt

$$\text{Var}_\theta S_n^2 = \frac{1}{n} \left(\mu'_4 - \frac{n-3}{n-1} \sigma^4 \right),$$

wobei $\mu'_4 = \mathbb{E}_\theta (X - \mu)^4$.

Beweis 1. Aus Lemma 2.2.1 1), 2) folgt, dass

$$S_n^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right),$$

und dass man o.B.d.A. $\mu = \mathbb{E}_\theta X_i = 0$ annehmen kann, woraus insbesondere $\mathbb{E}_\theta \bar{X}_n = 0$, $\theta \in \Theta$ folgt. Dann gilt

$$\begin{aligned} \mathbb{E}_\theta S_n^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n \mathbb{E}_\theta X_i^2 - n\mathbb{E}_\theta \bar{X}_n^2 \right) = \frac{1}{n-1} \left(\sum_{i=1}^n \text{Var}_\theta X_i - n\text{Var}_\theta \bar{X}_n \right) \\ &\stackrel{\text{s. 3.3.1, 4)}}{=} \frac{1}{n-1} \left(n\sigma^2 - n \cdot \frac{\sigma^2}{n} \right) = \sigma^2, \quad \theta \in \Theta. \end{aligned}$$

2. Berechnen wir $\text{Var}_\theta S_n^2 = \mathbb{E}_\theta(S_n^2)^2 - (\mathbb{E}_\theta S_n^2)^2 = \mathbb{E}_\theta(S_n^2)^2 - \sigma^4$. Es gilt

$$\begin{aligned} \mathbb{E}_\theta S_n^4 &= \frac{1}{(n-1)^2} \mathbb{E}_\theta \left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right)^2 \\ &= \frac{1}{(n-1)^2} \left(\underbrace{\mathbb{E}_\theta \left(\sum_{i=1}^n X_i^2 \right)^2}_{=I_1} - 2n \underbrace{\mathbb{E}_\theta \left(\bar{X}_n^2 \sum_{i=1}^n X_i^2 \right)}_{=I_2} + n^2 \underbrace{\mathbb{E}_\theta \bar{X}_n^4}_{=I_3} \right). \end{aligned}$$

Dabei gilt

$$\begin{aligned} I_1 &= \mathbb{E}_\theta \left(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2 \right) = \mathbb{E}_\theta \left(\sum_{i=1}^n X_i^4 + \sum_{i \neq j} X_i^2 X_j^2 \right) = \sum_{i=1}^n \mathbb{E}_\theta X_i^4 + \sum_{i \neq j} \mathbb{E}_\theta (X_i^2 X_j^2) \\ &\stackrel{X_i \text{ u.i.v.}, \mu=0}{=} \sum_{i=1}^n \mu'_4 + \sum_{i \neq j} \text{Var}_\theta X_i \cdot \text{Var}_\theta X_j = n\mu'_4 + n(n-1)\sigma^4, \end{aligned}$$

$$\begin{aligned} I_2 &= \mathbb{E}_\theta \left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \sum_{j=1}^n X_j^2 \right) = \frac{1}{n^2} \mathbb{E}_\theta \left(\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j \right) \sum_{j=1}^n X_j^2 \right) \\ &= \frac{1}{n^2} \mathbb{E}_\theta \left(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2 \right) + \frac{1}{n^2} \mathbb{E}_\theta \left(\sum_{i \neq j} X_i X_j \sum_{k=1}^n X_k^2 \right) \\ &= \frac{1}{n^2} I_1 + \frac{1}{n^2} \sum_{i \neq j} \sum_k \underbrace{\mathbb{E} (X_i X_j X_k^2)}_{=0, \text{ da } X_i \text{ u.i.v. und } \mu=0} = \frac{I_1}{n^2} = \frac{\mu'_4 + (n-1)\sigma^4}{n}, \end{aligned}$$

$$\begin{aligned} I_3 &= \mathbb{E}_\theta \left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \cdot \left(\frac{1}{n} \sum_{j=1}^n X_j \right)^2 \right) \\ &= \frac{1}{n^4} \mathbb{E}_\theta \left(\left(\sum_{k=1}^n X_k^2 + \sum_{i \neq j} X_i X_j \right) \cdot \left(\sum_{r=1}^n X_r^2 + \sum_{s \neq t} X_s X_t \right) \right) \\ &= \frac{1}{n^4} \mathbb{E}_\theta \left(\sum_{k,r=1}^n X_k^2 X_r^2 + 2 \sum_{k=1}^n X_k^2 \sum_{i \neq j} X_i X_j + \sum_{i \neq j} X_i X_j \sum_{s \neq t} X_s X_t \right) \\ &= \frac{1}{n^4} \left(\mathbb{E} \left(\sum_{k=1}^n X_k^4 \right) + \mathbb{E}_\theta \left(\sum_{k \neq r} X_k^2 X_r^2 \right) + 2 \underbrace{\mathbb{E}_\theta \left(\sum_{k=1}^n X_k^2 \sum_{i \neq j} X_i X_j \right)}_{=0, \text{ da } X_i \text{ u.i.v. und } \mu=0} \right) \\ &\quad + 2 \underbrace{\mathbb{E}_\theta \left(\sum_{i \neq j} X_i^2 X_j^2 \right)}_{\text{weil } (i,j) \text{ und } (j,i) \text{ zählen}} + \underbrace{\mathbb{E}_\theta \left(\sum_{i \neq j \neq t} X_i^2 X_j X_t \right)}_{=0, \text{ da } X_j \text{ u.i.v. und } \mu=0} + \underbrace{\mathbb{E}_\theta \left(\sum_{i \neq j \neq s \neq t} X_i X_j X_s X_t \right)}_{=0, \text{ da } X_i \text{ u.i.v. und } \mu=0} \\ &= \frac{1}{n^4} \left(n\mu'_4 + 3\mathbb{E}_\theta \sum_{i \neq j} X_i^2 X_j^2 \right) = \frac{n\mu'_4 + 3n(n-1)\sigma^4}{n^4} = \frac{\mu'_4 + 3(n-1)\sigma^4}{n^3}. \end{aligned}$$

Somit gilt insgesamt

$$\begin{aligned}\mathbb{E}_\theta S_n^4 &= \frac{1}{(n-1)^2} \left(n\mu'_4 + n(n-1)\sigma^4 - 2(\mu'_4 + (n-1)\sigma^4) + \frac{\mu'_4 + 3(n-1)\sigma^4}{n} \right) \\ &= \frac{(n^2 - 2n + 1)\mu'_4 + (n^2 - 2n + 3)(n-1)\sigma^4}{n(n-1)^2} \\ &= \frac{(n-1)^2 \mu'_4}{(n-1)^2 n} + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4 = \frac{\mu'_4}{n} + \frac{n^2 - 2n + 3}{n(n-1)} \sigma^4\end{aligned}$$

und deshalb

$$\text{Var}_\theta S_n^2 = \frac{\mu'_4}{n} + \frac{n^2 - 2n + 3 - n^2 + n}{n(n-1)} \sigma^4 = \frac{\mu'_4}{n} - \frac{n-3}{n(n-1)} \sigma^4 = \frac{1}{n} \left(\mu'_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

□

Satz 3.3.3 1. Der Schätzer \tilde{S}_n^2 für σ^2 ist erwartungstreu.

2. Es gilt $\text{Var}_\theta \tilde{S}_n^2 = 1/n(\mu'_4 - \sigma^4)$.

Beweis

1.
$$\mathbb{E}_\theta \tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}_\theta (X_i - \mu)^2}_{=\text{Var}_\theta X_i} = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2.$$

2. Setzen wir wie in Satz 3.3.2 o.B.d.A. $\mu = 0$ voraus. Dann gilt

$$\begin{aligned}\text{Var } \tilde{S}_n^2 &= \mathbb{E}_\theta \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right)^2 - \left(\mathbb{E}_\theta \tilde{S}_n^2 \right)^2 = \frac{1}{n^2} \mathbb{E} \left(\sum_{i=1}^n X_i^2 \right)^2 - \sigma^4 \\ &\stackrel{\text{I}_1 \text{ Beweis S. 3.3.2}}{=} \frac{n\mu'_4 + n(n-1)\sigma^4}{n^2} - \sigma^4 = \frac{\mu'_4 + (n-1)\sigma^4}{n} - \sigma^4 = \frac{\mu'_4 - \sigma^4}{n}.\end{aligned}$$

□

Folgerung 3.3.1

Der Schätzer \tilde{S}_n^2 für σ^2 ist besser als S_n^2 , weil beide erwartungstreu sind und

$$\text{Var}_\theta \tilde{S}_n^2 = \frac{\mu'_4 - \sigma^4}{n} < \frac{\mu'_4 - \frac{n-3}{n-1}\sigma^4}{n} = \text{Var}_\theta S_n^2.$$

Diese Eigenschaft von \tilde{S}_n^2 im Vergleich zu S_n^2 ist intuitiv klar, da man in \tilde{S}_n^2 mehr Informationen über die Verteilung der Stichprobenvariablen X_i (nämlich den bekannten Erwartungswert μ) eingesteckt hat.

Satz 3.3.4

Die Schätzer S_n^2 bzw. \tilde{S}_n^2 sind stark konsistent und asymptotisch normalverteilt:

$$\begin{aligned}S_n^2 &\xrightarrow[n \rightarrow \infty]{\text{f.s.}} \sigma^2, & \sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} &\xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1), \\ \tilde{S}_n^2 &\xrightarrow[n \rightarrow \infty]{\text{f.s.}} \sigma^2, & \sqrt{n} \frac{\tilde{S}_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} &\xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1),\end{aligned}$$

falls $\mu'_4 < \infty$.

Beweis Zeigen wir nur, dass S_n^2 die obigen Eigenschaften besitzt. Der Beweis für \tilde{S}_n^2 verläuft analog. Die starke Konsistenz von S_n^2 folgt aus dem starken Gesetz der großen Zahlen, nach dem

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}X^2 \quad \text{und} \quad \bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mu$$

gilt und somit auch

$$\bar{X}_n^2 \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mu^2.$$

Dann

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right) = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 \right) \\ &\xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}_\theta X^2 - \mu^2 = \text{Var}_\theta X = \sigma^2, \end{aligned}$$

und die starke Konsistenz ist bewiesen. Um die asymptotische Normalverteiltheit zu beweisen, nehmen wir o.B.d.A. an, dass $\mu = \mathbb{E}_\theta X = 0$. Dann folgt mit Hilfe des Satzes von Slutsky (vgl. Sätze 6.8 - 6.9 aus dem WR-Skript)

$$\begin{aligned} \sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} &= \sqrt{n} \frac{\frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \\ &= \sqrt{n} \frac{1}{n-1} \frac{\sum_{i=1}^n X_i^2 - n\sigma^2}{\sqrt{\mu'_4 - \sigma^4}} - \underbrace{\frac{\sqrt{n} \frac{n}{n-1} \bar{X}_n^2}{\sqrt{\mu'_4 - \sigma^4}}}_{=R_n^1} - \underbrace{\left(1 - \frac{n}{n-1}\right) \frac{\sigma^2 \sqrt{n}}{\sqrt{\mu'_4 - \sigma^4}}}_{=R_n^2} \\ &\stackrel{d}{\underset{n \rightarrow \infty}{\rightsquigarrow}} \frac{\sum_{i=1}^n X_i^2 - n\sigma^2}{\sqrt{n(\mu'_4 - \sigma^4)}}, \end{aligned}$$

weil

$$R_n^2 = \left(1 - \frac{n}{n-1}\right) \sigma^2 \frac{\sqrt{n}}{\sqrt{\mu'_4 - \sigma^4}} = -\frac{\sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \frac{\sqrt{n}}{n-1} \xrightarrow[n \rightarrow \infty]{\text{f.s.}} 0,$$

also auch stochastisch und in Verteilung. Es gilt

$$R_n^1 \sim \sqrt{n} \frac{\bar{X}_n^2}{\sqrt{\mu'_4 - \sigma^4}} \xrightarrow[n \rightarrow \infty]{d} 0,$$

weil

$$\mathbb{E}_\theta \left(\sqrt{n} \bar{X}_n^2 \right) \underset{\mu=0}{=} \sqrt{n} \text{Var}_\theta \bar{X}_n \underset{\text{s. 3.3.1, 4}}{=} \sqrt{n} \frac{\sigma^2}{n} = \frac{\sigma^2}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} 0$$

und somit

$$\sqrt{n} (\bar{X}_n)^2 \xrightarrow[n \rightarrow \infty]{L^1} 0 \implies \sqrt{n} (\bar{X}_n)^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \implies \sqrt{n} (\bar{X}_n)^2 \xrightarrow[n \rightarrow \infty]{d} 0.$$

Dann gilt

$$\lim_{n \rightarrow \infty} \sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu'_4 - \sigma^4}} \stackrel{d}{=} \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n X_i^2 - n\sigma^2}{\sqrt{n(\mu'_4 - \sigma^4)}} \stackrel{d}{=} Y \sim N(0, 1)$$

nach dem zentralen Grenzwertsatz für die Folge von unabhängigen identisch verteilten Zufallsvariablen $\{X_i^2\}_{i \in \mathbb{N}}$, weil $\mathbb{E}_\theta X_i^2 \stackrel{\mu=0}{=} \text{Var}_\theta X = \sigma^2$ und

$$\text{Var}_\theta X_i^2 = \mathbb{E}_\theta X^4 - (\mathbb{E}_\theta X^2)^2 = \mu'_4 - \sigma^4.$$

□

Folgerung 3.3.2

Es gilt

$$1. \quad \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1)$$

und somit

$$2. \quad \mathbb{P} \left(\mu \in \left[\bar{X}_n - \frac{z_{1-\alpha/2} S_n}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2} S_n}{\sqrt{n}} \right] \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha \quad (3.3.1)$$

für ein $\alpha \in (0, 1)$, wobei z_α das α -Quantil der $N(0, 1)$ -Verteilung ist.

Bemerkung 3.3.3

Das Intervall in (3.3.1) nennt man *asymptotisches Konfidenz- oder Vertrauensintervall* für den Parameter μ . Falls α klein ist (z.B. $\alpha = 0,05$), so liegt μ mit einer asymptotisch großen Wahrscheinlichkeit $1 - \alpha$ im vorgegebenen Intervall. Diese Art der Schätzung von μ stellt eine Alternative zu den Punktschätzern dar und wird ausführlich in der Vorlesung Statistik II behandelt.

Beweis der Folgerung 3.3.2

1. Aus Satz 3.3.4 folgt

$$S_n^2 \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \sigma^2 \implies \frac{\sigma}{S_n} \xrightarrow[n \rightarrow \infty]{\text{f.s.}} 1 \implies \frac{\sigma}{S_n} \xrightarrow[n \rightarrow \infty]{d} 1$$

und somit nach der Verwendung des Satzes von Slutsky

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \cdot \frac{\sigma}{S_n} \xrightarrow[n \rightarrow \infty]{} Y \cdot 1 = Y \sim N(0, 1),$$

wobei wir die asymptotische Normalverteiltheit von \bar{X}_n benutzt haben.

2. Aus 1) folgt

$$\mathbb{P}_\theta \left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \in [z_{\alpha/2}, z_{1-\alpha/2}] \right) \xrightarrow[n \rightarrow \infty]{} \Phi(z_{1-\alpha/2}) - \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

Daraus folgt das Intervall (3.3.1) nach der Auflösung der Ungleichung

$$z_{\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq z_{1-\alpha/2}$$

bzgl. μ .

□

Betrachten wir weiterhin den wichtigen Spezialfall der normalverteilten Stichprobenvariablen X_i , $i = 1, \dots, n$, also $X \sim N(\mu, \sigma^2)$.

Satz 3.3.5

Falls X_1, \dots, X_n normalverteilt sind mit Parametern μ und σ^2 , dann gilt

1.
$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2,$$
2.
$$\frac{n\tilde{S}_n^2}{\sigma^2} \sim \chi_n^2.$$

Beweis Beweisen wir den schwierigeren Fall 1, der Beweis im Fall 2 verläuft analog.

Da $X_i \sim N(\mu, \sigma^2)$, gilt, dass $\frac{X_i - \mu}{\sigma} \sim N(0, 1)$ unabhängige identisch verteilte Zufallsvariablen für $i = 1, \dots, n$ sind. Nach Lemma 2.2.1 gilt

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2$$

und somit

$$T_1 = \underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2}_{\sim \chi_n^2} = \frac{n-1}{\sigma^2} S_n^2 + \underbrace{\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \right)^2}_{= T_2 \sim \chi_1^2 \text{ aus S. 3.3.1, 4)}.$$

In Lemma 3.3.2 wird bewiesen, dass S_n^2 und \bar{X}_n unabhängig sind. Somit gilt

$$\varphi_{T_1}(s) = \varphi_{\frac{n-1}{\sigma^2} S_n^2}(s) \cdot \varphi_{T_2}(s), \quad \forall s \in \mathbb{R},$$

wobei $\varphi_Z(s)$ die charakteristische Funktion einer Zufallsvariablen Z ist. Da nach dem Satz 3.2.1

$$\varphi_{T_1}(s) = \frac{1}{(1 - 2is)^{n/2}}, \quad \varphi_{T_2}(s) = \frac{1}{(1 - 2is)^{1/2}},$$

folgt

$$\varphi_{\frac{n-1}{\sigma^2} S_n^2}(s) = \frac{\varphi_{T_1}(s)}{\varphi_{T_2}(s)} = \frac{1}{(1 - 2is)^{(n-1)/2}} = \varphi_{\chi_{n-1}^2}(s).$$

Aus dem Satz 3.2.1 und dem Eindeutigkeitsatz für charakteristische Funktionen (vgl. Folgerung 5.1 aus dem WR-Skript) folgt

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2.$$

□

Lemma 3.3.2

Falls $X \sim N(\mu, \sigma^2)$, X_1, \dots, X_n unabhängige identisch verteilte Zufallsvariablen, $X_i \stackrel{d}{=} X$, dann sind \bar{X}_n und S_n^2 unabhängig.

Dieses Lemma wird unter Anderem gebraucht, um folgendes Ergebnis zu beweisen:

Satz 3.3.6

Unter den Voraussetzungen von Lemma 3.3.2 gilt

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}.$$

Beweis von Lemma 3.3.2 Es folgt aus Lemma 2.2.1, dass

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X'_i + \mu \quad \text{und} \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X'_i - \bar{X}_n)^2$$

für $X'_i = X_i - \mu$, $i = 1, \dots, n$. Somit kann wegen des Satzes 3.2.5 o.B.d.A. $\mu = 0$ und $\sigma^2 = 1$ angenommen werden. Um die Unabhängigkeit von \bar{X}_n und S_n^2 zu zeigen, stellen wir S_n^2 in alternativer Form dar:

$$S_n^2 = \frac{1}{n-1} \left((X_1 - \bar{X}_n)^2 + \sum_{i=2}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n-1} \left(\left(\sum_{i=2}^n (X_i - \bar{X}_n) \right)^2 + \sum_{i=2}^n (X_i - \bar{X}_n)^2 \right),$$

weil $\sum_{i=1}^n (X_i - \bar{X}_n) = 0$ nach Abschnitt 2.2.1. Somit gilt

$$S_n^2 = \tilde{\varphi}(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n),$$

wobei

$$\tilde{\varphi}(x_2, \dots, x_n) = \frac{1}{n-1} \left(\left(\sum_{i=2}^n x_i \right)^2 + \sum_{i=2}^n x_i^2 \right), \quad (x_2, \dots, x_n) \in \mathbb{R}^{n-1}.$$

Es genügt (nach Satz 3.2.5) zu zeigen, dass der Zufallsvektor $(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ unabhängig von \bar{X}_n ist. Sei $X = (X_1, \dots, X_n)^T$, X_i unabhängige identisch verteilte Zufallsvariablen mit $X_i \sim N(0, 1)$ nach unserer Annahme. Dann gilt

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp \left(-\frac{1}{2} \sum_{i=1}^n x_i^2 \right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

für die Dichte von X . Sei $\varphi = (\varphi_1, \dots, \varphi_n) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ die lineare Abbildung mit

$$\begin{cases} \varphi_1(x) = \bar{x}_n, \\ \varphi_2(x) = x_2 - \bar{x}_n, \\ \vdots \\ \varphi_n(x) = x_n - \bar{x}_n, \end{cases} \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Um die Umkehrabbildung $\varphi^{-1} : (y_1, \dots, y_n) \mapsto (x_1, \dots, x_n)$ zu finden, setzen wir $y_i = \varphi_i(x)$, $i = 1, \dots, n$ und schreiben

$$\begin{cases} y_1 = \bar{x}_n \\ y_2 = x_2 - \bar{x}_n = x_2 - y_1 \\ \vdots \\ y_n = x_n - y_1 \end{cases}, \text{ woraus } \begin{cases} x_2 = y_1 + y_2 \\ \vdots \\ x_n = y_1 + y_n \\ x_2 + \dots + x_n = (n-1)y_1 + y_2 + \dots + y_n \\ x_1 + \dots + x_n = ny_1 = x_1 + (n-1)y_1 + y_2 + \dots + y_n \end{cases}$$

folgt und somit $x_1 = y_1 - \sum_{i=2}^n y_i$. Es gilt insgesamt

$$\begin{cases} \varphi_1^{-1}(y) = y_1 - \sum_{i=2}^n y_i, \\ \varphi_2^{-1}(y) = y_1 + y_2, \\ \vdots \\ \varphi_n^{-1}(y) = y_1 + y_n, \end{cases} \quad y = (y_1, \dots, y_n)^T \in \mathbb{R}^n.$$

Um den Dichtetransformationssatz 3.2.6 für $\varphi(X)$ zu verwenden, brauchen wir die Determinante der Jacobi-Matrix

$$J = \det \left(\frac{\partial \varphi_i^{-1}}{\partial y_j} \right)_{i,j=1,\dots,n} = \left| \begin{pmatrix} 1 & -1 & -1 & -1 & \dots & -1 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & 0 & 1 & 0 \\ 1 & 0 & \dots & \dots & 0 & 1 \end{pmatrix} \right|$$

$$= 1 \cdot 1 - (-1) \cdot 1 + (-1) \cdot (-1) - (-1) \cdot 1 + \dots = \underbrace{1 + \dots + 1}_n = n.$$

Somit gilt für die Dichte von $Y = \varphi(X) = (\bar{X}_n, X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$

$$\begin{aligned} f_{\varphi(Y)}(y_1, \dots, y_n) &= f_X(\varphi^{-1}(y)) \cdot |J| = \frac{n}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \left(y_1 - \sum_{i=2}^n y_i \right)^2 - \frac{1}{2} \sum_{i=2}^n (y_1 + y_i)^2 \right\} \\ &= \frac{n}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \left(y_1^2 - 2y_1 \sum_{i=2}^n y_i + \left(\sum_{i=2}^n y_i \right)^2 + \sum_{i=2}^n y_i^2 + 2y_1 \sum_{i=2}^n y_i + (n-1)y_1^2 \right) \right\} \\ &= \frac{n}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \left(ny_1^2 + \left(\sum_{i=2}^n y_i \right)^2 + \sum_{i=2}^n y_i^2 \right) \right\} \\ &= \underbrace{\left(\frac{n}{2\pi} \right)^{1/2} \exp \left\{ -\frac{1}{2} ny_1^2 \right\}}_{=f_{\varphi_1(X)}(y_1)} \cdot \underbrace{\left(\frac{n}{(2\pi)^{n-1}} \right)^{1/2} \exp \left\{ -\frac{1}{2} \left(\sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i \right)^2 \right) \right\}}_{f_{(\varphi_2(X), \dots, \varphi_n(X))}(y_2, \dots, y_n)}, \end{aligned}$$

woraus die Unabhängigkeit von

$$\varphi_1(X) = \bar{X}_n \sim N \left(\mu, \frac{\sigma^2}{n} \right) \stackrel{\mu=0, \sigma^2=1}{=} N \left(0, \frac{1}{n} \right)$$

und

$$(\varphi_2(X), \dots, \varphi_n(X)) = (X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$$

folgt. Somit sind auch \bar{X}_n und $S_n^2 = \tilde{\varphi}(X_2 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ unabhängig. \square

Beweis des Satzes 3.3.6 Aus den Sätzen 3.3.1, 4) und 3.3.5 folgt

$$\bar{X}_n \sim N \left(\mu, \frac{\sigma^2}{n} \right) \quad \text{und} \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2,$$

also

$$Y_1 = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1) \quad \text{und} \quad Y_2 = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Nach dem Lemma 3.3.2 und Satz 3.2.5 sind Y_1 und Y_2 unabhängig. Dann gilt

$$T = \frac{Y_1}{\sqrt{\frac{Y_2}{n-1}}} \sim t_{n-1}$$

nach der Definition einer t -Verteilung, wobei

$$T = \frac{\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2(n-1)}}} = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n}.$$

Somit gilt

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1}.$$

□

Bemerkung 3.3.4

Mit Hilfe des Satzes 3.3.6 kann folgendes Konfidenzintervall für den Erwartungswert μ einer normalverteilten Stichprobe (X_1, \dots, X_n) bei unbekannter Varianz σ^2 ($X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$) konstruiert werden:

$$\mathbb{P} \left(\mu \in \left[\bar{X}_n - \frac{t_{n-1, 1-\alpha/2}}{\sqrt{n}} S_n, \bar{X}_n + \frac{t_{n-1, 1-\alpha/2}}{\sqrt{n}} S_n \right] \right) = 1 - \alpha$$

für $\alpha \in (0, 1)$, denn

$$\begin{aligned} \mathbb{P} \left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \in \left[\underbrace{t_{n-1, \alpha/2}}_{=-t_{n-1, 1-\alpha/2} \text{ wg. Sym. } t\text{-Vert.}}, t_{n-1, 1-\alpha/2} \right] \right) &= F_{t_{n-1}}(t_{n-1, 1-\alpha/2}) - F_{t_{n-1}}(t_{n-1, \alpha/2}) \\ &= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha, \end{aligned} \quad (3.3.2)$$

wobei $t_{n-1, \alpha}$ das α -Quantil der t_{n-1} -Verteilung darstellt. Der Rest folgt aus (3.3.2) durch das Auflösen bzgl. μ .

3.3.4 Eigenschaften der Ordnungsstatistiken

In Abschnitt 2.2.2 haben wir bereits die Ordnungsstatistiken $x_{(1)}, \dots, x_{(n)}$ einer konkreten Stichprobe (x_1, \dots, x_n) betrachtet. Wenn wir nun auf der Modellebene arbeiten, also eine Zufallsstichprobe (X_1, \dots, X_n) von unabhängigen identisch verteilten Zufallsvariablen X_i mit Verteilungsfunktion $F(x)$ haben, welche Eigenschaften haben dann ihre Ordnungsstatistiken

$$X_{(1)}, \dots, X_{(n)}?$$

Satz 3.3.7

1. Die Verteilungsfunktion der Ordnungsstatistik $X_{(i)}$, $i = 1, \dots, n$ ist gegeben durch

$$F_{X_{(i)}}(x) = \sum_{k=i}^n \binom{n}{k} F^k(x) (1 - F(x))^{n-k}, \quad x \in \mathbb{R}. \quad (3.3.3)$$

2. Falls X_i eine diskrete Verteilung mit Wertebereich $E = \{\dots, a_{j-1}, a_j, a_{j+1}, \dots\}$ haben, $i = 1, \dots, n$, $a_i < a_j$ für $i < j$, dann gilt für die Zähldichte von $X_{(i)}$, $i = 1, \dots, n$:

$$\mathbb{P}(X_{(i)} = a_j) = \sum_{k=i}^n \binom{n}{k} \left(F^k(a_j) (1 - F(a_j))^{n-k} - F^k(a_{j-1}) (1 - F(a_{j-1}))^{n-k} \right),$$

wobei

$$F(a_j) = \sum_{a_k \in E, k \leq j} \mathbb{P}(X_i = a_k).$$

3. Falls X_i absolut stetig verteilt sind mit Dichte f , die stückweise stetig ist, dann ist auch $X_{(i)}$, $i = 1, \dots, n$ absolut stetig verteilt mit der Dichte

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) F^{i-1}(x) (1-F(x))^{n-i}, \quad x \in \mathbb{R}.$$

Beweis

1. Führen wir die Zufallsvariable

$$Y = \#\{i : X_i \leq x\} = \sum_{i=1}^n \mathbb{I}(X_i \leq x), \quad x \in \mathbb{R}$$

ein. Da X_1, \dots, X_n unabhängig identisch verteilt mit Verteilungsfunktion F sind, gilt $Y \sim \text{Bin}(n, F(x))$. Weiterhin gilt

$$F_{X_{(i)}}(x) = \mathbb{P}(X_{(i)} \leq x) = \mathbb{P}(Y \geq i) = \sum_{k=i}^n \binom{n}{k} F^k(x) (1-F(x))^{n-k}, \quad x \in \mathbb{R}.$$

2. folgt aus 1) durch

$$\mathbb{P}(X_{(i)} = a_j) = \mathbb{P}(a_{j-1} < X_{(i)} \leq a_j) = F_{X_{(i)}}(a_j) - F_{X_{(i)}}(a_{j-1}) \quad \forall j, i.$$

3. Beweisen Sie 3) als Übungsaufgabe. □

Bemerkung 3.3.5

1. Für $i = 1$ und $i = n$ sieht die Formel (3.3.3) besonders einfach aus:

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - (1 - F(x))^n, & x \in \mathbb{R} \\ F_{X_{(n)}}(x) &= F^n(x), & x \in \mathbb{R}. \end{aligned}$$

Diese Formeln lassen sich auch direkt herleiten:

$$\begin{aligned} F_{X_{(1)}}(x) &= \mathbb{P}\left(\min_{i=1, \dots, n} X_i \leq x\right) = 1 - \mathbb{P}\left(\min_{i=1, \dots, n} X_i > x\right) = 1 - \mathbb{P}(X_i \geq x, \quad \forall i = 1, \dots, n) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(X_i > x) = 1 - (1 - F(x))^n, \\ F_{X_{(n)}}(x) &= \mathbb{P}\left(\max_{i=1, \dots, n} X_i \leq x\right) = \mathbb{P}(X_i \leq x, \quad \forall i = 1, \dots, n) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) = F^n(x), \quad x \in \mathbb{R}. \end{aligned}$$

2. Falls X_i absolut stetig verteilt sind mit einer stückweise stetigen Dichte f , so lassen sich Formeln für die gemeinsame Dichte der Verteilung von $(X_{(i_1)}, \dots, X_{(i_k)})$, $i \leq k \leq n$ herleiten. Insbesondere gilt für $k = n$

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_1, \dots, x_n) = \begin{cases} n! \cdot f(x_1) \cdot \dots \cdot f(x_n), & \text{falls } -\infty < x_1 < \dots < x_n < \infty, \\ 0, & \text{sonst.} \end{cases}$$

Übungsaufgabe 3.3.1

Zeigen Sie für X_1, \dots, X_n unabhängig identisch verteilt, $X_i \sim U[0, \theta]$, $\theta > 0$, $i = 1, \dots, n$, dass

1. die Dichte von $X_{(i)}$ gleich

$$f_{X_{(i)}}(x) = \begin{cases} \frac{n!}{(i-1)!(n-i)!} \theta^{-n} x^{i-1} (\theta - x)^{n-i}, & x \in (0, \theta) \\ 0, & \text{sonst} \end{cases}$$

und

- 2.

$$\mathbb{E}X_{(i)}^k = \frac{\theta^k n! (i+k-1)!}{(n+k)!(i-1)!}, \quad k \in \mathbb{N}, \quad i = 1, \dots, n$$

sind. Insbesondere gilt $\mathbb{E}X_{(i)} = \frac{i}{n+1}\theta$ und $\text{Var} X_{(i)} = \frac{i(n-i+1)\theta^2}{(n+1)^2(n+2)}$.

3.3.5 Empirische Verteilungsfunktion

Im Folgenden betrachten wir die statistischen Eigenschaften der in Abschnitt 2.1.2 eingeführten empirischen Verteilungsfunktion $\hat{F}_n(x)$ einer Zufallsstichprobe (X_1, \dots, X_n) , wobei $X_i \stackrel{d}{=} X$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F(\cdot)$ sind.

Satz 3.3.8

Es gilt

1. $n\hat{F}_n(x) \sim \text{Bin}(n, F(x))$, $x \in \mathbb{R}$.
2. $\hat{F}_n(x)$ ist ein erwartungstreuer Schätzer für $F(x)$, $x \in \mathbb{R}$ mit

$$\text{Var} \hat{F}_n(x) = \frac{F(x)(1-F(x))}{n}.$$

3. $\hat{F}_n(x)$ ist stark konsistent.
4. $\hat{F}_n(x)$ ist asymptotisch normalverteilt:

$$\sqrt{n} \frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \xrightarrow{d} Y \sim N(0, 1), \quad \forall x : F(x) \in (0, 1).$$

Beweis 1. folgt aus der Darstellung

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x), \quad x \in \mathbb{R},$$

weil $\mathbb{I}(X_i \leq x) \sim \text{Bernoulli}(F(x))$, $\forall i = 1, \dots, n$. Somit ist

$$\sum_{i=1}^n \mathbb{I}(X_i \leq x) \sim \text{Bin}(n, F(x)).$$

2. Es folgt aus 1)

$$\begin{cases} \mathbb{E}(n\hat{F}_n(x)) = nF(x), & x \in \mathbb{R}, \\ \text{Var}(n\hat{F}_n(x)) = nF(x) \cdot (1 - F(x)), & x \in \mathbb{R}, \end{cases}$$

woraus $\mathbb{E}\hat{F}_n(x) = F(x)$ und $\text{Var}\hat{F}_n(x) = F(x)(1 - F(x))/n$ folgen.

3. Da $Y_i = \mathbb{I}(X_i \leq x)$, $i = 1, \dots, n$, $x \in \mathbb{R}$ unabhängige identisch verteilte Zufallsvariablen sind, gilt nach dem starken Gesetz der großen Zahlen

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}Y_i = F(x).$$

4. folgt aus der Anwendung des zentralen Grenzwertsatzes auf die oben genannte Folge $\{Y_i\}_{i \in \mathbb{N}}$. □

In Satz 3.3.8, 3) wird behauptet, dass

$$\hat{F}_n(x) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} F(x), \quad \forall x \in \mathbb{R}.$$

Der nachfolgende Satz von Gliwenko-Cantelli behauptet, dass diese Konvergenz gleichmäßig in $x \in \mathbb{R}$ stattfindet. Um diesen Satz formulieren zu können, betrachten wir den *gleichmäßigen Abstand* zwischen \hat{F}_n und F

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|.$$

Dieser Abstand ist eine Zufallsvariable, die auch *Kolmogorow-Abstand* genannt wird. Er gibt den maximalen Fehler an, den man bei der Schätzung von $F(x)$ durch $\hat{F}_n(x)$ macht.

Übungsaufgabe 3.3.2

Zeigen Sie, dass

$$D_n = \max_{i \in \{1, \dots, n\}} \max \left\{ F(X_{(i)} - 0) - \frac{i-1}{n}, \frac{i}{n} - F(X_{(i)}) \right\}. \quad (3.3.4)$$

Beachten Sie dabei die Tatsache, dass $\hat{F}_n(x)$ eine Treppenfunktion mit Sprungstellen $X_{(i)}$, $i = 1, \dots, n$ ist.

Satz 3.3.9 (Gliwenko-Cantelli):

Es gilt $D_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} 0$.

Beweis Für alle $m \in \mathbb{N}$ wähle beliebige Zahlen $-\infty = z_0 < z_1 < \dots < z_{m-1} < z_m = \infty$. Dann gilt

$$D_n = \sup_{z \in \mathbb{R}} |\hat{F}_n(z) - F(z)| = \sup_{j=0, \dots, m-1} \sup_{z \in [z_j, z_{j+1})} |\hat{F}_n(z) - F(z)|.$$

Zeigen wir, dass $\forall m \in \mathbb{N}$ z_0, \dots, z_m existieren, für die gilt

$$F(z_{j+1} - 0) - F(z_j) \leq \varepsilon = \frac{1}{m}. \quad (3.3.5)$$

Falls F stetig ist, genügt es, $z_j = F^{-1}(j/m)$, $j = 1 \dots m-1$ gleichzusetzen. Im allgemeinen Fall existieren $n < m/2$ Punkte x_j mit der Eigenschaft

$$F(x_j) - F(x_j - 0) > 2\varepsilon = 2/m$$

(weil $n \cdot 2\varepsilon \leq 1$ sein muss) und $k+1$ Punkte y_j zwischen diesen Punkten x_j mit Eigenschaft (3.3.5), wobei für k gilt:

$$n \cdot 2\varepsilon + (k+1)\varepsilon \leq 1 \implies 2n + k + 1 \leq m \implies k \leq m - 2n - 1.$$

Setzen wir $\{z_j\} = \{x_j\} \cup \{y_j\}$. Für alle $z \in [z_j, z_{j+1})$ gilt

$$\hat{F}_n(z) - F(z) \leq \hat{F}_n(z_{j+1} - 0) - F(z_j) \leq \hat{F}_n(z_{j+1} - 0) - F(z_{j+1} - 0) + \varepsilon,$$

weil aus (3.3.5) folgt, dass $-F(z_j) \leq \varepsilon - F(z_{j+1} - 0)$, $\forall j$.

Genauso gilt

$$\hat{F}_n(z) - F(z) \geq \hat{F}_n(z_j) - F(z_{j+1} - 0) \geq \hat{F}_n(z_j) - F(z_j) - \varepsilon,$$

weil aus (3.3.5) für alle j folgt, dass $-F(z_{j+1} - 0) \geq -F(z_j) - \varepsilon$ gilt. Für alle $m \in \mathbb{N}$, $j \in \{0, 1, \dots, m\}$ sei

$$A_{m,j} = \{\omega \in \Omega : \lim_{n \rightarrow \infty} \hat{F}_n(z_j) = F(z_j)\},$$

$$A'_{m,j} = \{\omega \in \Omega : \lim_{n \rightarrow \infty} \hat{F}_n(z_j - 0) = F(z_j - 0)\}.$$

Nach dem Satz 3.3.8, 3) gilt $\mathbb{P}(A_{m,j}) = 1$. Um $\mathbb{P}(A'_{m,j}) = 1$ zu zeigen, kann man die Verallgemeinerung von Aussage 3.3.8, 3) auf das Maß

$$\hat{F}_n(B) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \in B), \quad B \in \mathcal{B}_{\mathbb{R}}$$

benutzen: nach dem starken Gesetz der großen Zahlen gilt nämlich

$$\hat{F}_n(B) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} F(B) = \mathbb{P}(X \in B), \quad B \in \mathcal{B}_{\mathbb{R}}.$$

Da $(-\infty, z_j) \in \mathcal{B}_{\mathbb{R}} \quad \forall j$, ist $\mathbb{P}(A'_{m,j}) = 1$ bewiesen $\forall m \forall j$. Für

$$A'_m = \bigcap_{j=0}^m (A_{m,j} \cap A'_{m,j})$$

gilt $\mathbb{P}(A'_m) = 1 \quad \forall m$, weil

$$\mathbb{P}(A'_m) = 1 - \mathbb{P}(\bar{A}'_m) = 1 - \mathbb{P}\left(\bigcup_{j=0}^m (\bar{A}_{m,j} \cup \bar{A}'_{m,j})\right) \geq 1 - \sum_{j=0}^m \left(\underbrace{\mathbb{P}(\bar{A}_{m,j})}_{=0} + \underbrace{\mathbb{P}(\bar{A}'_{m,j})}_{=0}\right) = 1.$$

Weiterhin: für $\varepsilon = 1/m \forall \omega \in A'_m \exists n(\omega, m) : \forall n > n(\omega, m) \forall j \in \{0, \dots, m-1\} \forall z \in [z_j, z_{j+1})$

$$\left. \begin{aligned} \hat{F}_n(z) - F(z) &\leq \underbrace{\hat{F}_n(z_{j+1} - 0) - F(z_{j+1} - 0)}_{< \varepsilon \text{ aus } A'_{m,j}} + \varepsilon < 2\varepsilon, \\ \hat{F}_n(z) - F(z) &\geq \underbrace{\hat{F}_n(z_j) - F(z_j)}_{> -\varepsilon \text{ aus } A_{m,j}} - \varepsilon > -2\varepsilon, \end{aligned} \right\} \implies |\hat{F}_n(z) - F(z)| < 2\varepsilon.$$

$$\implies D_n = \sup_{j=0, \dots, m-1} \sup_{z \in [z_j, z_{j+1})} |\hat{F}_n(z) - F(z)| < 2\varepsilon.$$

Nun wählen wir ein beliebiges $m \in \mathbb{N}$ und betrachten $A' = \bigcap_{m=1}^{\infty} A'_m$. Es folgt, dass $\mathbb{P}(A') = 1$ und $\forall \omega \in A' \exists n_0 : \forall n \geq n_0$

$$D_n < 2\varepsilon = \frac{2}{m} \quad \forall m \in \mathbb{N} \quad \implies \quad D_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} 0.$$

□

Satz 3.3.10 (Ungleichung von Dvoretzky-Kiefer-Wolfowitz):

Seien X_1, \dots, X_n unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion F . Für alle $\varepsilon > 0$ gilt

$$\mathbb{P}(D_n > \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

(ohne Beweis)

Folgerung 3.3.3 (Konfidenzband für F):

Führen wir Statistiken

$$L(x) = \max\{\hat{F}_n(x) - \varepsilon_n, 0\} \text{ und } U(x) = \min\{\hat{F}_n(x) + \varepsilon_n, 1\}, \quad \varepsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}, \quad \alpha \in (0, 1)$$

ein. Dann gilt

$$\mathbb{P}(L(x) \leq F(x) \leq U(x) \quad \forall x \in \mathbb{R}) \geq 1 - \alpha \quad (3.3.6)$$

Beweis Beweisen Sie dieses Korollar als Übungsaufgabe! □

Bemerkung 3.3.6

Das simultane Konfidenzintervall $\{L(x) \leq F(x) \leq U(x), \quad x \in \mathbb{R}\}$ aus (3.3.6) heißt *Konfidenzband* für F zum Konfidenzniveau $1 - \alpha$ (vgl. Abb. 3.5).

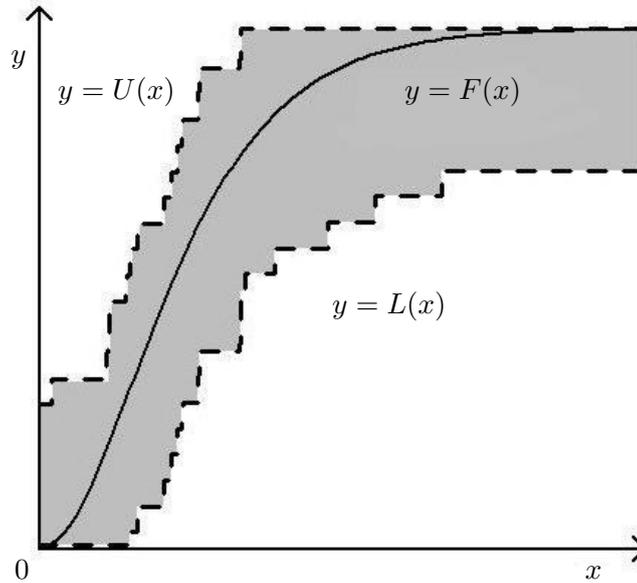
Falls die Verteilungsfunktion F stetig ist, kann man zeigen, dass die Zufallsvariable D_n nicht von F abhängt, also *verteilungsfrei* ist.

Satz 3.3.11

Für jede stetige Verteilungsfunktion F gilt

$$D_n \stackrel{d}{=} \sup_{y \in [0,1]} |\hat{G}_n(y) - y|, \text{ wobei } \hat{G}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(Y_i \leq y), \quad y \in \mathbb{R}$$

die empirische Verteilungsfunktion der Zufallsstichprobe (Y_1, \dots, Y_n) mit unabhängigen identisch verteilten Zufallsvariablen $Y_i \sim U[0, 1]$, $i = 1, \dots, n$ ist.

Abb. 3.5: Konfidenzband für F .

Beweis Zunächst definieren wir einen sogenannten *Konstanzbereich* $(a, b] \subset \mathbb{R}$ einer Verteilungsfunktion F als maximales Intervall mit der Eigenschaft $F(a) = F(b)$. Sei B die Vereinigung aller Konstanzbereiche von F . Auf B^C ist F eine monoton steigende eineindeutige Funktion. Damit folgt die Existenz ihrer Inversen $F^{-1} : (0, 1) \rightarrow B^C$. Gleichzeitig gilt

$$D_n = \sup_{x \in B^C} |\hat{F}_n(x) - F(x)|.$$

Führen wir $Y_i = F(X_i)$, $i = 1, \dots, n$ ein. Y_i sind unabhängig identisch verteilt und $Y_i \sim U[0, 1]$, denn

$$\mathbb{P}(Y_i \leq y) = \mathbb{P}(F(X_i) \leq y) = \mathbb{P}(X_i \leq F^{-1}(y)) = F(F^{-1}(y)) = y, \quad y \in (0, 1).$$

Somit gilt auch

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(\underbrace{F(X_i)}_{Y_i} \leq F(x)) = \hat{G}_n(F(x)), \quad x \in B^C.$$

Hieraus folgt

$$\begin{aligned} D_n &= \sup_{x \in B^C} |\hat{F}_n(x) - F(x)| = \sup_{x \in B^C} |\hat{G}_n(F(x)) - F(x)| = \sup_{x \in \mathbb{R}} |\hat{G}_n(F(x)) - F(x)| \\ &= \sup_{y \in [0, 1]} |\hat{G}_n(y) - y|, \end{aligned}$$

wobei die letzte Gleichheit die Stetigkeit von F ausnützt. □

Folgerung 3.3.4

Falls F eine stetige Verteilungsfunktion ist, dann gilt

$$D_n \stackrel{d}{=} \max_{i=1, \dots, n} \max \left\{ Y_{(i)} - \frac{i-1}{n}, \frac{i}{n} - Y_{(i)} \right\},$$

wobei $Y_{(1)}, \dots, Y_{(n)}$ die Ordnungsstatistiken der auf $[0, 1]$ gleichverteilten Stichprobenvariablen Y_1, \dots, Y_n sind.

Beweis Benutze dazu die Darstellung (3.3.4), den Satz 3.3.11 sowie die Tatsache, dass

$$F(x) = x, \quad x \in [0, 1]$$

für die Verteilungsfunktion der $U[0, 1]$ -Verteilung ist. \square

Folgende Ergebnisse werden ohne Beweis angegeben:

Bemerkung 3.3.7

1. Für die Zwecke des statistischen Testens (vgl. den Anpassungstest von Kolmogorow-Smirnow, Bemerkung 3.3.8, 3)) ist es notwendig, die Quantile der Verteilung von D_n zu nennen. Auf Grund der Komplexität der Verteilung von D_n ist es jedoch unmöglich, sie explizit anzugeben. Mit Hilfe des Satzes 3.3.11 ist es möglich, diese Quantile durch Monte-Carlo-Simulationen numerisch zu berechnen. Dazu simuliert man mehrere Stichproben (Y_1, \dots, Y_n) von $U[0, 1]$ -verteilten Pseudozufallszahlen, bildet $\hat{G}_n(x)$ und berechnet D_n nach Folgerung 3.3.4.
2. Für stetige Verteilungsfunktionen F kann folgende Integraldarstellung von Verteilungsfunktion von D_n bewiesen werden:

$$\mathbb{P} \left(D_n \leq x + \frac{1}{2n} \right) = \begin{cases} 0, & x \leq 0, \\ \int_{\frac{1}{2n}-x}^{\frac{1}{2n}+x} \int_{\frac{3}{2n}-x}^{\frac{3}{2n}+x} \dots \int_{\frac{2n-1}{2n}-x}^{\frac{2n-1}{2n}+x} g(y_1, \dots, y_n) dy_n \dots dy_1, & 0 < x < \frac{2n-1}{2n}, \\ 1, & x \geq \frac{2n-1}{2n}. \end{cases}$$

wobei

$$g(y_1, \dots, y_n) = \begin{cases} n!, & 0 < y_1 < \dots < y_n < 1, \\ 0, & \text{sonst} \end{cases}$$

die Dichte der Ordnungsstatistiken $(Y_{(1)}, \dots, Y_{(n)})$ von $U[0, 1]$ -verteilten Stichprobenvariablen (Y_1, \dots, Y_n) sind.

Satz 3.3.12 (Kolmogorow):

Falls die Verteilungsfunktion F der unabhängigen und identisch verteilten Stichprobenvariablen X_i , $i = 1, \dots, n$ stetig ist, dann gilt

$$\sqrt{n}D_n \xrightarrow[n \rightarrow \infty]{d} Y,$$

wobei Y eine Zufallsvariable mit der Verteilungsfunktion

$$K(x) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 x^2} = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}, & x > 0, \\ 0, & \text{sonst} \end{cases}$$

(Kolmogorow-Verteilung) ist.

Bemerkung 3.3.8

1. Die Verteilung von Kolmogorow ist die Verteilung des Maximums einer Brownschen Brücke, denn es gilt

$$Y \stackrel{d}{=} \sup_{t \in [0,1]} |w^\circ(t)|,$$

wobei $\{w^\circ(t), t \in [0,1]\}$ ein stochastischer Prozess ist, der die *Brownsche Brücke* genannt wird. Er wird als $w^\circ(t) = w(t) - w(1)t$, $t \in [0,1]$ definiert, wobei $\{w(t), t \in [0,1]\}$ die Brownsche Bewegung ist (für die unter anderem $w(t) \sim N(0,t)$ gilt). Der Name „Brücke“ ist der Tatsache $w^\circ(0) = w^\circ(1) = 0$ zu verdanken.

2. Aus Satz 3.3.12 folgt

$$\mathbb{P}(\sqrt{n}D_n \leq x) \underset{n \rightarrow \infty}{\approx} K(x), \quad x \in \mathbb{R}.$$

Die daraus resultierende Näherungsformel

$$\mathbb{P}(D_n \leq x) \approx K(x\sqrt{n})$$

ist ab $n > 40$ praktisch brauchbar.

3. *Kolmogorow-Smirnow-Anpassungstest*: Mit Hilfe der Aussage des Satzes 3.3.12 ist es möglich, folgenden *asymptotischen Anpassungstest von Komogorow-Smirnow* zu entwickeln. Es wird die Haupthypothese $H_0 : F = F_0$ (die unbekannte Verteilungsfunktion der Stichprobenvariablen X_1, \dots, X_n ist gleich F_0) gegen die Alternative $H_1 : F \neq F_0$ getestet. Dabei wird H_0 verworfen, falls

$$\sqrt{n}D_n \notin [k_{\alpha/2}, k_{1-\alpha/2}]$$

ist, wobei

$$D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)|$$

und k_α das α -Quantil der Kolmogorow-Verteilung ist. Somit ist die Wahrscheinlichkeit, die richtige Hypothese H_0 zu verwerfen (Wahrscheinlichkeit des *Fehlers 1. Art*) asymptotisch gleich

$$\mathbb{P}(\sqrt{n}D_n \notin [k_{\alpha/2}, k_{1-\alpha/2}] | H_0) \xrightarrow{n \rightarrow \infty} 1 - K(k_{1-\alpha/2}) + K(k_{\alpha/2}) = 1 - (1 - \alpha/2) + \alpha/2 = \alpha.$$

In der Praxis wird α klein gewählt, z.B. $\alpha \approx 0,05$. Somit ist im Fall, dass H_0 stimmt, die Wahrscheinlichkeit einer Fehlentscheidung in Folge des Testens klein.

Dieser Test ist nur ein Beispiel dessen, wie der Satz von Kolmogorow in der statistischen Testtheorie verwendet wird. Die allgemeine Philosophie des Testens wird in Statistik II erläutert.

Mit Hilfe von \hat{F}_n lassen sich sehr viele Schätzer durch die sogenannte *Plug-in-Methode* konstruieren. Dies werden wir jetzt näher erläutern: Sei $M = \{\text{Menge aller Verteilungsfunktionen}\}$.

Definition 3.3.9

Sei ein Parameter θ der Verteilungsfunktion F als Funktional $T : M \rightarrow \mathbb{R}$ von F gegeben: $\theta = T(F)$. Dann heißt $\hat{\theta} = T(\hat{F}_n)$ der *Plug-in-Schätzer* für θ .

Definition 3.3.10

Sei F eine beliebige Verteilungsfunktion. Das Funktional $T : M \rightarrow \mathbb{R}$ heißt *linear*, falls

$$T(aF_1 + bF_2) = aT(F_1) + bT(F_2) \quad \forall a, b \in \mathbb{R}, \quad F_1, F_2 \in M.$$

Betrachten wir eine spezielle Klasse der linearen Funktionalen

$$T(F) = \int_{\mathbb{R}} r(x) dF(x),$$

wobei $r(x)$ eine beliebige stetige Funktion ist. Beispiele für solche T sind

$$\mathbb{E}X^k = \int_{\mathbb{R}} x^k dF(x), \quad k \in \mathbb{N}.$$

Lemma 3.3.3

Der Plug-in Schätzer für $\theta = \int_{\mathbb{R}} r(x) dF(x)$ ist durch

$$\hat{\theta} = \int_{\mathbb{R}} r(x) d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n r(x_i)$$

gegeben.

Übungsaufgabe 3.3.3

Beweisen Sie Lemma 3.3.3!

Beispiel 3.3.1 (Plug-in-Schätzer):

1. \bar{X}_n ist ein Plug-in Schätzer für den Erwartungswert μ .
2. *Plug-in Schätzer* für $\sigma^2 = \text{Var } X$: Es gilt $\text{Var } X = \mathbb{E}X^2 - (\mathbb{E}X)^2$ und somit folgt

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2.$$

3. *Schätzer für Schiefe und Wölbung* $\hat{\gamma}_1$ und $\hat{\gamma}_2$ (vgl. Abschnitt 2.2.4) sind Plug-in Schätzer: Da der Koeffizient der Schiefe als

$$\gamma_1 = \mathbb{E} \left(\frac{X - \mu}{\sigma} \right)^3$$

definiert ist, wobei $\mu = \mathbb{E}X$, $\sigma^2 = \text{Var } X$, folgt

$$\hat{\gamma}_1 \stackrel{\mu \mapsto \bar{X}_n}{\sigma^2 \mapsto \hat{\sigma}^2} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^3}{(\hat{\sigma}_n^2)^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^3}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^{3/2}}.$$

Die Konstruktion von $\hat{\gamma}_2$ erfolgt analog.

4. Der *empirische Korrelationskoeffizient* ϱ_{XY} ist ein Plug-in Schätzer:

$$\hat{\varrho}_{XY} = \frac{S_{XY}^2}{\sqrt{S_{XX}^2} \sqrt{S_{YY}^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}};$$

in der Tat ist

$$\varrho_{XY} = \frac{\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{\sqrt{\text{Var } X \cdot \text{Var } Y}} = \frac{\mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y}{\sqrt{(\mathbb{E}X^2 - (\mathbb{E}X)^2)(\mathbb{E}Y^2 - (\mathbb{E}Y)^2)}}$$

und somit gilt für die linearen Funktionale

$$T_1(F) = \int x dF(x), \quad T_2(F) = \int x^2 dF(x), \quad T_{12}(F, G) = \int xy dF(x)dG(y),$$

$$\varrho_{XY} = \frac{T_{12}(F_X, F_Y) - T_1(F_X) \cdot T_1(F_Y)}{\sqrt{(T_2(F_X) - (T_1(F_X))^2)(T_2(F_Y) - (T_1(F_Y))^2)}}.$$

$\hat{\varrho}_{XY}$ bekommt man, in dem man T_1 , T_2 und T_{12} durch Plug-in Schätzer ersetzt:

$$\hat{\varrho}_{XY} = \frac{T_{12}(\hat{F}_{n,X}, \hat{F}_{n,Y}) - T_1(\hat{F}_{n,X}) \cdot T_1(\hat{F}_{n,Y})}{\sqrt{(T_2(\hat{F}_{n,X}) - (T_1(\hat{F}_{n,X}))^2)(T_2(\hat{F}_{n,Y}) - (T_1(\hat{F}_{n,Y}))^2)}}.$$

3.4 Methoden zur Gewinnung von Punktschätzern

Sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen identisch verteilten Zufallsvariablen X_i mit Verteilungsfunktion $F \in \{F_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$ (Parametrisches Modell). Sei die Parametrisierung $\theta \mapsto F_\theta$ unterscheidbar, d.h. $F_\theta \neq F_{\theta'} \iff \theta \neq \theta'$.

Zielstellung: Konstruiere einen Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für $\theta = (\theta_1, \dots, \theta_m)$.

3.4.1 Momentenschätzer

Aus der Wahrscheinlichkeitsrechnung (Satz 4.8) folgt, dass unter gewissen Voraussetzungen (z.B. Gleichverteilung auf einem kompakten Intervall) an die Verteilung F diese Verteilung aus der Kenntnis von Momenten $\mathbb{E}X^k$, $k \in \mathbb{N}$ wiedergewonnen werden kann. Auf dieser Idee der Schätzung von F aus den Momenten basiert die von Karl Pearson am Ende des XIX. Jh. vorgeschlagene *Momentenmethode*.

Annahme: Es existiert ein $r \geq m$, so dass $\mathbb{E}_\theta |X_i|^r < \infty$. Seien die Momente $\mathbb{E}_\theta X_i^k = g_k(\theta)$, $k = 1, \dots, r$ als Funktionen des Parametervektors $\theta = (\theta_1, \dots, \theta_m) \in \Theta$ gegeben.

Momenten-Gleichungssystem: $\hat{\mu}_k = g_k(\theta)$, $k = 1, \dots, r$, wobei $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ die k -ten empirischen Momente sind.

Definition 3.4.1

Falls das obige Gleichungssystem eindeutig lösbar bzgl. θ ist, so heißt die Lösung $\hat{\theta}(X_1, \dots, X_n)$ *Momentenschätzer (M-Schätzer)* von θ .

Lemma 3.4.1

Falls die Funktion $g = (g_1, \dots, g_r) : \Theta \rightarrow C \subset \mathbb{R}^r$ eineindeutig und ihre Inverse $g^{-1} : C \rightarrow \Theta$ stetig ist, dann ist der M-Schätzer $\hat{\theta}(X_1, \dots, X_n)$ von θ stark konsistent.

Beweis Es gilt $\hat{\theta}(X_1, \dots, X_n) = g^{-1}(\hat{\mu}_1, \dots, \hat{\mu}_r) \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \theta$, weil $\hat{\mu}_k \xrightarrow[n \rightarrow \infty]{\text{f.s.}} g_k(\theta)$, $k = 1, \dots, r$ (starke Konsistenz der empirischen Momente) und g^{-1} stetig. \square

Bemerkung 3.4.1

1. Unter gewissen Regularitätsbedingungen an F_θ ist der M-Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ asymptotisch normalverteilt:

$$\sqrt{n} \left(\hat{\theta}(X_1, \dots, X_n) - \theta \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \Sigma),$$

wobei $N(0, \Sigma)$ die multivariate Normalverteilung mit Kovarianzmatrix

$$\Sigma = G \mathbb{E}(YY^T) G^T \quad \text{ist mit} \quad Y = (X, X^2, \dots, X^r), \quad X \stackrel{d}{=} X_i,$$

und

$$G = \left(\frac{\partial g_i^{-1}}{\partial \theta_j} \right)_{\substack{i=1 \dots r, \\ j=1 \dots m}}.$$

2. Andere Eigenschaften gelten für M-Schätzer im Allgemeinen nicht. Zum Beispiel sind nicht alle M-Schätzer erwartungstreu (vgl. Beispiel 3.4.1, 1)).
3. Manchmal sind $r > m$ Gleichungen im Momentensystem notwendig, um einen M-Schätzer zu bekommen. Dies ist zum Beispiel dann der Fall, wenn manche Funktionen $g_i = \text{const}$ sind, d.h. sie enthalten keine Information über θ (vgl. Beispiel 3.4.1, 2)).

Beispiel 3.4.1

1. *Normalverteilung:* $X_i \stackrel{d}{=} X$, $i = 1, \dots, n$, $X \sim N(\mu, \sigma^2)$; Gesucht ist ein M-Schätzer für μ und σ^2 , also $\theta = (\mu, \sigma^2)$. Es gilt

$$\begin{aligned} g_1(\mu, \sigma^2) &= \mathbb{E}_\theta X = \mu, \\ g_2(\mu, \sigma^2) &= \mathbb{E}_\theta X^2 = \text{Var}_\theta X + (\mathbb{E}_\theta X)^2 = \sigma^2 + \mu^2. \end{aligned}$$

Somit ergibt sich das Gleichungssystem

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = \mu, \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \mu^2 + \sigma^2. \end{cases}$$

Damit folgt

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i^2 - \bar{X}_n^2) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2. \end{aligned}$$

Das heißt, das die M-Schätzer $\hat{\mu} = \bar{X}_n$, $\hat{\sigma}^2 = \frac{n-1}{n} S_n^2$ sind. Dabei ist $\hat{\sigma}^2$ nicht erwartungstreu:

$$\mathbb{E}_\theta \hat{\sigma}^2 = \frac{n-1}{n} \cdot \mathbb{E}_\theta S_n^2 = \frac{n-1}{n} \sigma^2.$$

2. Gleichverteilung: $X_i \stackrel{d}{=} X$, $i = 1, \dots, n$, $X \sim U[-\theta, \theta]$, $\theta > 0$. Gesucht ist ein Momentenschätzer für θ . Es gilt

$$g_1(\theta) = \mathbb{E}_\theta X = 0,$$

$$g_2(\theta) = \mathbb{E}_\theta X^2 = \text{Var}_\theta X = \frac{(\theta - (-\theta))^2}{12} = \frac{(2\theta)^2}{12} = \frac{\theta^2}{3}.$$

Damit ergibt sich das Gleichungssystem

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = 0 & \text{unbrauchbar,} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\theta^2}{3}. \end{cases}$$

Es folgt, dass $\hat{\theta} = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}$ der Momentenschätzer für θ ist. Wir haben somit 2 Gleichungen für die Schätzung eines einzigen Parameters θ benötigt, d.h. $r = 2 > m = 1$.

3.4.2 Maximum-Likelihood-Schätzer

Diese wurden von Carl Friedrich Gauss (Anfang des XIX. Jh.) und Sir Ronald Fisher (1922) entdeckt. Seien entweder alle Verteilungen aus der parametrischen Familie $\{F_\theta : \theta \in \Theta\}$ diskret oder alle absolut stetig.

Definition 3.4.2

1. Falls die Stichprobenvariablen X_i , $i = 1, \dots, n$ absolut stetig verteilt mit Dichte $f_\theta(x)$ sind, dann heißt

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i), \quad (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \theta \in \Theta$$

die *Likelihood-Funktion* der Stichprobe (x_1, \dots, x_n) .

2. Falls die Stichprobenvariablen X_i , $i = 1, \dots, n$ diskret verteilt mit Zähldichte $p_\theta(x) = \mathbb{P}_\theta(X_i = x)$, $x \in C$ sind (C ist der Wertebereich von X), dann heißt

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n p_\theta(x_i), \quad (x_1, \dots, x_n) \in C^n, \quad \theta \in \Theta$$

die *Likelihood-Funktion* der Stichprobe (x_1, \dots, x_n) .

Nach dieser Definition gilt im

- *diskreten Fall* $L(x_1, \dots, x_n, \theta) = \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n)$
- *absolut stetigen Fall*

$$L(x_1, \dots, x_n, \theta) \Delta x_1 \cdot \dots \cdot \Delta x_n = f_{(X_1, \dots, X_n), \theta}(x_1, \dots, x_n) \Delta x_1 \cdot \dots \cdot \Delta x_n$$

$$\approx \mathbb{P}_\theta(X_1 \in [x_1, x_1 + \Delta x_1], \dots, X_n \in [x_n, x_n + \Delta x_n]), \quad \Delta x_i \rightarrow 0, \quad i = 1, \dots, n.$$

Nun wird ein Schätzer für θ so gewählt, dass die Wahrscheinlichkeit

$$\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n) \quad \text{bzw.} \quad \mathbb{P}_\theta(X_i \in [x_i, x_i + \Delta x_i], \quad i = 1, \dots, n)$$

maximal wird. \implies Maximum-Likelihoodmethode:

Definition 3.4.3

Sei das Maximierungsproblem $L(x_1, \dots, x_n, \theta) \mapsto \max_{\theta \in \Theta}$ eindeutig lösbar. Dann heißt

$$\hat{\theta}(x_1, \dots, x_n) = \arg \max_{\theta \in \Theta} L(x_1, \dots, x_n, \theta)$$

der *Maximum-Likelihood-Schätzer* von θ (*ML-Schätzer*).

Bemerkung 3.4.2

1. In relativ wenigen Fällen ist ein ML-Schätzer $\hat{\theta}$ für θ explizit auffindbar. In diesen Fällen wird meistens der konstante Faktor von $L(x_1, \dots, x_n, \theta)$ weggeworfen und vom Rest der Logarithmus gebildet:

$$\log L(x_1, \dots, x_n, \theta) \quad (\text{die sog. Loglikelihood-Funktion}).$$

Dadurch wird

$$\prod_{i=1}^n f_\theta(x_i) \quad \text{bzw.} \quad \prod_{i=1}^n p_\theta(x_i)$$

zu einer Summe

$$\sum_{i=1}^n \log f_\theta(x_i) \quad \text{bzw.} \quad \sum_{i=1}^n \log p_\theta(x_i),$$

die leichter bzgl. θ zu differenzieren ist. Danach betrachtet man

$$\frac{\partial \log L(x_1, \dots, x_n, \theta)}{\partial \theta_j} = 0, \quad j = 1 \dots m.$$

Dies ist die notwendige Bedingung eines Extremums von $\log L$ (und somit von L , weil $\log \nearrow$). Falls dieses System eindeutig lösbar ist, und die Lösung eine Maximum-Stelle ist, dann wird sie zum ML-Schätzer $\hat{\theta}(X_1, \dots, X_n)$ erklärt.

2. In den meisten praxisrelevanten Fällen sind ML-Schätzer jedoch nur numerisch auffindbar.

Beispiel 3.4.2

1. *Bernoulli-Verteilung*: $X_i \stackrel{d}{=} X$, $i = 1, \dots, n$, $X \sim \text{Bernoulli}(p)$, für ein $p \in [0, 1]$. Da

$$X = \begin{cases} 1, & \text{mit Wkt. } p \\ 0, & \text{sonst} \end{cases}$$

mit Zähldichte

$$p_\theta(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\},$$

ist die *Likelihood-Funktion* der Stichprobe (X_1, \dots, X_n) gegeben durch

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \stackrel{\text{def.}}{=} h(p).$$

- a) Falls $\sum_{i=1}^n x_i = 0$ ($\Leftrightarrow x_1 = x_2 = \dots = x_n = 0$), es folgt $h(p) = (1-p)^n \rightarrow \max_{p \in [0,1]}$ bei $p = 0$. Dann ist der ML-Schätzer $\hat{p}(0, \dots, 0) = 0$.
- b) Falls $\sum_{i=1}^n x_i = n$ ($\Leftrightarrow x_1 = x_2 = \dots = x_n = 1$), es folgt $h(p) = p^n \rightarrow \max_{p \in [0,1]}$ bei $p = 1$. Dann ist der ML-Schätzer $\hat{p}(1, 1, \dots, 1) = 1$.
- c) Falls $0 < \sum_{i=1}^n x_i < n$, dann gilt

$$\log L(x_1, \dots, x_n, p) = n\bar{x}_n \log p + n(1 - \bar{x}_n) \log(1 - p) = n \cdot g(p).$$

Da $g(p) \xrightarrow{p \rightarrow 0,1} -\infty$ und

$$\frac{\partial g(p)}{\partial p} = \frac{\bar{x}_n}{p} + \frac{1 - \bar{x}_n}{1 - p} \cdot (-1) = \frac{\bar{x}_n}{p} + \frac{\bar{x}_n - 1}{1 - p} = 0$$

$$\Leftrightarrow (1 - p)\bar{x}_n + (\bar{x}_n - 1)p = 0 \quad \Rightarrow \quad p = \bar{x}_n,$$

folgt aufgrund der Stetigkeit von g , dass g genau ein Extremum $\arg \max_p g(p) = \bar{x}_n$ besitzt.

Der ML-Schätzer ist also gegeben durch $\hat{p}(X_1, \dots, X_n) = \bar{X}_n$.

2. *Gleichverteilung*: $X \sim U[0, \theta]$, $\theta > 0$, (X_1, \dots, X_n) unabhängig identisch verteilt, gesucht ist ein ML-Schätzer für θ . Es gilt

$$f_{X_i}(x) = 1/\theta \cdot \mathbb{I}(x \in [0, \theta]), \quad i = 1, \dots, n.$$

Somit ist die Likelihood-Funktion durch

$$L(x_1, \dots, x_n, \theta) = \begin{cases} (1/\theta)^n, & 0 \leq x_1, \dots, x_n \leq \theta \\ 0, & \text{sonst} \end{cases}$$

$$= \begin{cases} (1/\theta)^n, & \text{falls } \min\{x_1, \dots, x_n\} \geq 0, \quad \max\{x_1, \dots, x_n\} \leq \theta \\ 0, & \text{sonst} \end{cases}$$

$$= g(\theta), \quad \theta > 0$$

gegeben. Damit folgt $\hat{\theta} = \arg \max_{\theta > 0} g(\theta) = \max\{x_1, \dots, x_n\} = x_{(n)}$, wodurch der ML-

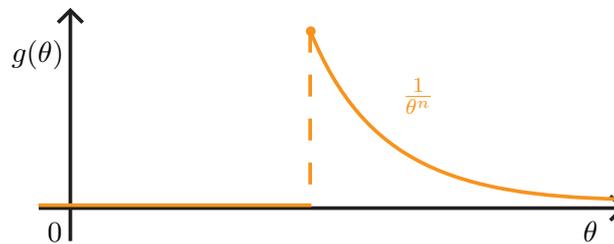


Abb. 3.6: Illustration der Funktion g .

Schätzer durch $\hat{\theta}(X_1, \dots, X_n) = X_{(n)}$ gegeben ist.

Nun wollen wir zeigen, dass ML-Schätzer unter gewissen Voraussetzungen schwach konsistent und asymptotisch normalverteilt sind.

Definition 3.4.4

Sei

$$L(x, \theta) = \begin{cases} f_\theta(x), & \text{im absolut stetigen Fall,} \\ p_\theta(x), & \text{im diskreten Fall} \end{cases}$$

die Likelihood-Funktion von x . Für $\theta, \theta' \in \Theta$ und $X \stackrel{d}{=} X_i$, $\mathbb{P}_\theta(L(X, \theta') = 0) = 0$ definieren wir die *Information* (Abstand) $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'})$ von *Kullback-Leibler* im absolut stetigen Fall als

$$H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \mathbb{E}_\theta \log L(X, \theta) - \mathbb{E}_{\theta'} \log L(X, \theta') = \int_{\mathbb{R}} \log \frac{L(x, \theta)}{L(x, \theta')} \cdot L(x, \theta) dx.$$

Für den Fall $\mathbb{P}_\theta(L(X, \theta') = 0) > 0$ setzen wir $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \infty$. Im diskreten Fall betrachte statt des Integrals die Summe über die nicht trivialen $p_\theta(x)$.

Wir werden gleich zeigen, dass $H(\cdot, \cdot)$ die Eigenschaften $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = 0 \iff \theta = \theta'$ und $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \geq 0 \quad \forall \theta, \theta' \in \Theta$ besitzt. Es ist allerdings offensichtlich, dass $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'})$ nicht symmetrisch bzgl. θ und θ' ist. Somit ist $H(\cdot, \cdot)$ keine Metrik.

Lemma 3.4.2

Es gilt

1. $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'})$ ist wohldefiniert und ≥ 0 .
2. Falls $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = 0$, dann gilt $\theta = \theta'$.

Beweis Wir betrachten zum Beispiel den Fall absolut stetiger \mathbb{P}_θ , $\theta \in \Theta$ (diskreter Fall folgt analog).

1. Definieren wir

$$f(x) = \begin{cases} \frac{L(x, \theta)}{L(x, \theta')}, & \text{falls } L(x, \theta') > 0, \\ 1, & \text{sonst.} \end{cases}$$

Betrachten wir den Fall $\mathbb{P}_\theta(L(X, \theta') = 0) = 0$, so folgt $\mathbb{P}_\theta(L(X, \theta') > 0) = 1$. Ansonsten ist $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = \infty > 0$, also positiv und wohldefiniert. Dann folgt mit Wahrscheinlichkeit 1, dass $L(x, \theta) = f(x) \cdot L(x, \theta')$. Sei $g(x) = 1 - x + x \log x$, $x > 0$. Man kann zeigen, dass g konvex mit $g(x) \geq 0$ ist. Tatsächlich, es gilt

$$g'(x) = -1 + \log x + 1 = \log x, \quad g''(x) = 1/x > 0.$$

Somit besitzt g genau eine Nullstelle bei $x = 1$, die gleichzeitig ihr Minimum ist. Betrachten wir $g(f(X))$, $X \sim L(x, \theta')$. Dann gilt

$$\begin{aligned} 0 &\leq \mathbb{E}_{\theta'} g(f(X)) = 1 - \mathbb{E}_{\theta'} f(X) + \mathbb{E}_{\theta'} (f(X) \log f(X)) \\ &= 1 - \int \frac{L(x, \theta)}{L(x, \theta')} \cdot L(x, \theta') dx + \int \frac{L(x, \theta)}{L(x, \theta')} \cdot \log \frac{L(x, \theta)}{L(x, \theta')} \cdot L(x, \theta') dx = H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}). \end{aligned}$$

Somit gilt $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) \geq 0$, was zu zeigen war.

2. Falls $H(\mathbb{P}_\theta, \mathbb{P}_{\theta'}) = 0 \implies \mathbb{E}_{\theta'} g(f(X)) = 0$, $g(f(X)) \geq 0$. Somit folgt θ' -fast sicher $g(f(X)) = 0 \implies f(X) \stackrel{\theta'\text{-f.s.}}{=} 1$, damit entweder $L(x, \theta') = 0$ oder $L(x, \theta) = L(x, \theta')$ für alle x und daher $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$.

□

Satz 3.4.1 (Schwache Konsistenz von ML-Schätzern):

Sei $m = 1$ und Θ ein offenes Intervall aus \mathbb{R} . Sei $L(x_1, \dots, x_n, \theta)$ *unimodal*, d.h. für $\hat{\theta}$ ML-Schätzer für θ gilt

$$\begin{cases} \forall \theta < \hat{\theta}(x_1, \dots, x_n) & \implies L(x_1, \dots, x_n, \theta) \text{ ist steigend} \\ \forall \theta > \hat{\theta}(x_1, \dots, x_n) & \implies L(x_1, \dots, x_n, \theta) \text{ ist fallend} \end{cases}$$

(d.h. es existiert genau ein $\max_{\theta \in \Theta} L(x_1, \dots, x_n, \theta)$). Dann gilt $\hat{\theta}(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{P} \theta$.

Beweis Es ist zu zeigen, dass

$$\mathbb{P}_\theta \left(\left| \hat{\theta}(X_1, \dots, X_n) - \theta \right| > \varepsilon \right) \xrightarrow[n \rightarrow \infty]{} 0, \quad \varepsilon > 0. \quad (3.4.1)$$

Wählen wir beliebiges $\varepsilon > 0$: $\theta \pm \varepsilon \in \Theta$. Dann gilt $H(\mathbb{P}_\theta, \mathbb{P}_{\theta \pm \varepsilon}) > \delta > 0$, wegen der Unterscheidbarkeit der Parametrisierung von \mathbb{P}_θ und Lemma 3.4.2. Betrachten wir $\{|\hat{\theta} - \theta| \leq \varepsilon\}$. Um (3.4.1) zu zeigen, ist es hinreichend, eine untere Schranke für $\mathbb{P}_\theta(|\hat{\theta} - \theta| \leq \varepsilon)$ zu konstruieren, die für $n \rightarrow \infty$ gegen 1 konvergiert. Es gilt

$$\begin{aligned} \{|\hat{\theta} - \theta| < \varepsilon\} &\stackrel{\text{Unimod}}{\supseteq} \{L(X_1, \dots, X_n, \theta - \varepsilon) < L(X_1, \dots, X_n, \theta) > L(X_1, \dots, X_n, \theta + \varepsilon)\} \\ &= \left\{ \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta \pm \varepsilon)} > 1 \right\} \stackrel{\delta > 0 \implies e^{n\delta} > 1}{\supseteq} \left\{ \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta \pm \varepsilon)} > e^{n\delta} \right\} \\ &= \left\{ \frac{1}{n} \log \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta \pm \varepsilon)} > \delta \right\} = A_+ \cap A_-, \end{aligned}$$

wobei

$$A_\pm = \left\{ \frac{1}{n} \log \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta \pm \varepsilon)} > \delta \right\}.$$

Somit gilt also

$$\mathbb{P}_\theta \left(|\hat{\theta} - \theta| < \varepsilon \right) \geq \mathbb{P}_\theta(A_+ \cap A_-) = \mathbb{P}_\theta(A_+) + \mathbb{P}_\theta(A_-) - \mathbb{P}_\theta(A_+ \cup A_-).$$

Wenn wir zeigen können, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta(A_\pm) = 1, \quad (3.4.2)$$

dann folgt daraus

$$1 \geq \lim_{n \rightarrow \infty} \mathbb{P}_\theta(A_+ \cup A_-) \geq \lim_{n \rightarrow \infty} \mathbb{P}_\theta(A_\pm) = 1 \implies \lim_{n \rightarrow \infty} \mathbb{P}_\theta(A_+ \cup A_-) = 1$$

und

$$1 \geq \lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(|\hat{\theta} - \theta| < \varepsilon \right) \geq 1 + 1 - 1 = 1,$$

womit folgt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(|\hat{\theta} - \theta| > \varepsilon \right) = 1 - \underbrace{\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(|\hat{\theta} - \theta| < \varepsilon \right)}_{=1} = 0$$

und somit $\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta} - \theta| \geq \varepsilon) = 0$, d.h., $\hat{\theta} \xrightarrow[n \rightarrow \infty]{P} \theta$.

Jetzt zeigen wir, dass $\mathbb{P}_\theta(A_+) \xrightarrow[n \rightarrow \infty]{} 1$ (für $\mathbb{P}_\theta(A_-)$ ist es analog).

1. Sei $H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) < \infty$. Sei

$$f(x) = \begin{cases} \frac{L(x, \theta)}{L(x, \theta + \varepsilon)}, & \text{falls } L(x, \theta + \varepsilon) > 0, \\ 1, & \text{sonst.} \end{cases}$$

Dann folgt aus Definition 3.4.4, dass $\mathbb{P}_\theta(x : L(x, \theta + \varepsilon) > 0) = 1$. Weiter gilt

$$\begin{aligned} \frac{1}{n} \log \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta + \varepsilon)} &= \frac{1}{n} \sum_{i=1}^n \log \frac{L(X_i, \theta)}{L(X_i, \theta + \varepsilon)} = \frac{1}{n} \sum_{i=1}^n \log f(X_i) \\ &\xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}_\theta \log f(X_1) = \int L(x, \theta) \cdot \log \frac{L(x, \theta)}{L(x, \theta + \varepsilon)} dx = H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) > \delta > 0 \end{aligned}$$

nach dem starken Gesetz der großen Zahlen, weil $\log f(X_1) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ wegen

$$\mathbb{E}_\theta \log f(X_1) = H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) < \infty \implies \mathbb{P}(A_+) \xrightarrow[n \rightarrow \infty]{} 1.$$

2. Sei $H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) = \infty$ und $\mathbb{P}_\theta(x \in \mathbb{R} : L(x, \theta + \varepsilon) = 0) = 0$, dann folgt

$$f(x) \stackrel{\text{f.s.}}{=} \frac{L(x, \theta)}{L(x, \theta + \varepsilon)}.$$

Es gilt $\log \min\{f(X_1), c\} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ für alle $c > 0$. Somit folgt wie in Punkt 1:

$$\frac{1}{n} \sum_{i=1}^n \log \min\{f(X_i), c\} \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}_\theta \log \min\{f(X_1), c\} \in (0, \infty) \xrightarrow[c \rightarrow \infty]{} H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) = \infty$$

und damit

$$\begin{aligned} A_+ &\supset \left\{ \frac{1}{n} \sum_{i=1}^n \log \min\{f(X_i), c\} > \delta \right\} \\ \implies \mathbb{P}(A_+) &\geq \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n \log \min\{f(X_i), c\} > \delta \right) \xrightarrow[n \rightarrow \infty]{} 1. \end{aligned}$$

3. Sei $H(\mathbb{P}_\theta, \mathbb{P}_{\theta+\varepsilon}) = \infty$ und $\mathbb{P}_\theta(x \in \mathbb{R} : L(x, \theta + \varepsilon) = 0) = a > 0$, dann folgt

$$\begin{aligned} \mathbb{P}_\theta \left(\frac{1}{n} \log \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta + \varepsilon)} = \infty \right) &= 1 - \mathbb{P} \left(\frac{1}{n} \log \frac{L(X_1, \dots, X_n, \theta)}{L(X_1, \dots, X_n, \theta + \varepsilon)} < \infty \right) \\ &= 1 - \mathbb{P} \left(\bigcap_{i=1}^n \{L(X_i, \theta + \varepsilon) > 0\} \right) \stackrel{X_i \text{ u.i.v.}}{=} 1 - (1 - a)^n \xrightarrow[n \rightarrow \infty]{} 1 \end{aligned}$$

Insgesamt also $\mathbb{P}(A_+) \xrightarrow[n \rightarrow \infty]{} 1$.

□

Definition 3.4.5

Sei $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe von unabhängigen identisch verteilten Zufallsvariablen $X_i \sim F_\theta$, $\theta \in \Theta$. Sei $L(x, \theta)$ die Likelihood-Funktion von X_i . Dann heißt der Ausdruck

$$I(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \theta) \right)^2, \quad \theta \in \Theta \quad (3.4.3)$$

die *Fisher-Information* der Stichprobe (X_1, \dots, X_n) .

Es wird in Zukunft vorausgesetzt, dass $0 < I(\theta) < \infty$. Wir stellen nun einige Bedingungen auf, die für die asymptotische Normalverteiltheit von ML-Schätzern notwendig sind.

1. $\Theta \subset \mathbb{R}$ ist ein offenes Intervall ($m = 1$).
2. Es gelte $\mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$ genau dann, wenn $\theta \neq \theta'$.
3. Die Familie $\{\mathbb{P}_\theta, \theta \in \Theta\}$, $\theta \in \Theta$ bestehe nur aus diskreten oder nur aus absolut stetigen Verteilungen, also nicht aus Mischungen von diskreten und absolut stetigen Verteilungen.
4. $B = \text{supp } L(x, \theta) = \{x \in \mathbb{R} : L(x, \theta) > 0\}$ hängt nicht von $\theta \in \Theta$ ab. Dabei heißt supp (von englisch „support“) der „Träger“ einer Funktion f und ist definiert als

$$\text{supp } f = \{x \in \mathbb{R} : f(x) \neq 0\}$$

und die Likelihood-Funktion $L(x, \theta)$ ist durch

$$L(x, \theta) = \begin{cases} p(x, \theta), & \text{im diskreten Fall,} \\ f(x, \theta), & \text{im absolut stetigen Fall} \end{cases} \quad (3.4.4)$$

gegeben, wobei $p(x, \theta)$ bzw. $f(x, \theta)$ die Wahrscheinlichkeitsfunktion bzw. Dichte von \mathbb{P}_θ ist.

5. Die Abbildung $L(x, \theta)$ ist dreimal stetig differenzierbar und es gilt

$$0 = \frac{d^k}{d\theta^k} \int_B L(x, \theta) dx = \int_B \frac{\partial^k}{\partial \theta^k} L(x, \theta) dx, \quad k = 1, 2, \theta \in \Theta.$$

Da das Integral über die Dichte $L(x, \theta)$ gleich 1 ist, ist die Ableitung gleich 0. Dabei sind im diskreten Fall die Integrale durch Summen zu ersetzen.

6. Für alle $\theta_0 \in \Theta$ existiert eine Konstante $\delta_{\theta_0} > 0$ und eine messbare Funktion $g_{\theta_0} : B \rightarrow [0, \infty)$, so dass

$$\left| \frac{\partial^3 \log L(x, \theta)}{\partial \theta^3} \right| \leq g_{\theta_0}(x), \quad \forall x \in B, \quad |\theta - \theta_0| < \delta_{\theta_0},$$

wobei $\mathbb{E}_{\theta_0} g_{\theta_0}(X_1) < \infty$.

Bemerkung 3.4.3

Es gilt folgende Relation:

$$n \cdot I(\theta) = \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \dots, X_n, \theta) \right),$$

wobei

$$L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n L(X_i, \theta) \quad (3.4.5)$$

die Likelihood-Funktion der Stichprobe (X_1, \dots, X_n) ist mit $L(X_i, \theta)$ nach (3.4.4).

Beweis Es gilt

$$\frac{\partial}{\partial \theta} \log L(X_1, \dots, X_n, \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^n \log L(X_i, \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log L(X_i, \theta) = \sum_{i=1}^n \frac{L'(X_i, \theta)}{L(X_i, \theta)}.$$

Ferner

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \dots, X_n, \theta) \right) = \sum_{i=1}^n \mathbb{E}_\theta \frac{L'(X_i, \theta)}{L(X_i, \theta)} = \sum_{i=1}^n \int_B \frac{L'(X, \theta)}{L(X, \theta)} \cdot L(X, \theta) dx \stackrel{5)}{=} 0.$$

Insgesamt gilt also

$$\begin{aligned} \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \dots, X_n, \theta) \right) &= \text{Var}_\theta \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \log L(X_i, \theta) \right) \\ &\stackrel{X_i \text{ unabhg.}}{=} \sum_{i=1}^n \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_i, \theta) \right) \stackrel{X_i \text{ ident. vert.}}{=} n \cdot \text{Var}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \theta) \right) \\ &= n \cdot \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log L(X_1, \theta) \right)^2 = n \cdot I(\theta). \end{aligned}$$

□

Satz 3.4.2

Sei (X_1, \dots, X_n) eine Stichprobe von Zufallsvariablen, für die die Bedingungen 1) bis 6) erfüllt sind und $0 < I(\theta) < \infty$, $\theta \in \Theta$. Falls $\hat{\theta}(X_1, \dots, X_n)$ ein schwach konsistenter ML-Schätzer für θ ist, dann ist $\hat{\theta}(X_1, \dots, X_n)$ asymptotisch normalverteilt:

$$\sqrt{n \cdot I(\theta)} \left(\hat{\theta}(X_1, \dots, X_n) - \theta \right) \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

Beweis Führen wir die Bezeichnung $l_n(\theta) = \log L(X_1, \dots, X_n, \theta)$, $\theta \in \Theta$ ein. Sei

$$l_n^{(k)}(\theta) = \frac{d^k}{d\theta^k} l_n(\theta), \quad k = 1, 2, 3.$$

Ist $\hat{\theta}$ ein ML-Schätzer, so folgt $l_n^{(1)}(\hat{\theta}) = 0$. Schreiben wir die Taylor-Entwicklung von $l_n^{(1)}(\hat{\theta})$ in der Umgebung von θ auf:

$$0 = l_n^{(1)}(\hat{\theta}) = l_n^{(1)}(\theta) + (\hat{\theta} - \theta) \cdot l_n^{(2)}(\theta) + (\hat{\theta} - \theta)^2 \cdot \frac{l_n^{(3)}(\theta^*)}{2},$$

wobei θ^* zwischen θ und $\hat{\theta}$ liegt. Dabei ist

$$-(\hat{\theta} - \theta) \left(l_n^{(2)}(\theta) + (\hat{\theta} - \theta) \frac{l_n^{(3)}(\theta^*)}{2} \right) = l_n^{(1)}(\theta) \implies \sqrt{n}(\hat{\theta} - \theta) = \frac{\frac{l_n^{(1)}(\theta)}{\sqrt{n}}}{-\frac{l_n^{(2)}(\theta)}{n} - (\hat{\theta} - \theta) \frac{l_n^{(3)}(\theta^*)}{2n}}$$

Falls wir zeigen können, dass

$$1. \quad \frac{l_n^{(1)}(\theta)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, I(\theta)),$$

$$2. \quad -\frac{l_n^{(2)}(\theta)}{n} \xrightarrow[n \rightarrow \infty]{\text{f.s.}} I(\theta),$$

$$3. \quad (\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{P} 0 \quad \text{und} \quad \frac{l_n^{(3)}(\theta^*)}{2n}$$

beschränkt ist, das heißt

$$\exists c > 0 : \quad \lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\left| \frac{l_n^{(3)}(\theta^*)}{2n} \right| < c \right) = 1,$$

dann konvergiert der Ausdruck

$$(\hat{\theta} - \theta) \cdot \frac{l_n^{(3)}(\theta^*)}{2n} \xrightarrow[n \rightarrow \infty]{P} 0, \quad \text{weil} \quad \left| \frac{l_n^{(3)}(\theta^*)}{n} \right| \leq g_\theta(X_1) \text{ integrierbar}$$

und somit gilt

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{\frac{l_n^{(1)}(\theta)}{\sqrt{n}}}{-\frac{l_n^{(2)}(\theta)}{n} - (\hat{\theta} - \theta) \frac{l_n^{(3)}(\theta^*)}{2n}} \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim N\left(0, \frac{1}{I(\theta)}\right)$$

nach dem Satz von Slutsky. Damit folgt $\sqrt{n}\sqrt{I(\theta)}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1)$

1. Es gilt

$$\frac{l_n^{(1)}}{\sqrt{n}} = \frac{\sum_{i=1}^n \frac{\partial}{\partial \theta} \log L(X_i, \theta)}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} Y_1 \sim N\left(0, \underbrace{\text{Var}_\theta\left(\frac{\partial}{\partial \theta} L(X_i, \theta)\right)}_{=I(\theta)}\right)$$

nach dem zentralen Grenzwertsatz, weil $\frac{\partial}{\partial \theta} \log L(X_i, \theta)$ unabhängig identisch verteilte Zufallsvariablen mit Erwartungswert 0 (siehe Bemerkung 3.4.3) sind.

2.

$$\begin{aligned} -\frac{1}{n} l_n^{(2)}(\theta) &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log L(X_i, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{\left(L^{(1)}(X_i, \theta)\right)^2 - L(X_i, \theta) \cdot L^{(2)}(X_i, \theta)}{\left(L(X_i, \theta)\right)^2} \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{L^{(1)}(X_i, \theta)}{L(X_i, \theta)}\right)^2 - \frac{1}{n} \sum_{i=1}^n \frac{L^{(2)}(X_i, \theta)}{L(X_i, \theta)} \\ &\xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}_\theta \left(\frac{L^{(1)}(X_1, \theta)}{L(X_1, \theta)}\right)^2 - \mathbb{E}_\theta \left(\frac{L^{(2)}(X_1, \theta)}{L(X_1, \theta)}\right) = I(\theta) - 0 = I(\theta) \end{aligned}$$

nach dem Gesetz der großen Zahlen, wobei

$$L^{(k)}(X_i, \theta) = \frac{\partial^k}{\partial \theta^k} L(X_i, \theta)$$

und

$$\mathbb{E}_\theta \left(\frac{L^{(2)}(X_1, \theta)}{L(X_1, \theta)}\right) = \int_B \frac{\partial^2}{\partial \theta^2} L(x, \theta) dx \stackrel{5)}{=} \frac{d^2}{d\theta^2} \int_B L(x, \theta) dx = 0.$$

3. $\hat{\theta} \xrightarrow[n \rightarrow \infty]{P} \theta$, weil $\hat{\theta}$ schwach konsistent ist. Zeigen wir, dass

$$\frac{l_n^{(3)}(\theta^*)}{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Aus $\hat{\theta} \xrightarrow[n \rightarrow \infty]{P} \theta$ folgt für alle $\varepsilon > 0$

$$\mathbb{P}\left(|\hat{\theta} - \theta| \leq \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 1.$$

Damit folgt, dass mit asymptotisch großer Wahrscheinlichkeit $|\hat{\theta} - \theta| \leq \delta$, $\delta > 0$ gilt, welches aus der Bedingung 6) folgt. Damit gilt, dass für alle θ : $|\hat{\theta} - \theta| < \delta$

$$\left| \frac{l_n^{(3)}(\theta^*)}{n} \right| \leq \frac{1}{n} \sum_{i=1}^n \underbrace{\left| \frac{\partial^3}{\partial \theta^3} \log L(X_i, \theta) \right|}_{\leq g_\theta(X_i)} \leq \frac{1}{n} \sum_{i=1}^n g_\theta(X_i) \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}_\theta g_\theta(X_1) < \infty.$$

So folgt, dass eine Konstante $c > 0$ existiert, sodass

$$\mathbb{P}_\theta \left(\left| \frac{l_n^{(3)}(\theta^*)}{n} \right| < c \right) \xrightarrow[n \rightarrow \infty]{} 1 \quad \text{und somit} \quad \frac{l_n^{(3)}(\theta^*)}{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Der Beweis ist beendet. □

3.4.3 Bayes-Schätzer

Sei (X_1, \dots, X_n) eine Zufallsstichprobe, wobei X_i unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion F_θ , $\theta \in \Theta$ sind. Sei F_θ entweder eine diskrete oder eine absolut stetige Verteilung. Sei aber auch θ eine Zufallsvariable $\tilde{\theta}$ mit Verteilung $Q(\cdot)$ auf dem Messraum $(\Theta, \mathcal{B}_\Theta)$, die entweder diskret mit Zähldichte $q(\cdot)$ oder absolut stetig mit Dichte $q(\cdot)$ ist. Nach wie vor werden beide Fälle gemeinsam betrachtet, dabei entsprechen sich die Summation und Integration im diskreten bzw. absolut stetigen Fall.

Definition 3.4.6

Die Verteilung $Q(\cdot)$ heißt *a-priori-Verteilung* des Parameters θ (von $\tilde{\theta}$) (a-priori bedeutet hier „vor dem Experiment (X_1, \dots, X_n) “).

Definition 3.4.7

Die *a-posteriori-Verteilung* des Parameters θ (von $\tilde{\theta}$) ist gegeben durch die (Zähl-)Dichte

$$q_{X_1, \dots, X_n}(\theta) = \begin{cases} \mathbb{P}(\tilde{\theta} = \theta | X_1 = x_1, \dots, X_n = x_n), & \text{falls die Verteilung } Q \text{ diskret ist,} \\ f_{\tilde{\theta}|X_1, \dots, X_n}(\theta, x_1, \dots, x_n), & \text{falls die Verteilung } Q \text{ absolut stetig ist.} \end{cases}$$

Dabei ist

$$\begin{aligned} \mathbb{P}(\tilde{\theta} = \theta | X = x_1, \dots, X = x_n) &= \frac{\mathbb{P}(\tilde{\theta} = \theta, X_1 = x_1, \dots, X_n = x_n)}{\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)} \\ &= \frac{\mathbb{P}_\theta(X_i = x_i, i = 1, \dots, n) \cdot q(\theta)}{\sum_{\theta_1 \in \Theta} \mathbb{P}_{\theta_1}(X_i = x_i, i = 1, \dots, n) \cdot q(\theta_1)} \end{aligned}$$

die *Bayesche Formel*, bzw.

$$f_{\tilde{\theta}|X_1, \dots, X_n}(\theta, x_1, \dots, x_n) = \frac{f_{(\tilde{\theta}, X_1, \dots, X_n)}(\theta, x_1, \dots, x_n)}{f_{X_1, \dots, X_n}(x_1, \dots, x_n)} = \frac{L(x_1, \dots, x_n, \theta) \cdot q(\theta)}{\int_{\Theta} L(x_1, \dots, x_n, \theta_1) \cdot q(\theta_1) d\theta_1},$$

mit $L(x_1, \dots, x_n, \theta)$ nach (3.4.5).

Definition 3.4.8

Eine *Verlustfunktion* $V : \Theta^2 \rightarrow \mathbb{R}_+$ ist eine Θ^2 -messbare Funktion.

Verlustfunktionen spielen in unseren Betrachtungen folgende Rolle: $\mathbb{E}_*V(\tilde{\theta}, a)$ stellt den *erwarteten Verlust* (mittleres Risiko) dar, der bei der Schätzung des Parameters θ durch a entsteht. Dabei stellt \mathbb{E}_* den Erwartungswert bezüglich der *a-posteriori-Verteilung* von $\tilde{\theta}$ dar. Es sind offensichtlich die konkreten Stichprobenwerte x_1, \dots, x_n in die a-posteriori-Verteilung eingegangen, deshalb ist $\mathbb{E}_*V(\tilde{\theta}, a)$ eine Funktion von a und x_1, \dots, x_n :

$$\mathbb{E}_*V(\tilde{\theta}, a) = \varphi(x_1, \dots, x_n, a).$$

Definition 3.4.9

Ein Schätzer $\hat{\theta}$ heißt *Bayes-Schätzer* des Parameters θ , falls

$$\hat{\theta}(x_1, \dots, x_n) = \arg \min_a \mathbb{E}_*V(\tilde{\theta}, a) \tag{3.4.6}$$

existiert und eindeutig ist.

Bemerkung 3.4.4

1. Manchmal gilt $\hat{\theta} \notin \Theta$, was mit der Existenz des Minimums von $\varphi(x_1, \dots, x_n, a)$ auf Θ zu tun hat.
2. Der Name „Bayesscher Ansatz“ stammt von dem englischen Mathematiker Thomas Bayes (1702–1761), der die Bayessche Formel

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j) \cdot \mathbb{P}(B_j)} \tag{3.4.7}$$

nur ideenhaft eingeführt hat. Der eigentliche Entdecker der Formel (3.4.7) ist Pierre-Simon Laplace (1749–1827) (Ende des XVIII. Jahrhunderts). Diese Formel wurde bei der Herleitung der *a-posteriori-Verteilung* von $\tilde{\theta}$ implizit benutzt.

3. Die Vorgehensweise in Definition 3.4.9 ist in konkreten praxisrelevanten Fällen meistens nur numerisch möglich. Es gibt sehr wenige Beispiele für analytische Lösungen des in (3.4.6) gestellten Minimierungsproblems.

Beispiel 3.4.3 (Quadratische Verlustfunktion):

Ist $V(\theta_1, \theta_2) = (\theta_1 - \theta_2)^2$, so ist

$$\min_a (\varphi(x_1, \dots, x_n, a)) = \min_a (\mathbb{E}_*(\tilde{\theta} - a)^2) = \min_a (\mathbb{E}_*\tilde{\theta}^2 - 2a\mathbb{E}_*\tilde{\theta} + a^2) = \mathbb{E}_*\tilde{\theta}$$

und daher der *Bayes-Schätzer* $\hat{\theta}(x_1, \dots, x_n)$ für θ durch $\mathbb{E}_*\tilde{\theta}$ gegeben.

Beispiel 3.4.4 (Bernoulli-Verteilung):

Sei (X_1, \dots, X_n) eine unabhängig identisch verteilte Stichprobe von $X_i \sim \text{Bernoulli}(p)$, $p \in (0, 1)$. Weiter sei die a-priori-Verteilung

$$\tilde{p} \sim \text{Beta}(\alpha, \beta), \quad \alpha, \beta > 0, \text{ mit Zähldichte } q(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \cdot \mathbb{I}(p \in [0, 1]),$$

die a-posteriori-Verteilung von \tilde{p} ist dann gleich

$$q^*(p) = f_{\tilde{p}|X_1=x_1, \dots, X_n=x_n}(p) = \frac{\mathbb{P}_p(X_1 = x_1, \dots, X_n = x_n) \cdot q(p)}{\int_0^1 \mathbb{P}_{p_1}(X_1 = x_1, \dots, X_n = x_n) \cdot q(p_1) dp_1}.$$

Es ist immer möglich die a-posteriori-Verteilung nicht bezüglich des Vektors (X_1, \dots, X_n) , sondern bezüglich einer Funktion $g(X_1, \dots, X_n)$, zu berechnen (*Komplexitätsreduktion*).

Hier ist $Y = g(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ die Gesamtanzahl aller Erfolge in n Experimenten, wobei

$$X_i = \begin{cases} 1, & \text{mit Wahrscheinlichkeit } p, \\ 0, & \text{sonst.} \end{cases}$$

Daher gilt für die a-posteriori-Verteilung bzgl. Y :

$$\begin{aligned} q^*(p) = f_{\tilde{p}|Y=k}(p) &= \frac{\mathbb{P}_p(Y = k) \cdot q(p)}{\int_0^1 \mathbb{P}_{p_1}(Y = k) q(p_1) dp_1} \\ &\stackrel{Y \sim \text{Bin}(n, p)}{\text{falls } \tilde{p}=p} \frac{\binom{n}{k} p^k (1-p)^{n-k} \cdot (B(\alpha, \beta))^{-1} \cdot p^{\alpha-1} (1-p)^{\beta-1}}{\frac{\binom{n}{k}}{B(\alpha, \beta)} \cdot \int_0^1 p_1^{k+\alpha-1} (1-p_1)^{n-k+\beta-1} dp_1} \\ &= \frac{p^{k+\alpha-1} (1-p)^{n-k+\beta-1}}{B(k+\alpha, n-k+\beta)}, \quad p \in [0, 1]. \end{aligned}$$

Daher ist die a-posteriori-Verteilung von \tilde{p} unter der Bedingung $Y = k$ durch

$$\text{Beta}(k + \alpha, n - k + \beta)$$

gegeben.

Für den *Bayes-Schätzer* gilt:

$$\begin{aligned} \hat{p}(x_1, \dots, x_n) = \mathbb{E}_* \tilde{p} &= \int_0^1 p \cdot q^*(p) dp = \frac{\int_0^1 p^{k+\alpha} (1-p)^{n-k+\beta-1} dp}{B(k+\alpha, n-k+\beta)} \\ &= \frac{B(k+\alpha+1, n-k+\beta)}{B(k+\alpha, n-k+\beta)} = \dots = \frac{k+\alpha}{\alpha+\beta+n} = \frac{\sum_{i=1}^n x_i + \alpha}{\alpha+\beta+n} = \frac{\alpha + n\bar{x}_n}{\alpha+\beta+n}. \end{aligned}$$

Interpretation:

$$\hat{p}(X_1, \dots, X_n) = \underbrace{\frac{n}{\alpha+\beta+n}}_{=:c_1} \bar{X}_n + \underbrace{\frac{\alpha+\beta}{\alpha+\beta+n}}_{=:c_2} \cdot \frac{\alpha}{\alpha+\beta} = c_1 \cdot \bar{X}_n + c_2 \cdot \mathbb{E}_{apr} \tilde{\theta},$$

wobei $c_1 + c_2 = 1$ ist. Dies heißt, dass die Bayessche Methode einen Mittelweg zwischen dem Schätzer $\mathbb{E}_{apr} \tilde{\theta}$ (in Abwesenheit der Information über die Stichprobe (X_1, \dots, X_n)) und dem M-Schätzer \bar{X}_n (in Abwesenheit der a-priori-Information über die Verteilung von \tilde{p}) für p einschlägt.

3.4.4 Resampling-Methoden zur Gewinnung von Punktschätzern

Sei (X_1, \dots, X_n) eine Stichprobe im parametrischen Modell. Gesucht ist ein Schätzer $\hat{\theta}$ für den Parameter θ . Um diesen Schätzer zu konstruieren, werden bei Resampling-Methoden neue Stichproben (X_1^*, \dots, X_n^*) durch das unabhängige Ziehen mit Zurücklegen aus der alten Stichprobe (X_1, \dots, X_n) generiert und auf ihrer Basis Mittelwerte, Stichprobenvarianzen und andere Schätzer gebildet. Dabei ist die Dimension m des Parameterraums Θ beliebig.

Wir werden im Folgenden die *Resampling*-Methoden

1. *Jackknife* (dt. „Taschenmesser“, weist auf Mittel, die jedem immer zur Hand sein sollten)
2. *Bootstrap* (engl. „self-sufficient“, dt. „mit eigenen Ressourcen“)

betrachten.

1. *Jackknife-Methoden zur Schätzung der Varianz bzw. der Verzerrung von Schätzern:*

Als einführendes Beispiel betrachten wir $\theta = \mathbb{E}X = \mu$ bzw. $\theta = \text{Var} X = \sigma^2$ und ihre (erwartungstreue) Schätzer $\hat{\mu} = \bar{X}_n$ bzw. $\hat{\sigma}^2 = S_n^2$.

Wie wir bereits wissen, gilt

$$\text{Var} \hat{\mu} = \frac{\sigma^2}{n}, \quad \text{Var} \hat{\sigma}^2 = \frac{1}{n} \left(\mu'_4 - \frac{n-3}{n-1} \sigma^4 \right).$$

Nun ist ein Schätzer für die Varianz von $\hat{\mu}$ bzw. $\hat{\sigma}^2$ gesucht. Dazu verwenden wir die Plug-in Methode

$$\widehat{\text{Var}} \hat{\mu} = \frac{S_n^2}{n}, \quad \widehat{\text{Var}} \hat{\sigma}^2 = \frac{1}{n} \left(\hat{\mu}'_4 - \frac{n-3}{n-1} S_n^4 \right),$$

wobei $\hat{\mu}'_4$ das vierte zentrierte empirische Moment ist.

Im Allgemeinen sind jedoch *keine* Formeln von $\text{Var} \hat{\theta}$ bekannt. Hier kommt nun die *Jackknife*-Methode zum Einsatz:

- Sei $X_{[i]}$ die Stichprobe $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$, $i = 1, \dots, n$. Falls

$$\hat{\theta}(X_1, \dots, X_n) = \varphi_n(X_1, \dots, X_n),$$

so bilden wir

$$\hat{\theta}_{[i]} = \varphi_{n-1}(X_{[i]}), \quad \bar{\theta}_{[i]} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{[i]}, \quad \widehat{\text{Var}}_{jn}(\hat{\theta}) \stackrel{\text{def.}}{=} \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{[i]} - \bar{\theta}_{[i]})^2.$$

Definition 3.4.10

Der Schätzer $\hat{\theta}_{[i]}$ bzw. $\widehat{\text{Var}}_{jn}(\hat{\theta})$ heißt *Jackknife-Schätzer* für den Erwartungswert bzw. die Varianz des Schätzers $\hat{\theta}$ von θ .

Beispiel 3.4.5

Sei $\theta = \mu$, $\hat{\theta} = \hat{\mu} = \bar{X}_n$, so gilt

$$\varphi_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i,$$

womit folgt, dass

$$\begin{aligned}\hat{\theta}_{[i]} &= \frac{1}{n-1} \sum_{j \neq i} X_j = \frac{1}{n-1} \left(-X_i + \sum_{j=1}^n X_j \right) = \frac{n}{n-1} \bar{X}_n - \frac{1}{n-1} X_i, \quad \forall i = 1, \dots, n, \\ \bar{\theta}_{[\cdot]} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{[i]} = \frac{n}{n-1} \bar{X}_n - \frac{1}{n(n-1)} \sum_{i=1}^n X_i = \frac{n \cdot \bar{X}_n}{n-1} - \frac{\bar{X}_n}{n-1} = \frac{n-1}{n-1} \bar{X}_n = \bar{X}_n.\end{aligned}$$

Daher ist ein *Jackknife-Schätzer* für μ gleich \bar{X}_n .

Konstruieren wir nun einen *Jackknife-Schätzer der Varianz*:

$$\begin{aligned}\widehat{\text{Var}}_{jn}(\hat{\theta}) &= \frac{n-1}{n} \sum_{i=1}^n \left(\frac{n}{n-1} \bar{X}_n - \frac{1}{n-1} X_i - \bar{X}_n \right)^2 = \frac{n-1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} (\bar{X}_n - X_i) \right)^2 \\ &= \frac{n-1}{n(n-1)^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} S_n^2,\end{aligned}$$

wobei dies genau der Plug-in Schätzer der Varianz von $\hat{\mu}$ ist.

- *Jackknife-Schätzer für die Verzerrung eines Schätzers*

Sei $\hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ . Der Bias von $\hat{\theta}$ ist $\mathbb{E}_\theta \hat{\theta} - \theta = \text{Bias}(\hat{\theta})$.

Definition 3.4.11

Ein *Jackknife-Schätzer der Verzerrung (Bias)* von $\hat{\theta}$ ist durch

$$\widehat{\text{Bias}}_{jn}(\hat{\theta}) = (n-1)(\bar{\theta}_{[\cdot]} - \hat{\theta})$$

gegeben.

An folgenden Beispielen wird klar, dass der oben beschriebene Vorgang zur Verringerung der Verzerrung beiträgt:

Der *Schätzer*

$$\tilde{\theta} = \hat{\theta} - \widehat{\text{Bias}}_{jn}(\hat{\theta}) = n\hat{\theta} - (n-1)\bar{\theta}_{[\cdot]} \quad (3.4.8)$$

hat in der Regel einen *kleineren Bias* als $\hat{\theta}$. Dabei ist wiederum

$$\hat{\theta}_{[i]} = \varphi_{n-1}(X_{[i]}) \quad \text{und} \quad \bar{\theta}_{[\cdot]} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{[i]} \quad \text{mit} \quad \hat{\theta}(X_1, \dots, X_n) = \varphi_n(X_1, \dots, X_n).$$

Beispiel 3.4.6

- a) Ist $\theta = \mathbb{E}X_i = \mu$, so ist $\hat{\theta} = \bar{X}_n$ ein unverzerrter Schätzer für μ . Was ist der Bias-korrigierte Schätzer $\tilde{\mu}$? (Dieser sollte schließlich nicht schlechter werden!)

Es gilt $\bar{\theta}_{[\cdot]} = \bar{X}_n$, daher ist der Bias-Schätzer von Jackknife $\widehat{\text{Bias}}_{jn}(\hat{\theta}) = (n-1)(\bar{X}_n - \bar{X}_n) = 0$ und somit $\tilde{\theta} = \hat{\theta} - 0 = \bar{X}_n$. Wir haben also gesehen, dass die Jackknife-Methode die unverzerrten Schätzer (zumindest in diesem Beispiel) richtig behandelt, indem sie keinen zusätzlichen Bias einbaut.

- b) $\theta = \sigma^2 = \text{Var}X_i$, $\hat{\theta} = \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ein verzerrter M-Schätzer der Varianz. Was ist $\tilde{\theta}$ in diesem Fall?

Übungsaufgabe 3.4.1

Zeigen Sie, dass $\tilde{\theta} = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \hat{\sigma}^2$ ein erwartungstreuer Schätzer der Varianz ist. Somit wurde der Bias von $\hat{\sigma}^2$ durch die Anwendung der Jackknife-Methode vollständig beseitigt.

Beweisidee: Zeigen Sie hierzu zunächst, dass

$$\widehat{\text{Bias}}_{jn}(\hat{\theta}) = -\frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Bemerkung 3.4.5

Die Beispiele 3.4.6 a), b), in denen sich der Jackknife-Schätzer analytisch bestimmen ließ, sind eher eine Ausnahme als die Regel. In den meisten Fällen erfolgt die Bias-Reduktion mit Hilfe der Monte-Carlo-Methoden auf Basis der Formel (3.4.8).

2. *Bootstrap-Schätzer:*

Die Bootstrap-Methode besteht in dem Erzeugen einer neuen Stichprobe (X_1^*, \dots, X_n^*) , die aus einer approximativen Verteilung \hat{F} der Stichprobenvariablen X_i , $i = 1, \dots, n$ gewonnen wird. Seien \mathbb{E}_* und Var_* die wahrscheinlichkeitstheoretischen Größen, die auf dem Verteilungsgesetz \mathbb{P}_* der neuen Stichprobe (X_1^*, \dots, X_n^*) beruhen. Dabei gibt es folgende Möglichkeiten, \hat{F} zu konstruieren:

- i) $\hat{F}(x) = \hat{F}_n(x)$ die empirische Verteilungsfunktion von X_i , falls X_i unabhängig identisch verteilt sind.
- ii) \hat{F} ist ein parametrischer Schätzer von F , der parametrischen Verteilungsfunktion von X_i . Das heißt, falls $X_i \sim F_\theta$, $i = 1, \dots, n$ für ein $\theta \in \Theta$ und $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ ist, so setzen wir $\hat{F} = F_{\hat{\theta}}$ (Plug-in Methode).

Definition 3.4.12

Ein *Bootstrap-Schätzer* für den *Erwartungswert* (bzw. *Bias* oder *Varianz*) von Schätzer $\hat{\theta}(X_1, \dots, X_n)$ ist gegeben durch

- a) $\hat{\mathbb{E}}_{boot}(\hat{\theta}) = \mathbb{E}_* \hat{\theta}(X_1^*, \dots, X_n^*)$.
- b) $\widehat{\text{Bias}}_{boot}(\hat{\theta}) = \hat{\mathbb{E}}_{boot} \hat{\theta} - \hat{\theta}$.
- c) $\widehat{\text{Var}}_{boot}(\hat{\theta}) = \text{Var}_*(\hat{\theta}(X_1^*, \dots, X_n^*))$.

Beispiel 3.4.7

Sei $\theta = \mu = \mathbb{E}X_i$ und $\hat{F} = \hat{F}_n$ die empirische Verteilungsfunktion. Wie generiert man eine Stichprobe X_1^*, \dots, X_n^* , wobei $X_i^* \sim \hat{F}_n$?

\hat{F}_n gewichtet jede Beobachtung x_i der ursprünglichen Stichprobe mit dem Gewicht $1/n$, deshalb genügt es, einen der Einträge (x_1, \dots, x_n) auszuwählen (mit Wahrscheinlichkeit $1/n$, Urnenmodell „Ziehen mit Zurücklegen“), um X_j^* , $j = 1, \dots, n$ zu generieren.

Bootstrap-Schätzer für den Erwartungswert von $\hat{\mu} = \bar{X}_n$:

$$\hat{\mathbb{E}}_{boot} \hat{\mu} = \mathbb{E}_* \left(\frac{1}{n} \sum_{i=1}^n X_i^* \right) \stackrel{X_i^* \text{ u.i.v.}}{=} \frac{1}{n} \cdot n \mathbb{E}_*(X_1^*) = \int x d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n.$$

Somit folgt $\widehat{\text{Bias}}_{boot} \hat{\mu} = 0$.

$$\widehat{\text{Var}}_{boot}(\hat{\mu}) = \text{Var}_* \left(\frac{1}{n} \sum_{i=1}^n X_i^* \right) \stackrel{X_i^* \text{ u.i.v.}}{=} \frac{1}{n^2} \cdot n \cdot \text{Var}_*(X_1^*) = \frac{1}{n} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{\hat{\sigma}^2}{n},$$

ein Plug-in Schätzer für $\text{Var}\bar{X}_n = \sigma^2/n$.

Monte-Carlo-Methoden zur numerischen Berechnung von Bootstrap-Schätzern:

Was kann man tun, wenn keine expliziten Formeln für z.B. $\widehat{\text{Var}}_{\text{Boot}}(\hat{\theta})$ vorliegen (der Regelfall in der Statistik)?

Generiere M unabhängige Stichproben $(X_{i1}^*, \dots, X_{in}^*)$, $i = 1, \dots, M$ nach der Regel i) oder ii) mit Hilfe der Monte-Carlo-Simulation. Dann berechne

$$\hat{\theta}_i = \hat{\theta}(X_{i1}^*, \dots, X_{in}^*), \quad i = 1, \dots, M \quad \text{und setze} \quad \hat{\mathbb{E}}_{\text{boot}}\hat{\theta} \approx \frac{1}{M} \sum_{i=1}^M \hat{\theta}_i.$$

Ähnlich gewinnt man approximative Bootstrap-Schätzer für Bias $\hat{\theta}$ und $\text{Var}\hat{\theta}$:

$$\widehat{\text{Bias}}_{\text{boot}}\hat{\theta} \approx \hat{\mathbb{E}}_{\text{boot}}\hat{\theta} - \hat{\theta}, \quad \widehat{\text{Var}}_{\text{boot}}\hat{\theta} \approx \frac{1}{M-1} \sum_{i=1}^M (\hat{\theta}_i - \hat{\mathbb{E}}_{\text{boot}}\hat{\theta})^2.$$

Mehr sogar, man kann die Verteilungsfunktion von X_{ij}^* durch die empirische Verteilungsfunktion bestimmen:

$$\hat{F}_{\text{boot}}(x) = \frac{1}{M} \sum_{i=1}^M \frac{1}{n} \sum_{j=1}^n \mathbb{I}(X_{ij}^* \leq x), \quad x \in \mathbb{R}.$$

Ferner lassen sich mit Hilfe von oben genannten Methoden *Bootstrap-Konfidenzintervalle* für $\hat{\theta}$ ableiten:

Dafür lassen sich Quantile von $\hat{F}_{\text{boot}}(x)$ empirisch bestimmen. Damit gilt

$$\mathbb{P}\left(\hat{F}_{\text{boot}}^{-1}(\alpha_1) \leq \hat{\theta}(X_1^*, \dots, X_n^*) \leq \hat{F}_{\text{boot}}^{-1}(\alpha_2)\right) \approx 1 - \alpha_1 - \alpha_2 = 1 - \alpha,$$

wobei $\alpha = \alpha_1 + \alpha_2$ klein ist. Beachte dabei, dass man hofft, dass X_i^* sehr ähnlich verteilt ist wie X_i und somit

$$\mathbb{P}\left(\hat{F}_{\text{boot}}^{-1}(\alpha_1) \leq \hat{\theta}(X_1, \dots, X_n) \leq \hat{F}_{\text{boot}}^{-1}(\alpha_2)\right) \approx 1 - \alpha_1 - \alpha_2 = 1 - \alpha$$

gilt.

3.5 Weitere Güteeigenschaften von Punktschätzern

3.5.1 Ungleichung von Cramér-Rao

Sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen identisch verteilten Zufallsvariablen X_i mit Verteilungsfunktion F_θ , $\theta \in \Theta$. Sei $\hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ . Falls $\hat{\theta}$ erwartungstreu ist, dann misst man die Güte eines anderen erwartungstreuen Schätzers $\tilde{\theta}$ von θ am Wert seiner Varianz. Das bedeutet, falls $\text{Var}_\theta \tilde{\theta} < \text{Var}_\theta \hat{\theta}$, dann ist der Schätzer $\tilde{\theta}$ besser. Wir werden uns nun mit der Frage befassen, ob immer wieder neue, bessere Schätzer $\tilde{\theta}$ mit immer kleinerer Varianz konstruiert werden können. Die Antwort hierauf ist unter gewissen Voraussetzungen negativ. Die untere Schranke der Varianz $\text{Var}_\theta \hat{\theta}$ hierzu liefert der Satz von Cramér-Rao.

Sei $L(x, \theta)$ die Likelihood-Funktion von X_i , d.h.

$$L(x, \theta) = \begin{cases} \mathbb{P}_\theta(x), & \text{im diskreten Fall,} \\ f_\theta(x), & \text{im stetigen Fall} \end{cases}$$

und $L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n L(x_i, \theta)$ die Likelihood-Funktion von der gesamten Stichprobe (X_1, \dots, X_n) . Es gelten die Bedingungen 1) bis 5), die für die asymptotische Normalverteilung von ML-Schätzern auf Seite 80 gestellt wurden, wobei die Bedingung 5) für $k = 1$ gilt.

Satz 3.5.1 (Ungleichung von Cramér-Rao):

Sei $\hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ mit den folgenden Eigenschaften:

1. $\mathbb{E}_\theta \hat{\theta}^2(X_1, \dots, X_n) < \infty \quad \forall \theta \in \Theta$.
2. Für alle $\theta \in \Theta$ existiert

$$\frac{d}{d\theta} \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) = \begin{cases} \int_{\mathbb{R}} \hat{\theta}(x_1, \dots, x_n) \frac{\partial}{\partial \theta} L(x_1, \dots, x_n, \theta) dx_1 \dots dx_n, & \text{im stetigen Fall,} \\ \sum_{x_1, \dots, x_n} \hat{\theta}(x_1, \dots, x_n) \frac{\partial}{\partial \theta} L(x_1, \dots, x_n, \theta), & \text{im diskr. Fall.} \end{cases}$$

Dann gilt

$$\text{Var}_\theta \hat{\theta}(X_1, \dots, X_n) \geq \frac{\left(\frac{d}{d\theta} \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) \right)^2}{n \cdot I(\theta)}, \quad \theta \in \Theta,$$

wobei $I(\theta)$ die Fisher-Information aus (3.4.3) ist.

Beweis Führen wir die Funktion

$$\varphi_\theta(x_1, \dots, x_n) = \frac{\partial}{\partial \theta} \log L(x_1, \dots, x_n, \theta)$$

ein. In Bemerkung 3.4.3 haben wir bewiesen, dass

$$\mathbb{E}_\theta \varphi_\theta(X_1, \dots, X_n) = 0, \quad \text{Var}_\theta \varphi_\theta(X_1, \dots, X_n) = n \cdot \mathbb{I}(\theta).$$

Wenden wir die Ungleichung von Cauchy-Schwarz auf $\text{Cov}_\theta(\varphi_\theta(X_1, \dots, X_n), \hat{\theta}(X_1, \dots, X_n))$ an:

$$\begin{aligned} \text{Cov}_\theta \left(\varphi_\theta(X_1, \dots, X_n), \hat{\theta}(X_1, \dots, X_n) \right) &= \mathbb{E}_\theta \left(\varphi_\theta(X_1, \dots, X_n) \cdot \hat{\theta}(X_1, \dots, X_n) \right) - 0 \\ &\leq \sqrt{\text{Var}_\theta \varphi_\theta(X_1, \dots, X_n)} \sqrt{\text{Var}_\theta \hat{\theta}(X_1, \dots, X_n)} \end{aligned}$$

Somit folgt

$$\text{Var}_\theta \hat{\theta}(X_1, \dots, X_n) \geq \frac{\overbrace{\left(\mathbb{E}_\theta \left(\varphi_\theta(X_1, \dots, X_n) \cdot \hat{\theta}(X_1, \dots, X_n) \right) \right)^2}^{=: A}}{\text{Var}_\theta \varphi_\theta(X_1, \dots, X_n)} = \frac{A^2}{n \cdot I(\theta)}.$$

Es bleibt zu zeigen, dass

$$A = \frac{d}{d\theta} \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n).$$

Wir zeigen die Aussage für den absolut stetigen Fall (im diskreten Fall sind die Integrale durch Summen zu ersetzen):

$$\begin{aligned} A &= \int \frac{\partial}{\partial \theta} \log L(x_1, \dots, x_n, \theta) \cdot \hat{\theta}(x_1, \dots, x_n) \cdot L(x_1, \dots, x_n, \theta) dx_1 \dots dx_n \\ &= \int \frac{\partial}{\partial \theta} L(x_1, \dots, x_n, \theta) \cdot \hat{\theta}(x_1, \dots, x_n) dx_1 \dots dx_n \stackrel{\text{Vor. 2)}}{=} \frac{d}{d\theta} \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n). \end{aligned}$$

□

Folgerung 3.5.1

Falls $\hat{\theta}$ ein erwartungstreuer Schätzer für θ ist und die Voraussetzungen des Satzes 3.5.1 erfüllt sind, so gilt

$$\text{Var}_\theta \hat{\theta}(X_1, \dots, X_n) \geq \frac{1}{n \cdot I(\theta)}.$$

Beweis Wende die Ungleichung von Cramér-Rao an $\hat{\theta}$ mit

$$\frac{d}{d\theta} \left(\mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) \right) = \frac{d}{d\theta} \theta = 1$$

an.

□

An folgenden Beispielen werden wir sehen, dass der Schätzer \bar{X}_n des Erwartungswertes μ in der Klasse aller Schätzer für μ , die die Voraussetzungen des Satzes 3.5.1 erfüllen, die kleinste Varianz besitzt. Somit ist \bar{X}_n der beste erwartungstreue Schätzer in dieser Klasse für mindestens zwei parametrische Familien von Verteilungen:

- Normalverteilung und
- Poisson-Verteilung.

Beispiel 3.5.1

1. $X_i \sim N(\mu, \sigma^2)$, $\hat{\mu} = \bar{X}_n$ als Schätzer für μ . Dabei ist $\hat{\mu}$ erwartungstreu mit $\text{Var} \hat{\mu} = \sigma^2/n$. Zeigen wir, dass die Cramér-Rao-Schranke für die Varianz eines erwartungstreuen Schätzers $\hat{\theta}$ für μ ebenso gleich σ^2/n ist. Prüfen wir zunächst die Voraussetzungen des Satzes 3.5.1:

Zeigen wir, dass

$$\begin{aligned} 0 &= \frac{d}{d\mu} \int_{\mathbb{R}} L(x, \mu) dx = \int_{\mathbb{R}} \frac{\partial}{\partial \mu} L(x, \mu) dx \quad \text{mit} \quad L(x, \mu) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} : \\ \frac{\partial}{\partial \mu} L(x, \mu) &= \frac{2(x-\mu)}{2\sigma^2} \cdot \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{x-\mu}{\sigma^2} \cdot L(x, \mu), \\ \int_{\mathbb{R}} \frac{\partial}{\partial \mu} L(x, \mu) dx &= \mathbb{E} \left(\frac{X-\mu}{\sigma^2} \right) = 0. \end{aligned}$$

Zeigen wir weiterhin die Gültigkeit der Bedingung 2) des Satzes 3.5.1:

$$\frac{d}{d\mu} \mathbb{E} \bar{X}_n = \frac{d}{d\mu} (\mu) = 1 \stackrel{?}{=} \frac{1}{n} \int_{\mathbb{R}^n} (x_1 + \dots + x_n) \frac{\partial}{\partial \mu} \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \right) dx_1 \dots dx_n.$$

Induktion bzgl. n :

- Induktionsanfang $n = 1$:

$$\int_{\mathbb{R}} x \frac{\partial}{\partial \mu} L(x, \mu) dx = \int_{\mathbb{R}} \frac{x(x - \mu)}{\sigma^2} L(x, \mu) dx = \frac{1}{\sigma^2} (\mathbb{E}_{\mu} X^2 - \mu^2) = \frac{\text{Var}_{\mu} X}{\sigma^2} = 1.$$

- Induktionshypothese: Für n gilt

$$\int_{\mathbb{R}^n} (x_1 + \dots + x_n) \cdot \frac{\partial}{\partial \mu} L(x_1, \dots, x_n, \mu) dx_1 \dots dx_n = n.$$

- Induktionsschritt $n \rightarrow n + 1$:

$$A = \int_{\mathbb{R}^{n+1}} (x_1 + \dots + x_{n+1}) \frac{\partial}{\partial \mu} \underbrace{L(x_1, \dots, x_{n+1}, \mu)}_{=L(x_1, \dots, x_n, \mu) \cdot L(x_{n+1}, \mu)} dx_1 \dots dx_{n+1} \stackrel{?}{=} n + 1.$$

Dabei gilt für A :

$$\begin{aligned} A &= \int_{\mathbb{R}^{n+1}} (x_1 + \dots + x_n) \cdot \left(\frac{\partial}{\partial \mu} L(x_1, \dots, x_n, \mu) \cdot L(x_{n+1}, \mu) + L(x_1, \dots, x_n, \mu) \cdot \frac{\partial}{\partial \mu} L(x_{n+1}, \mu) \right) dx_1 \dots dx_n dx_{n+1} \\ &\quad + \int_{\mathbb{R}^{n+1}} x_{n+1} \left(\frac{\partial}{\partial \mu} L(x_1, \dots, x_n, \mu) \cdot L(x_{n+1}, \mu) + L(x_1, \dots, x_n, \mu) \cdot \frac{\partial}{\partial \mu} L(x_{n+1}, \mu) \right) dx_1 \dots dx_n dx_{n+1} \\ &= n \cdot \underbrace{\int_{\mathbb{R}} L(x_{n+1}, \mu) dx_{n+1}}_{=1} + \int_{\mathbb{R}^n} (x_1 + \dots + x_n) \cdot L(x_1, \dots, x_n, \mu) dx_1 \dots dx_n \\ &\quad \cdot \underbrace{\int_{\mathbb{R}} \frac{\partial}{\partial \mu} L(x_{n+1}, \mu) dx_{n+1}}_{=0} + \int_{\mathbb{R}} x_{n+1} L(x_{n+1}, \mu) dx_{n+1} \\ &\quad \cdot \underbrace{\int_{\mathbb{R}^n} \frac{\partial}{\partial \mu} L(x_1, \dots, x_n, \mu) dx_1 \dots dx_n}_{=0} + \underbrace{\int_{\mathbb{R}} x_{n+1} \frac{\partial}{\partial \mu} L(x_{n+1}, \mu) dx_{n+1}}_{=\frac{d}{d\mu} \mathbb{E}_{\mu} X = \frac{d}{d\mu} \mu = 1} \\ &\quad \cdot \underbrace{\int_{\mathbb{R}^n} L(x_1, \dots, x_n, \mu) dx_1 \dots dx_n}_{=1} = n + 1. \end{aligned}$$

Nachdem alle Voraussetzungen erfüllt sind, berechnen wir die Schranke

$$\frac{1}{n \cdot I(\mu)} \quad \text{mit} \quad I(\mu) = \mathbb{E}_{\mu} \left(\frac{\partial}{\partial \mu} \log L(X, \mu) \right)^2.$$

Es gilt

$$\frac{\partial}{\partial \mu} \log L(x, \mu) = \frac{\partial}{\partial \mu} \left(-\log \sqrt{2\pi\sigma^2} - \frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right) = -\frac{2(x - \mu)}{2\sigma^2} \cdot (-1) = \frac{x - \mu}{\sigma^2},$$

woraus folgt, dass

$$I(\mu) = \frac{1}{\sigma^4} \mathbb{E}_\mu (X - \mu)^2 = \frac{1}{\sigma^4} \cdot \text{Var}_\mu X = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} \quad \implies \quad n \cdot I(\mu) = \frac{n}{\sigma^2}.$$

Insgesamt gilt also

$$\text{Var}_\mu \hat{\theta} \geq \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n} = \text{Var}_\mu \bar{X}_n$$

für einen beliebigen erwartungstreuen Schätzer $\hat{\theta}$ für μ , der die Voraussetzungen des Satzes 3.5.1 erfüllt.

2. Das zweite Beispiel sei folgende Übungsaufgabe:

Übungsaufgabe 3.5.1

Seien $X_i \sim \text{Poisson}(\lambda)$, $i = 1, \dots, n$. Zeigen Sie, dass die Schranke von Cramér-Rao

$$\frac{1}{n \cdot I(\lambda)} = \frac{\lambda}{n} = \text{Var}_\lambda \bar{X}_n$$

ist. Dies bedeutet, dass auch hier \bar{X}_n der beste erwartungstreue Schätzer ist, der die Voraussetzungen des Satzes 3.5.1 erfüllt.

An Hand des nächsten Beispiels wollen wir zeigen, dass die Konstruktion von Schätzern mit einer Varianz, die kleiner als die Cramér-Rao-Schranke ist, möglich ist, falls die Voraussetzungen von Satz 3.5.1 nicht erfüllt sind.

Beispiel 3.5.2

Seien $X_i \sim U[0, \theta]$, $\theta > 0$. Dann ist die Bedingung „ $\text{supp} f_\theta(x) = [0, \theta]$ unabhängig von θ “ verletzt und auch eine weitere Bedingung:

$$0 \neq \int_{\mathbb{R}} \frac{\partial}{\partial \theta} L(x, \theta) dx = \int_0^\theta \left(\frac{1}{\theta} \right)' dx = -\frac{1}{\theta^2} \cdot \theta = -\frac{1}{\theta}.$$

Sei $\hat{\theta}$ ein erwartungstreuer Schätzer für θ , so würde nach der Ungleichung von Cramér-Rao folgen, dass $\text{Var}_\theta \hat{\theta} \geq (n \cdot I(\theta))^{-1}$, wobei

$$I(\theta) = \mathbb{E} \left(\frac{\partial}{\partial \theta} \log L(X, \theta) \right)^2 = \int_0^\theta \frac{1}{\theta} \left(\frac{\partial}{\partial \theta} \log \left(\frac{1}{\theta} \right) \right)^2 dx = \frac{1}{\theta} \int_0^\theta dx \cdot \left(-\frac{1}{\theta} \right)^2 = \frac{1}{\theta^2}.$$

Damit hätten wir

$$\text{Var}_\theta \hat{\theta} \geq \frac{\theta^2}{n}.$$

Betrachten wir

$$\hat{\theta}(X_1, \dots, X_n) = \frac{n+1}{n} \max\{X_1, \dots, X_n\} = \frac{n+1}{n} X_{(n)}.$$

Zeigen wir, dass

$$\mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) = \theta \quad \text{und} \quad \text{Var}_\theta \hat{\theta}(X_1, \dots, X_n) < \frac{\theta^2}{n}.$$

Berechnen wir dazu $\mathbb{E}_\theta X_{(n)}^k$, $k \in \mathbb{N}$. Es gilt

$$F_{X_{(n)}}(x) = F_{X_i}^n(x) = \begin{cases} \frac{x^n}{\theta^n}, & x \in [0, \theta], \\ 1, & x \geq \theta, \\ 0, & x < 0, \end{cases}$$

$$f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} \cdot \mathbb{I}(x \in [0, \theta]),$$

$$\mathbb{E}_\theta X_{(n)}^k = \int_0^\theta x^k \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \int_0^\theta x^{n+k-1} dx = \frac{n \cdot \theta^{n+k}}{\theta^n \cdot (n+k)} = \frac{n\theta^k}{n+k}.$$

Damit folgt

$$\mathbb{E}_\theta \hat{\theta} = \frac{n+1}{n} \cdot \mathbb{E}_\theta X_{(n)} = \frac{n+1}{n} \cdot \frac{n\theta}{n+1} = \theta,$$

das heißt, $\hat{\theta}$ ist erwartungstreu. Weiterhin gilt

$$\begin{aligned} \text{Var}_\theta \hat{\theta} &= \left(\frac{n+1}{n}\right)^2 \cdot \text{Var}_\theta X_{(n)} = \left(\frac{n+1}{n}\right)^2 \cdot \left(\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}\right)^2 \\ &= \frac{(n+1)^2}{n^2} \cdot \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2} \cdot \theta^2 \\ &= \frac{\theta^2}{n(n+2)}(n^2 + 2n + 1 - n^2 - 2n) = \frac{\theta^2}{n(n+2)} \end{aligned}$$

und somit

$$\text{Var}_\theta \hat{\theta} = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}.$$

3.5.2 Bedingte Erwartung

Seien X und Y zwei Zufallsvariablen, wobei Y eine absolut stetige Verteilung besitzt. Dann folgt $\mathbb{P}(Y = y) = 0 \quad \forall y \in \mathbb{R}$. Deshalb kann die bedingte Wahrscheinlichkeit $\mathbb{P}(X \in B | Y = y)$ auf dem gewöhnlichen Wege

$$\mathbb{P}(X \in B | Y = y) = \frac{\mathbb{P}(X \in B, Y = y)}{\mathbb{P}(Y = y)}$$

nicht definiert werden. Aus der Praxis ist aber eine Reihe von Fragestellungen bekannt (z.B. Bayessche Analyse), in denen Wahrscheinlichkeiten $\mathbb{P}(X \in B | Y = y)$ ausgewertet werden müssen. Deswegen werden wir eine neue Definition der bedingten Wahrscheinlichkeit geben, die solche Situationen berücksichtigt. Diese Definition erfolgt durch die Definition der bedingten Erwartung.

Schema:

1. Es wird die bedingte Erwartung von der Zufallsvariablen X bzgl. der σ -Algebra \mathcal{B} als Zufallsvariable $\mathbb{E}(X|\mathcal{B})$ eingeführt, wobei \mathcal{B} eine Teil- σ -Algebra von \mathcal{F} und $(\Omega, \mathcal{F}, \mathbb{P})$ der Wahrscheinlichkeitsraum ist.

2. Die bedingte Erwartung von X unter der Bedingung Y wird als $\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma_Y)$ eingeführt, wobei σ_Y die von Y erzeugte σ -Algebra ist.
3. $\mathbb{P}(X \in B|Y = y)$ wird als Zufallsvariable $\mathbb{E}(\mathbb{I}(X \in B)|Y)$ auf der Menge $\{\omega \in \Omega : Y(\omega) = y\}$ eingeführt.

Gehen wir nun dieses Schema im Detail durch:

1. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und \mathcal{B} eine Teil- σ -Algebra von \mathcal{F} , d.h. $\mathcal{B} \subseteq \mathcal{F}$.

Definition 3.5.1

Der *bedingte Erwartungswert* einer Zufallsvariablen X definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ bezüglich einer σ -Algebra $\mathcal{B} \subseteq \mathcal{F}$ ist in dem Fall $\mathbb{E}|X| < \infty$ als eine \mathcal{B} -messbare Zufallsvariable Y definiert, die die Eigenschaft

$$\int_B Y(\omega) \mathbb{P}(d\omega) = \int_B X \mathbb{P}(d\omega), \quad \forall B \in \mathcal{B}$$

besitzt. Dabei wird die Bezeichnung $Y = \mathbb{E}(X|\mathcal{B})$ verwendet.

Warum existiert diese Zufallsvariable Y ?

- Zerlegen wir X in den positiven X_+ und negativen X_- Anteil $X = X_+ - X_-$ und beweisen die Existenz von $\mathbb{E}(X_{\pm}|\mathcal{B})$. Danach setzen wir $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X_+|\mathcal{B}) - \mathbb{E}(X_-|\mathcal{B})$.
- Somit genügt es zu zeigen, dass der Erwartungswert $\mathbb{E}(X|\mathcal{B})$ einer nicht negativen Zufallsvariablen $X \geq 0$ fast sicher existiert.
- Sei $Q(B) = \int_B X(\omega) \mathbb{P}(d\omega)$. Man kann zeigen, dass $Q(\cdot)$ ein Maß auf (Ω, \mathcal{F}) ist. Dabei folgt aus $\mathbb{P}(B) = 0$ die Gleichheit $Q(B) = 0$ für $B \in \mathcal{B}_{\mathbb{R}}$ (bzw. $B \in \mathcal{B}$). Somit ist Q absolut stetig bzgl. \mathbb{P} . Weiter existiert nach dem Satz von Radon-Nikodym eine Dichte $Y(\omega)$, die messbar bzgl. \mathcal{B} ist und für die

$$Q(B) = \int_B Y(\omega) \mathbb{P}(d\omega) \implies Y(\omega) = \mathbb{E}(X|\mathcal{B})$$

gilt.

Bemerkung 3.5.1

Aus der obigen Beweisskizze wird ersichtlich, dass $Y(\omega) = \mathbb{E}(X|\mathcal{B})$ nur \mathbb{P} -fast sicher definiert ist. Somit kann man mehrere Versionen von $Y(\omega)$ angeben, die sich auf einer Menge der Wahrscheinlichkeit 0 unterscheiden.

Satz 3.5.2 (Eigenschaften des bedingten Erwartungswertes):

Seien X und Y Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit der Eigenschaft $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$ und $\mathbb{E}|XY| < \infty$ (dies kann noch ein wenig abgeschwächt werden, ist hier allerdings ausreichend). Seien \mathcal{B} , \mathcal{B}_1 und \mathcal{B}_2 Teil- σ -Algebren von \mathcal{F} . Es gelten folgende Eigenschaften (im fast sicheren Sinne):

- a) $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}X$, $\mathbb{E}(X|\mathcal{F}) = X$ fast sicher.
- b) Falls $X \leq Y$ fast sicher, dann gilt ebenso $\mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(Y|\mathcal{B})$ fast sicher.
- c) Es gilt $\mathbb{E}(XY|\mathcal{B}) = X \cdot \mathbb{E}(Y|\mathcal{B})$, falls X \mathcal{B} -messbar ist.
- d) $\mathbb{E}(c|\mathcal{B}) = c$ für $c = \text{const.}$

- e) Es gilt $\mathbb{E}(\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1) = \mathbb{E}(X|\mathcal{B}_1)$ und $\mathbb{E}(\mathbb{E}(X|\mathcal{B}_1)|\mathcal{B}_2) = \mathbb{E}(X|\mathcal{B}_1)$, falls $\mathcal{B}_1 \subseteq \mathcal{B}_2$.
- f) Falls X unabhängig von \mathcal{B} ist (d.h., die σ -Algebren $\sigma_X = X^{-1}(\mathcal{B}_{\mathbb{R}})$ und \mathcal{B} sind unabhängig), dann gilt $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}X$.

Ohne Beweis (siehe Beweis in [26]).

Beispiel 3.5.3

Sei $\mathcal{B} = \sigma(\{A_1, \dots, A_n\})$, wobei $\{A_1, \dots, A_n\}$ eine messbare Zerlegung des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, \mathbb{P})$ ist, d.h. $\bigcup_{i=1}^n A_i = \Omega$, $A_i \cap A_j = \emptyset$, $i \neq j$, $\mathbb{P}(A_i) > 0$, $i = 1, \dots, n$. Was ist $\mathbb{E}(X|\mathcal{B})$? Da $\mathbb{E}(X|\mathcal{B})$ \mathcal{B} -messbar ist, können wir die allgemeine Form der Funktionen ausnutzen, die messbar bzgl. einer endlich erzeugten σ -Algebra $\mathcal{B} = \sigma(\{A_1, \dots, A_n\})$ sind: $\mathbb{E}(X(\omega)|\mathcal{B}) = \sum_{i=1}^n k_i \mathbb{I}(\omega \in A_i)$ (ohne Beweis).

Berechnen wir k_i : Aus der Definition 3.5.1 folgt für $B = A_j$

$$\begin{aligned} \int_B \mathbb{E}(X|\mathcal{B}) \mathbb{P}(d\omega) &= \int_{A_j} \sum_{i=1}^n k_i \cdot \mathbb{I}(\omega \in A_i) \mathbb{P}(d\omega) = k_j \cdot \mathbb{P}(A_j) \\ &= \int_B X \mathbb{P}(d\omega) = \int_{A_j} X \mathbb{P}(d\omega) = \mathbb{E}(X \cdot \mathbb{I}_{A_j}) \\ \implies k_j &= \frac{\mathbb{E}(X \cdot \mathbb{I}_{A_j})}{\mathbb{P}(A_j)}, \quad j = 1, \dots, n. \\ \implies \mathbb{E}(X(\omega)|\mathcal{B}) &= \frac{\mathbb{E}(X \cdot \mathbb{I}_{A_j})}{\mathbb{P}(A_j)}, \quad \text{falls } \omega \in A_j, \quad j = 1, \dots, n. \end{aligned}$$

2. *Bedingte Erwartung bzgl. einer Zufallsvariablen Y :*

Definition 3.5.2

Seien X und Y zwei Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Der *bedingte Erwartungswert von X unter der Bedingung Y* wird als $\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma_Y)$ eingeführt, wobei σ_Y die von Y erzeugte σ -Algebra ist: $\sigma_Y = Y^{-1}(\mathcal{B}_{\mathbb{R}})$.

Lemma 3.5.1

Es existiert eine Borel-messbare Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$, für die gilt, dass $\mathbb{E}(X|Y) = g(Y)$ fast sicher (Ohne Beweis).

Daher wird die Schreibweise $\mathbb{E}(X|Y = y)$ als $g(y)$ verstanden: $\mathbb{E}(X|Y = y) = g(y)$ oder $\mathbb{E}(X|Y = y)$ ist der Wert von $\mathbb{E}(X|Y)$ auf der Menge $\{\omega \in \Omega : Y(\omega) = y\}$.

3. *Bedingte Wahrscheinlichkeit bzgl. einer σ -Algebra bzw. einer Zufallsvariable.*

Definition 3.5.3

Die *bedingte Wahrscheinlichkeit* von $A \in \mathcal{F}$ unter der Bedingung \mathcal{B} ist gegeben durch $\mathbb{P}(A|\mathcal{B}) = \mathbb{E}(\mathbb{I}_A|\mathcal{B})$ fast sicher. Analog dazu definieren wir $\mathbb{P}(A|Y) = \mathbb{E}(\mathbb{I}_A|Y)$ für eine Zufallsvariable Y .

Bemerkung 3.5.2

Die so definierte Familie von Zufallsvariablen $\mathbb{P}(\cdot|\mathcal{B})$ erfüllen (fast sicher) nicht die Eigenschaften eines Maßes: Es gilt

$$0 \leq \mathbb{P}(A|\mathcal{B}) \leq 1, \quad \forall A \in \mathcal{F} \text{ fast sicher,}$$

aber die Eigenschaft der σ -Additivität

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \mid \mathcal{B}\right) \stackrel{\text{f.s.}}{=} \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid \mathcal{B})$$

für disjunkte $\{A_i\}$ hängt von der Version $\mathbb{P}(\cdot \mid \mathcal{B})$ ab. Das bedeutet, es existiert kein $M \in \mathcal{F} : \mathbb{P}(M) = 0$, so dass die obige Eigenschaft für alle $\omega \in M^C$ gilt.

3.5.3 Suffizienz

Sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen identisch verteilten Zufallsvariablen X_i mit Verteilungsfunktion F_θ , $\theta \in \Theta \subseteq \mathbb{R}^m$. Wenn man von der vollen Information $\{X_1 = x_1, \dots, X_n = x_n\}$ zum Schätzer $\hat{\theta}(X_1, \dots, X_n)$ des Parameters θ übergeht, dann entsteht durch die Abbildung

$$\hat{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad m \ll n$$

ein Informationsverlust, weil man normalerweise (X_1, \dots, X_n) nicht aus $\hat{\theta}(X_1, \dots, X_n)$ zurückrechnen kann. Die sogenannten *suffizienten* Schätzer minimieren diesen Informationsverlust im stochastischen Sinne:

Definition 3.5.4

1. Seien Zufallsvariablen X_1, \dots, X_n und $\hat{\theta}(X_1, \dots, X_n)$ diskret verteilt. Ein Schätzer $\hat{\theta}$ des Parameters θ heißt *suffizient*, falls

$$\mathbb{P}_\theta \left(X_1 = x_1, \dots, X_n = x_n \mid \hat{\theta}(X_1, \dots, X_n) = t \right)$$

nicht von θ abhängt für beliebige x_1, \dots, x_n und t aus den Trägern der Zähldichten von (X_1, \dots, X_n) bzw. $\hat{\theta}(X_1, \dots, X_n)$.

2. Falls X_1, \dots, X_n und $\hat{\theta}(X_1, \dots, X_n)$ absolut stetig verteilt sind, dann heißt der Schätzer $\hat{\theta}$ *suffizient* für θ , falls die Wahrscheinlichkeit

$$\mathbb{P} \left((X_1, \dots, X_n) \in B \mid \hat{\theta}(X_1, \dots, X_n) = t \right)$$

für beliebige $B \in \mathcal{B}_{\mathbb{R}^n}$ und $t \in \text{supp} f_{\hat{\theta}}$ nicht von $\theta \in \Theta$ abhängt, wobei $f_{\hat{\theta}}$ die Dichte von $\hat{\theta}$ ist.

Bemerkung 3.5.3

1. Betrachten wir im diskreten Fall die bedingte Likelihood-Funktion

$$L_{\hat{\theta}}(x_1, \dots, x_n, \theta) = \mathbb{P}_\theta \left(X_1 = x_1, \dots, X_n = x_n \mid \hat{\theta}(X_1, \dots, X_n) = t \right).$$

Aus Definition 3.5.4 folgt, dass wir keinen neuen ML-Schätzer für θ aus dieser bedingten Likelihood $L_{\hat{\theta}}(x_1, \dots, x_n, \theta)$ gewinnen werden können, da sie nicht von θ abhängt. Das heißt, der Schätzer $\hat{\theta}$ enthält bereits die volle Information über θ , die man aus der Stichprobe (x_1, \dots, x_n) gewinnen kann.

2. Falls $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$ eine bijektive Borel-messbare Abbildung und $\hat{\theta}(X_1, \dots, X_n)$ ein suffizienter Schätzer von $\theta \in \Theta \subset \mathbb{R}^m$ ist, dann ist der Schätzer $g(\hat{\theta}(X_1, \dots, X_n))$ auch ein suffizienter Schätzer für θ . Dies wird aus der Tatsache ersichtlich, dass

$$\{\omega \in \Omega : g(\hat{\theta}(X_1, \dots, X_n)) = t\} = \{\omega \in \Omega : \hat{\theta}(X_1, \dots, X_n) = g^{-1}(t)\}, \quad \forall t.$$

Lemma 3.5.2 (Suffizienz):

Seien Zufallsvariablen X_1, \dots, X_n und $\hat{\theta}(X_1, \dots, X_n)$ entweder alle diskret oder absolut stetig verteilt mit den Likelihood-Funktionen

$$L(x_1, \dots, x_n, \theta) = \begin{cases} \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n), & \text{im diskreten Fall,} \\ f_{X_1, \dots, X_n}(x_1, \dots, x_n), & \text{im absolut stetigen Fall,} \end{cases}$$

$$L_{\hat{\theta}}(t, \theta) = \begin{cases} \mathbb{P}_\theta(\hat{\theta}(X_1, \dots, X_n) = t), & \text{im diskreten Fall,} \\ f_{\hat{\theta}}(t, \theta), & \text{im absolut stetigen Fall.} \end{cases}$$

Seien die Träger um L bzw. $L_{\hat{\theta}}$ gegeben durch

$$\begin{aligned} \text{supp}L &= \{(x_1, \dots, x_n) \in \mathbb{R}^n : L(x_1, \dots, x_n, \theta) > 0\}, \\ \text{supp}L_{\hat{\theta}} &= \{t \in \mathbb{R} : L_{\hat{\theta}}(t, \theta) > 0\}. \end{aligned}$$

Der Schätzer $\hat{\theta}$ ist suffizient für θ genau dann, wenn

$$\frac{L(x_1, \dots, x_n, \theta)}{L_{\hat{\theta}}(\hat{\theta}(x_1, \dots, x_n), \theta)} \tag{3.5.1}$$

nicht von θ abhängig ist für alle $(x_1, \dots, x_n) \in \text{supp}L$.

Beweis Wir beweisen lediglich den diskreten Fall:

„ \implies “ Ist $\hat{\theta}$ suffizient, so überprüfen wir, ob damit folgt, dass (3.5.1) von θ abhängt für alle $(x_1, \dots, x_n) \in \text{supp}L$. Es gilt:

$$\begin{aligned} \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n \mid \hat{\theta}(X_1, \dots, X_n) = t) \\ &= \frac{\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n, \hat{\theta}(X_1, \dots, X_n) = t)}{\mathbb{P}_\theta(\hat{\theta}(X_1, \dots, X_n) = t)} \\ &= \begin{cases} 0, & \text{falls } \hat{\theta}(x_1, \dots, x_n) \neq t \\ \frac{\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n)}{\mathbb{P}_\theta(\hat{\theta}(X_1, \dots, X_n) = \hat{\theta}(x_1, \dots, x_n))}, & \text{falls } \hat{\theta}(x_1, \dots, x_n) = t. \end{cases} \end{aligned}$$

Somit hängt (3.5.1) nicht von θ ab.

„ \impliedby “ Folgt aus dem 1. Fall durch Betrachtung von hinten.

□

Beispiel 3.5.4

1. *Bernoulli-Verteilung*: Seien $X_i \sim \text{Bernoulli}(p)$, $p \in [0, 1]$, $i = 1, \dots, n$, $\hat{p} = \bar{X}_n$ ein erwartungstreuer Schätzer für p . Wir zeigen nun, dass \hat{p} suffizient ist. Es gilt

$$\hat{p} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} Y,$$

wobei $Y \sim \text{Bin}(n, p)$. Es genügt nach Bemerkung 3.5.3 2) zu zeigen, dass Y ein suffizienter Schätzer für p ist. Nach Lemma 3.5.2 gilt für $x_i \in \{0, 1\}$, $i = 1, \dots, n$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}.$$

Definieren wir nun L_Y als

$$L_Y(y, p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, \dots, n.$$

Setzen wir nun statt y die Summe $\sum_{i=1}^n x_i$ ein und betrachten

$$\frac{L(x_1, \dots, x_n, p)}{L_Y(\sum_{i=1}^n x_i, p)} = \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}}{\binom{n}{\sum_{i=1}^n x_i} p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}} = \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}.$$

Dies hängt offensichtlich nicht von p ab, womit folgt, dass Y und somit \hat{p} suffizient sind.

2. *Normalverteilung mit bekannter Varianz*: Seien $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$, σ^2 bekannt. So ist $\hat{\mu} = \bar{X}_n$ ein erwartungstreuer Schätzer für μ . Zeigen wir nun, dass $\hat{\mu}$ suffizient ist: Betrachten wir

$$\begin{aligned} L(x_1, \dots, x_n, \mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma}\right)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

und nach Lemma 2.2.1

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \mu)^2}{2\sigma^2}\right).$$

Ferner gilt bekanntermaßen $\hat{\mu} \sim N(\mu, \sigma^2/n)$, und somit

$$\begin{aligned} L_{\hat{\mu}}(x, \mu) &= \frac{\sqrt{n}}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{n}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right), \\ \frac{L(x_1, \dots, x_n, \mu)}{L_{\hat{\mu}}(\bar{x}_n, \mu)} &= \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \cdot \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - \mu)^2}{2\sigma^2}\right)}{\frac{\sqrt{n}}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{n(\bar{x}_n - \mu)^2}{2\sigma^2}\right)} \\ &= \frac{\sqrt{n}}{(2\pi\sigma^2)^{n/2-1}} \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right), \end{aligned}$$

was von μ unabhängig ist. Somit folgt nach Lemma 3.5.2, dass $\hat{\mu} = \bar{X}_n$ ein suffizienter Schätzer für μ ist.

Mit Hilfe des nächsten Satzes von Neyman-Fisher wird es möglich sein zu zeigen, dass bei unbekannter Varianz der Schätzer (\bar{X}_n, S_n^2) für (μ, σ^2) suffizient ist.

Satz 3.5.3 (Faktorisierungssatz von Neyman-Fisher):

Unter den Voraussetzungen von Lemma 3.5.2 ist $\hat{\theta}(X_1, \dots, X_n)$ ein suffizienter Schätzer für θ genau dann, wenn zwei messbare Funktionen $g : \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}$ und $h : \mathbb{R}^n \rightarrow \mathbb{R}$ existieren, so dass folgende Faktorisierung der Likelihood-Funktion $L(x_1, \dots, x_n, \theta)$ der Stichprobe (X_1, \dots, X_n) gilt:

$$L(x_1, \dots, x_n, \theta) = g(\hat{\theta}(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in \text{supp}L, \quad \theta \in \Theta.$$

Beweis Wir beweisen nur den diskreten Fall.

1. Falls $\hat{\theta}$ suffizient ist, dann hängt nach Lemma 3.5.2

$$\frac{L(x_1, \dots, x_n, \theta)}{\underbrace{L_{\hat{\theta}}(\hat{\theta}(x_1, \dots, x_n), \theta)}_{=g(\hat{\theta}(x_1, \dots, x_n), \theta)}} = h(x_1, \dots, x_n)$$

nicht von θ ab. Somit bekommen wir die Faktorisierung von Neyman-Fisher.

2. Sei nun $L(x_1, \dots, x_n, \theta) = g(\hat{\theta}(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)$ für alle $(x_1, \dots, x_n) \in \text{supp}L$, $\theta \in \Theta$. Führen wir eine Menge

$$C = \{(y_1, \dots, y_n) \in \mathbb{R}^n : \hat{\theta}(y_1, \dots, y_n) = \hat{\theta}(x_1, \dots, x_n)\} = \hat{\theta}^{-1}(\hat{\theta}(x_1, \dots, x_n))$$

ein. So gilt

$$\begin{aligned} \frac{\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n)}{\underbrace{L_\theta(\hat{\theta}(x_1, \dots, x_n), \theta)}_{=\mathbb{P}_\theta(\hat{\theta}(X_1, \dots, X_n) = \hat{\theta}(x_1, \dots, x_n))}} &= \frac{g(\hat{\theta}(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n) \in C} \mathbb{P}_\theta(X_1 = y_1, \dots, X_n = y_n)} \\ &= \frac{g(\hat{\theta}(x_1, \dots, x_n), \theta) \cdot h(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n) \in C} \underbrace{g(\hat{\theta}(y_1, \dots, y_n), \theta)}_{=\hat{\theta}(x_1, \dots, x_n)} \cdot h(y_1, \dots, y_n)} \\ &= \frac{h(x_1, \dots, x_n)}{\sum_{(y_1, \dots, y_n) \in C} h(y_1, \dots, y_n)}, \end{aligned}$$

welches nicht von θ abhängt. Daher ist $\hat{\theta}$ nach Lemma 3.5.2 suffizient. □

Beispiel 3.5.5

1. *Poisson-Verteilung:* Seien $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$, $\hat{\lambda} = \bar{X}_n$ ein erwartungstreuer Schätzer für λ . Zeigen wir, dass $\hat{\lambda}$ suffizient ist. Es gilt für $x_i \in \{0, 1, 2, \dots\}$, $i = 1, \dots, n$

$$\begin{aligned} L(x_1, \dots, x_n, \lambda) &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \frac{e^{-\lambda n} \cdot \lambda^{\sum_{i=1}^n x_i}}{x_1! \cdot \dots \cdot x_n!} = \frac{e^{-n\lambda} \lambda^{n\bar{x}_n}}{x_1! \cdot \dots \cdot x_n!} \\ &= g(\bar{x}_n, \lambda) \cdot h(x_1, \dots, x_n), \end{aligned}$$

wobei $g(\bar{x}_n, \lambda) = e^{-n\lambda} \cdot \lambda^{n\bar{x}_n}$, $h(x_1, \dots, x_n) = \frac{1}{x_1! \cdot \dots \cdot x_n!}$ ist. Somit ist $\hat{\lambda} = \bar{X}_n$ nach Satz 3.5.3 suffizient.

2. *Exponentialverteilung:* Seien $X_i \sim \text{Exp}(\lambda)$, $\lambda > 0$, $\hat{\lambda} = \bar{X}_n^{-1}$ ein Momentenschätzer für λ , der zwar nicht erwartungstreu ist, jedoch stark konsistent, denn $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{f.s.}} \mathbb{E}X_i = \frac{1}{\lambda}$ nach dem starken Gesetz der großen Zahlen. Zeigen wir, dass $\hat{\lambda}$ suffizient ist. Für $x_1 \geq 0, \dots, x_n \geq 0$ gilt

$$\begin{aligned} L(x_1, \dots, x_n, \lambda) &= \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = \lambda^n e^{-\lambda n \bar{x}_n} \\ &= \lambda^n e^{-\frac{\lambda n}{\hat{\lambda}}} = g(\hat{\lambda}, \lambda) \cdot \underbrace{h(x_1, \dots, x_n)}_{=1}, \end{aligned}$$

wobei $g(\hat{\lambda}, \lambda) = \lambda^n e^{-\frac{\lambda n}{\hat{\lambda}}}$ und $h(x_1, \dots, x_n) \equiv 1$ ist. Somit ist $\hat{\lambda}$ nach dem Satz 3.5.3 suffizient.

Übungsaufgabe 3.5.2

Zeigen Sie mit Hilfe des Satzes 3.5.3, dass der Schätzer (\bar{X}_n, S_n^2) suffizient für (μ, σ^2) im Falle der normal und unabhängig identisch verteilten Stichprobe (X_1, \dots, X_n) , $X_i \sim N(\mu, \sigma^2)$ ist.

Bemerkung 3.5.4

Der Vorteil des Satzes von Neyman-Fisher ist, dass man für die Überprüfung der Suffizienz-eigenschaft von $\hat{\theta}$ die Likelihood-Funktion von $\hat{\theta}$ nicht explizit zu kennen braucht. Dies ist insbesondere in den Fällen vorteilhaft, in denen der Schätzer $\hat{\theta}$ kompliziert ist und seine Likelihood-Funktion nicht analytisch angegeben werden kann (bzw. unbekannt ist).

3.5.4 Vollständigkeit**Definition 3.5.5**

Ein Schätzer $\hat{\theta}(X_1, \dots, X_n)$ des Parameters $\theta \in \Theta \subset \mathbb{R}^m$ heißt *vollständig*, falls für beliebige messbare Funktionen $g: \mathbb{R}^m \rightarrow \mathbb{R}$ mit der Eigenschaft $\mathbb{E}_\theta g(\hat{\theta}(X_1, \dots, X_n)) = 0$, $\theta \in \Theta$ folgt

$$g(\hat{\theta}(X_1, \dots, X_n)) \stackrel{\text{f.s.}}{\equiv} 0.$$

Bemerkung 3.5.5

1. Seien $g_1, g_2: \mathbb{R}^m \rightarrow \mathbb{R}$ Funktionen, für die $\forall \theta \in \Theta$ gilt

$$\mathbb{E}_\theta |g_i(\hat{\theta}(X_1, \dots, X_n))| < \infty, \quad \mathbb{E}_\theta g_1(\hat{\theta}(X_1, \dots, X_n)) = \mathbb{E}_\theta g_2(\hat{\theta}(X_1, \dots, X_n)),$$

wobei $\hat{\theta}$ vollständig ist. So folgt aus der Definition 3.5.5

$$g_1(\hat{\theta}(X_1, \dots, X_n)) = g_2(\hat{\theta}(X_1, \dots, X_n))$$

fast sicher (nehme $g = g_1 - g_2$).

Fazit: Die Eigenschaft der Vollständigkeit erlaubt aus dem Vergleich der Schätzer $g_1(\hat{\theta})$ und $g_2(\hat{\theta})$ im Mittel eine Aussage über ihre fast sichere Gleichheit zu machen.

2. Falls $\hat{\theta}$ ein vollständiger Schätzer für θ ist, dann ist auch $g(\hat{\theta})$ ein vollständiger Schätzer für θ für eine beliebige messbare Funktion $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$.

Beispiel 3.5.6

1. *Bernoulli-Verteilung:* Seien $X_i \sim \text{Bernoulli}(p)$, $p \in [0, 1]$. Zeigen wir, dass $\hat{p} = \bar{X}_n$ vollständig ist:

Sei g eine beliebige Funktion $\mathbb{R} \rightarrow \mathbb{R}$. Es genügt zu zeigen, dass $Y = \sum_{i=1}^n X_i$ vollständig ist. Es gilt $Y \sim \text{Bin}(n, p)$, womit folgt, dass

$$\mathbb{E}_p g(Y) = \sum_{k=0}^n g(k) \binom{n}{k} p^k (1-p)^{n-k}.$$

Weiter gilt $\mathbb{E}_p g(Y) = 0$ genau dann, wenn

$$\sum_{k=0}^n g(k) \binom{n}{k} \underbrace{\left(\frac{p}{1-p} \right)^k}_{=t} = p_n(t) = 0$$

für $p \in (0, 1)$, also $t \in (0, \infty)$. $p_n(t)$ ist ein Polynom des Grades n , womit folgt

$$g(k) \binom{n}{k} = 0 \quad \text{für alle } k \implies g(k) = 0, \quad k = 0, \dots, n \implies g(Y) = 0 \quad \mathbb{P}_p\text{-fast sicher.}$$

Somit ist Y vollständig und daher auch $\hat{p} = \bar{X}_n$.

2. *Gleichverteilung:* Sei $X_i \sim U[0, \theta]$, $i = 1, \dots, n$. Wie wir bereits gezeigt haben, ist der Schätzer $\hat{\theta}(X_1, \dots, X_n) = \frac{n+1}{n} X_{(n)}$ erwartungstreu. Zeigen wir nun, dass er ein vollständiger Schätzer ist. Es genügt zu zeigen, dass $X_{(n)} = \max_{i=1, \dots, n} X_i$ vollständig ist. Es ist zu zeigen, dass für alle messbaren $g : \mathbb{R} \rightarrow \mathbb{R}$ aus $\mathbb{E}_\theta g(X_{(n)}) = 0$ folgt $g(X_{(n)}) = 0$ fast sicher. Die Dichte von $X_{(n)}$ ist nach Beispiel 3.5.2 gegeben durch $f_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} \cdot \mathbb{I}_{[0, \theta]}(x)$.

$$\begin{aligned} 0 &= \frac{d}{d\theta} \mathbb{E}_\theta g(X_{(n)}) = \frac{d}{d\theta} \int_0^\theta g(x) f_{X_{(n)}}(x) dx = \frac{d}{d\theta} \frac{1}{\theta^n} \int_0^\theta nx^{n-1} g(x) dx \\ &= -n \frac{1}{\theta^{n+1}} \int_0^\theta g(x) nx^{n-1} dx + \frac{1}{\theta^n} n \theta^{n-1} g(\theta) = -\frac{n}{\theta} \underbrace{\mathbb{E}_\theta g(X_{(n)})}_{=0} + \frac{n}{\theta} g(\theta) \\ &= \frac{n}{\theta} g(\theta) = 0, \quad \theta > 0 \implies g(x) = 0, \quad x > 0. \end{aligned}$$

Daher gilt $g(X_{(n)}) = 0$ fast sicher.

3.5.5 Bester erwartungstreuer Schätzer

Aus Definition 3.3.7 folgt: Sei (X_1, \dots, X_n) eine Zufallsstichprobe, $X_i \sim F_\theta$, $\theta \in \Theta \subset \mathbb{R}$ ($m = 1$), X_i unabhängig identisch verteilte Zufallsvariablen. Dann heißt $\hat{\theta}(X_1, \dots, X_n)$ bester erwartungstreuer Schätzer, falls

$$\mathbb{E}_\theta \hat{\theta}^2(X_1, \dots, X_n) < \infty \quad \mathbb{E}_\theta \hat{\theta}(X_1, \dots, X_n) = \theta, \quad \theta \in \Theta.$$

und $\hat{\theta}$ die minimale Varianz unter allen erwartungstreuen Schätzern besitzt.

Lemma 3.5.3 (Eindeutigkeit der besten erwartungstreuen Schätzer):

Falls $\hat{\theta}$ ein bester erwartungstreuer Schätzer für θ ist, dann ist er eindeutig bestimmt.

Beweis Sei $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ ein bester erwartungstreuer Schätzer für θ und $\tilde{\theta}$ ein weiterer bester erwartungstreuer Schätzer für θ . Zeigen wir, dass $\hat{\theta} = \tilde{\theta}$.

Ex adverso: Nehmen wir an, dass $\hat{\theta} \neq \tilde{\theta}$ ist und betrachten $\theta^* = 1/2(\hat{\theta} + \tilde{\theta})$. Offensichtlich ist θ^* erwartungstreu. Untersuchen wir

$$\text{Var}_\theta \theta^* = \frac{1}{4} \text{Var}_\theta (\hat{\theta} + \tilde{\theta}) = \frac{1}{4} \text{Var}_\theta \hat{\theta} + \frac{1}{4} \text{Var}_\theta \tilde{\theta} + \frac{1}{2} \text{Cov}_\theta (\hat{\theta}, \tilde{\theta}).$$

Da $\hat{\theta}, \tilde{\theta}$ beste erwartungstreue Schätzer sind und mit der Ungleichung von Cauchy-Schwarz $|\text{Cov}_\theta (\hat{\theta}, \tilde{\theta})| \leq \sqrt{\text{Var}_\theta \hat{\theta} \cdot \text{Var}_\theta \tilde{\theta}} = \text{Var}_\theta \hat{\theta}$ gilt, folgt

$$\text{Var}_\theta \theta^* \leq \frac{1}{2} \text{Var}_\theta \hat{\theta} + \frac{1}{2} \text{Var}_\theta \tilde{\theta} = \text{Var}_\theta \hat{\theta}.$$

Da $\hat{\theta}$ der beste erwartungstreue Schätzer ist folgt $\text{Var}_\theta \theta^* = \text{Var}_\theta \hat{\theta}$ und somit $\rho(\hat{\theta}, \tilde{\theta}) = 1 \implies \hat{\theta}$ und $\tilde{\theta}$ sind linear abhängig, d.h. es existieren Konstanten a und b , für die gilt $\hat{\theta} = a\tilde{\theta} + b$. Es folgt $a = 1$ aus $\text{Var}_\theta \hat{\theta} = a^2 \text{Var}_\theta \tilde{\theta} = \text{Var}_\theta \hat{\theta}$ und $b = 0$, weil $\hat{\theta}$ und $\tilde{\theta}$ erwartungstreu sind: $\theta = \mathbb{E}_\theta \hat{\theta} = \mathbb{E}_\theta \tilde{\theta} + b = \theta + b$. Das bedeutet, dass $\hat{\theta} = \tilde{\theta}$. \square

Lemma 3.5.4

Ein erwartungstreuer Schätzer $\hat{\theta}$, dessen zweites Moment endlich ist, ist genau dann der beste erwartungstreue Schätzer für θ , wenn $\text{Cov}_\theta (\hat{\theta}, \varphi) = 0$, $\theta \in \Theta$ für eine beliebige Stichprobenfunktion $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ mit der Eigenschaft $\mathbb{E}_\theta \varphi(X_1, \dots, X_n) = 0$, $\forall \theta \in \Theta$.

Beweis Wir beweisen den Satz für beide Richtungen getrennt:

„ \implies “ Sei $\hat{\theta}$ der beste erwartungstreue Schätzer für θ , $\varphi(X_1, \dots, X_n)$ eine Stichprobenfunktion mit $\mathbb{E}_\theta \varphi(X_1, \dots, X_n) = 0$, $\forall \theta \in \Theta$. So ist zu zeigen, dass $\text{Cov}_\theta (\hat{\theta}, \varphi) = \mathbb{E}_\theta (\hat{\theta} \varphi) = 0$, $\theta \in \Theta$ gilt.

Definieren wir $\tilde{\theta} = \hat{\theta} + a\varphi$, $a \in \mathbb{R}$. Berechnen wir

$$\text{Var}_\theta \tilde{\theta} = \text{Var}_\theta \hat{\theta} + a^2 \text{Var}_\theta \varphi + 2a \text{Cov}_\theta (\hat{\theta}, \varphi)$$

für $a \in \mathbb{R}$. Sei $g(a) = a^2 \text{Var}_\theta \varphi + 2a \text{Cov}_\theta (\varphi, \hat{\theta})$. Falls $\text{Cov}_\theta (\varphi, \hat{\theta}) \neq 0$, dann existiert ein $a \in \mathbb{R}$ mit $g(a) < 0$. Da $\tilde{\theta}$ ein erwartungstreuer Schätzer für θ ist ($\mathbb{E}_\theta \tilde{\theta} = \mathbb{E}_\theta \hat{\theta} + a \mathbb{E}_\theta \varphi = \theta + 0 = \theta$) folgt $\text{Var}_\theta \tilde{\theta} \geq \text{Var}_\theta \hat{\theta}$ für alle $a \in \mathbb{R}$. Dies ist jedoch ein Widerspruch mit $g(a) < 0$ für ein $a \in \mathbb{R}$. Damit folgt $\text{Cov}_\theta (\varphi, \hat{\theta}) = 0$, $\theta \in \Theta$.

„ \Leftarrow “ Sei $\hat{\theta}$ erwartungstreu, $\mathbb{E}_\theta \hat{\theta}^2 < \infty$, $\theta \in \Theta$, $\text{Cov}_\theta(\varphi, \hat{\theta}) = 0$, $\theta \in \Theta$, falls $\mathbb{E}_\theta \varphi = 0$, $\theta \in \Theta$. Sei $\tilde{\theta}$ ein anderer erwartungstreuer Schätzer für θ . Zeigen wir, dass $\text{Var}_\theta \tilde{\theta} \geq \text{Var}_\theta \hat{\theta}$. Es gilt

$$\tilde{\theta} = \hat{\theta} + \underbrace{(\tilde{\theta} - \hat{\theta})}_{=: \varphi}, \quad \mathbb{E}_\theta \varphi = \mathbb{E}_\theta \tilde{\theta} - \mathbb{E}_\theta \hat{\theta} = \theta - \theta = 0, \quad \forall \theta \in \Theta.$$

Somit

$$\text{Var}_\theta \tilde{\theta} = \text{Var}_\theta \hat{\theta} + \underbrace{\text{Var}_\theta \varphi}_{\geq 0} + 2 \underbrace{\text{Cov}_\theta(\hat{\theta}, \varphi)}_{=0} \geq \text{Var}_\theta \hat{\theta},$$

woraus folgt, dass $\hat{\theta}$ der beste Erwartungstreuer Schätzer für θ ist. □

Satz 3.5.4 (Lehmann-Scheffé):

Sei $\hat{\theta}$ ein erwartungstreuer vollständiger und suffizienter Schätzer für θ , $\mathbb{E}_\theta \hat{\theta}^2 < \infty$, $\forall \theta \in \Theta$. Dann ist $\hat{\theta}$ der beste erwartungstreue Schätzer für θ .

Beweis Nach Lemma 3.5.4 ist zu zeigen, dass $\text{Cov}_\theta(\hat{\theta}, \varphi) = \mathbb{E}_\theta(\hat{\theta}\varphi) = 0$, $\theta \in \Theta$, falls $\mathbb{E}_\theta \varphi = 0$, $\theta \in \Theta$. Es ist

$$\mathbb{E}_\theta(\hat{\theta}\varphi) = \mathbb{E}_\theta(\mathbb{E}(\hat{\theta}\varphi | \hat{\theta})) \stackrel{\hat{\theta} \text{ } \sigma(\hat{\theta})\text{-messbar}}{=} \mathbb{E}_\theta(\hat{\theta} \cdot \mathbb{E}_\theta(\varphi | \hat{\theta})) = \mathbb{E}_\theta(\hat{\theta} \cdot g(\hat{\theta})) \stackrel{?}{=} 0,$$

falls $g(\hat{\theta}) = 0$ fast sicher. Da $\hat{\theta}$ suffizient ist, ist $g(t) = \mathbb{E}_\theta(\varphi | \hat{\theta} = t)$ unabhängig von θ . Betrachten wir $\mathbb{E}_\theta g(\hat{\theta})$. Wir wollen zeigen, dass $\mathbb{E}_\theta g(\hat{\theta}) = 0$, $\theta \in \Theta$. Daraus und aus der Vollständigkeit von $\hat{\theta}$ wird folgen, dass $g(\hat{\theta}) = 0$ fast sicher für alle $\theta \in \Theta$.

$$\mathbb{E}_\theta g(\hat{\theta}) = \mathbb{E}_\theta(\mathbb{E}_\theta(\varphi | \hat{\theta})) = \mathbb{E}_\theta \varphi = 0$$

nach Voraussetzung. Somit folgt $\mathbb{E}_\theta(\varphi \hat{\theta}) = 0$ und $\hat{\theta}$ ist unkorreliert mit φ : $\mathbb{E}_\theta \varphi = 0$, $\theta \in \Theta$, womit folgt, dass nach Lemma 3.5.4 $\hat{\theta}$ der beste erwartungstreue Schätzer ist. □

Satz 3.5.5

Sei $\hat{\theta}$ ein erwartungstreuer Schätzer für θ , $\mathbb{E}_\theta \hat{\theta}^2 < \infty$, $\theta \in \Theta$. Sei $\tilde{\theta}$ ein vollständiger und suffizienter Schätzer für θ . Dann ist der Schätzer $\theta^* = \mathbb{E}(\hat{\theta} | \tilde{\theta})$ der beste erwartungstreue Schätzer für θ .

Beweis 1. Zeigen wir, dass $\mathbb{E}_\theta \theta^{*2} < \infty$ $\forall \theta \in \Theta$. Es gilt

$$\mathbb{E}_\theta (\theta^{*2}) = \mathbb{E}_\theta \left(\mathbb{E}(\hat{\theta} | \tilde{\theta}) \right)^2 \leq \mathbb{E}_\theta \left(\mathbb{E}(\hat{\theta}^2 | \tilde{\theta}) \right) = \mathbb{E}_\theta \hat{\theta}^2 < \infty,$$

da mit der Ungleichung von Jensen für bedingte Erwartung gilt

$$f(\mathbb{E}(X | \mathcal{B})) \stackrel{\text{f.s.}}{\leq} \mathbb{E}(f(X) | \mathcal{B})$$

für jede Zufallsvariable X , σ -Algebra \mathcal{B} und konvexe Funktion f .

2. Zeigen wir, dass θ^* erwartungstreu ist: $\mathbb{E}_\theta \theta^* = \mathbb{E}_\theta(\mathbb{E}(\hat{\theta} | \tilde{\theta})) = \mathbb{E}_\theta \hat{\theta} = \theta$, $\theta \in \Theta$, weil $\hat{\theta}$ erwartungstreu ist.

3. Nach Lemma 3.5.4 genügt es zu zeigen, dass $\mathbb{E}_\theta(\theta^*\varphi) = 0$ für $\theta \in \Theta$, falls $\mathbb{E}_\theta\varphi = 0$, $\theta \in \Theta$.

$$\begin{aligned}\mathbb{E}_\theta(\theta^*\varphi) &= \mathbb{E}_\theta(\underbrace{\mathbb{E}(\hat{\theta}|\tilde{\theta})}_{=g(\tilde{\theta}), \tilde{\theta} \text{ suf.}} \varphi) = \mathbb{E}_\theta(g(\tilde{\theta})\varphi) = \mathbb{E}_\theta(\mathbb{E}(g(\tilde{\theta})\varphi|\tilde{\theta})) \\ &\stackrel{g(\tilde{\theta}) \tilde{\theta}\text{-messbar}}{=} \mathbb{E}_\theta(g(\tilde{\theta}) \cdot \underbrace{\mathbb{E}(\varphi|\tilde{\theta})}_{=g_1(\tilde{\theta})}) = 0,\end{aligned}$$

falls $g_1(\tilde{\theta}) \stackrel{\text{f.s.}}{=} 0, \theta \in \Theta$. Zeigen wir, dass $\mathbb{E}_\theta g_1(\tilde{\theta}) = 0$. Es gilt $\mathbb{E}_\theta g_1(\tilde{\theta}) = \mathbb{E}_\theta(\mathbb{E}(\varphi|\tilde{\theta})) = \mathbb{E}_\theta\varphi = 0$ nach Voraussetzung. Daraus und aus der Vollständigkeit von $\tilde{\theta}$ folgt genauso wie im Beweis des Satzes 3.5.4, dass $g_1(\tilde{\theta}) = 0$ fast sicher. \square

Lemma 3.5.5 (Ungleichung von Blackwell-Rao):

Sei $\hat{\theta}$ ein erwartungstreuer Schätzer für θ , $\mathbb{E}_\theta\hat{\theta}^2 < \infty, \theta \in \Theta$. Sei $\tilde{\theta}$ ein suffizienter Schätzer für θ . Dann besitzt der erwartungstreue Schätzer $\theta^* := \mathbb{E}(\hat{\theta}|\tilde{\theta})$ eine Varianz, die kleiner oder gleich als $\text{Var}_\theta\hat{\theta}$ ist.

Beweis Siehe Beweis des Satzes 3.5.5. Dabei folgt die Erwartungstreue von θ^* aus Beweispunkt 2) des Satzes 3.5.5 und $\text{Var}_\theta\theta^* = \mathbb{E}_\theta\theta^{*2} - \theta^2 \leq \mathbb{E}_\theta\hat{\theta}^2 - \theta^2 = \text{Var}_\theta\hat{\theta}$ aus Beweispunkt 1) des Satzes 3.5.5. \square

Bemerkung 3.5.6

Die Suffizienz $\tilde{\theta}$ kommt im Beweis des Lemmas 3.5.5 explizit nicht vor. Dennoch ist sie notwendig, damit der Schätzer $\theta^* = \mathbb{E}(\hat{\theta}|\tilde{\theta}) = g(\tilde{\theta})$ nicht von θ abhängt.

Folgerung 3.5.2

Falls $\hat{\theta}$ ein vollständiger und suffizienter Schätzer für θ ist und falls eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ so existiert, dass $\mathbb{E}_\theta g(\hat{\theta}) = \theta \quad \forall \theta \in \Theta$, dann ist $g(\hat{\theta})$ der beste erwartungstreue Schätzer für θ .

Beweis $g(\hat{\theta}) = \mathbb{E}(g(\hat{\theta})|\hat{\theta})$, welcher nach Satz 3.5.5 der beste erwartungstreue Schätzer ist. \square

4 Konfidenzintervalle

4.1 Einführung

Konfidenz- oder Vertrauensintervalle wurden bereits in Kapitel 3 exemplarisch behandelt (vgl. Folgerung 3.3.2 und Bemerkung 3.3.4 des). In diesem Kapitel werden wir eine formale Definition eines Konfidenzintervalles angeben, um Vertrauensintervalle in größerer Tiefe studieren zu können. Dabei werden sowohl *Ein-* als auch *Zweistichprobenprobleme* behandelt.

Rufen wir uns die Annahmen eines parametrischen Modells in Erinnerung: es sei eine Stichprobe (X_1, \dots, X_n) von unabhängigen, identisch verteilten Zufallsvariablen mit $X_i \sim F_\theta$ gegeben, wobei F_θ eine Verteilungsfunktion aus einer parametrischen Familie von Verteilungen $\{F_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$ ist, dem m -dimensionalen Parameterraum, $m \geq 1$.

Die Punktschätzer von θ liefern jeweils einen Wert für den Parametervektor. Es wäre allerdings auch vorteilhaft, die Genauigkeit solcher Schätzansätze zu nennen, das heißt, einen Bereich anzugeben, in dem θ mit hoher Wahrscheinlichkeit $1 - \alpha$ liegt. Dabei heißt α *Irrtumswahrscheinlichkeit*; übliche Werte für α sind $\alpha = 0,01; 0,05; 0,1$. Die Wahrscheinlichkeit $1 - \alpha$, daß θ im vorgegebenen *Konfidenzintervall* liegt, heißt dann *Überdeckungswahrscheinlichkeit* oder *Konfidenzniveau* und soll dann entsprechend hoch ausfallen, z.B. $0,99; 0,95; 0,9$.

Definition 4.1.1

Es sei $1 - \alpha$ ein Konfidenzniveau und $\underline{\theta} : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, $\bar{\theta} : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ zwei Stichprobenfunktionen mit der Eigenschaft

$$\underline{\theta}(x_1, \dots, x_n) \leq \bar{\theta}(x_1, \dots, x_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Falls

1. $P_\theta \left(\theta \in \left[\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n) \right] \right) \geq 1 - \alpha, \quad \theta \in \Theta$
2. $\inf_{\theta \in \Theta} P_\theta \left(\theta \in \left[\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n) \right] \right) = 1 - \alpha$
3. $\lim_{n \rightarrow \infty} P_\theta \left(\theta \in \left[\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n) \right] \right) = 1 - \alpha, \quad \theta \in \Theta$

dann heißt $I = \left[\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n) \right]$ ein

1. *Konfidenzintervall*
2. *minimales Konfidenzintervall*
3. *asymptotisches Konfidenzintervall*

zum Konfidenzniveau $1 - \alpha$. Dabei heißt $l_\theta(X_1, \dots, X_n) = \bar{\theta}(X_1, \dots, X_n) - \underline{\theta}(X_1, \dots, X_n)$ die *Länge* des Konfidenzintervalls. Es ist erwünscht, möglichst kleine Konfidenzintervalle (mit minimaler Länge) bei großem Konfidenzniveau für θ zu konstruieren.

Wie bereits bei den Beispielen im Statistik I-Skript ersichtlich ist, folgt die Konstruktion eines Konfidenzintervalls einem bestimmten Muster, das wir jetzt genauer studieren werden:

1. Finde eine Statistik $T(X_1, \dots, X_n, \theta)$, die
 - vom Parameter θ abhängt und
 - eine bekannte (Prüf-) Verteilung F besitzt (möglicherweise asymptotisch für $n \rightarrow \infty$).
2. Bestimme von der Verteilung F die Quantile $F^{-1}(\alpha_1)$ und $F^{-1}(1 - \alpha_2)$ für Niveaus α_1 und $1 - \alpha_2$, sodaß $\alpha_1 + \alpha_2 = \alpha$.
3. Löse (falls möglich) die Ungleichung $F^{-1}(\alpha_1) \leq T(X_1, \dots, X_n, \theta) \leq F^{-1}(1 - \alpha_2)$ bzgl. θ auf. Das entsprechende Ergebnis $I = [T^{-1}(F^{-1}(\alpha_1)), T^{-1}(F^{-1}(1 - \alpha_2))]$ (im Falle einer monoton in θ steigenden Statistik T) ist ein Konfidenzintervall für θ zum Niveau $1 - \alpha$, denn es gilt

$$\begin{aligned}
 \mathbb{P}_\theta(\theta \in I) &= \mathbb{P}_\theta\left(T_\theta^{-1}(F^{-1}(\alpha_1)) \leq \theta \leq T^{-1}(F^{-1}(1 - \alpha_2))\right) \\
 &= \mathbb{P}_\theta\left(F^{-1}(\alpha_1) \leq T_\theta(X_1, \dots, X_n, \theta) \leq F^{-1}(1 - \alpha_2)\right) \\
 &= F(F^{-1}(1 - \alpha_2)) - F(F^{-1}(\alpha_1)) \\
 &= 1 - \alpha_2 - \alpha_1 \\
 &= 1 - \alpha \text{ für alle } \theta \in \Theta.
 \end{aligned}$$

Für asymptotische Konfidenzintervalle soll überall noch $\lim_{n \rightarrow \infty}$ geschrieben werden: $\lim_{n \rightarrow \infty} \mathbb{P}_\theta(\theta \in I) = \dots = 1 - \alpha$. Hierbei ist T_θ^{-1} die Inverse von $T(X_1, \dots, X_n, \theta)$ bezüglich θ . Grafisch kann dies auf Abb. 4.1 veranschaulicht werden.

Definition 4.1.2

1. Falls $\alpha_1 = \alpha_2 = \alpha/2$, dann heißt das Konfidenzintervall $I = [T^{-1}(F^{-1}(\frac{\alpha}{2})), T^{-1}(F^{-1}(1 - \frac{\alpha}{2}))]$ *symmetrisch*.
2. Falls $\alpha_1 = 0$ (bzw. $\underline{\theta}(X_1, \dots, X_n) = -\infty$), dann heißt das Konfidenzintervall $(-\infty, \bar{\theta}(X_1, \dots, X_n)]$ *einseitig*. Das selbe gilt für $\alpha_2 = 0$ (bzw. $\bar{\theta}(X_1, \dots, X_n) = +\infty$) und das Vertrauensintervall $[\underline{\theta}(X_1, \dots, X_n), +\infty)$.

In der Zukunft werden wir oft, ohne Beschränkung der Allgemeinheit, symmetrische Konfidenzintervalle konstruieren, obwohl man auch ein allgemeineres, nicht-symmetrisches Intervall leicht angeben kann.

Bemerkung 4.1.1

Man sieht leicht, daß der Algorithmus zur Konstruktion eines Vertrauensbereiches sich sehr dem eines statistischen Tests ähnelt. Im letzten Fall heißt $T(X_1, \dots, X_n)$ *Teststatistik*. Im Allgemeinen kann man für jedes Konfidenzintervall einen entsprechenden statistischen Test angeben, aber nicht umgekehrt. In der Vorlesung Stochastik III werden wir einige Beispiele dieser Übertragung „Konfidenzintervall \mapsto Test“ sehen.

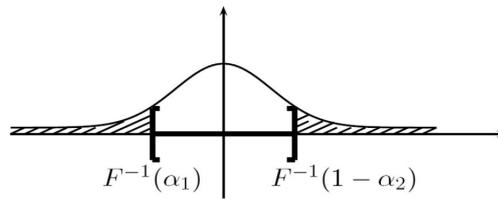


Abb. 4.1: asymptotisches Konfidenzintervall

4.2 Ein-Stichproben-Probleme

In diesem Abschnitt werden wir einige Beispiele von Vertrauensbereichen für Parameter einiger bekannter Verteilungen nach dem oben genannten Schema konstruieren. Dabei werden wir immer mit einer Stichprobe (X_1, \dots, X_n) wie in Abschnitt 4.1 arbeiten.

4.2.1 Normalverteilung

Es seien X_1, \dots, X_n unabhängig, identisch verteilt, mit $X_i \sim N(\mu, \sigma^2)$.

Konfidenzintervalle für den Erwartungswert μ

- **bei bekannter Varianz σ^2** Wenn wir annehmen, daß σ^2 bekannt ist, so ermöglicht uns der Satz 3.3.1, 4., ein exaktes Konfidenzintervall für μ zum Niveau $1 - \alpha$ zu berechnen. Denn es gilt $\bar{X}_n \sim N(\mu, \sigma^2/n)$ und somit

$$T(X_1, \dots, X_n, \mu) = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1)$$

Es seien z_{α_1} und $z_{1-\alpha_2}$ Quantile der $N(0, 1)$ -Verteilung, $\alpha_1 + \alpha_2 = \alpha$ und $1 - \alpha$ das vorgegebene Konfidenzniveau.

Dann gilt

$$\begin{aligned} 1 - \alpha &= \mathbb{P}(z_{\alpha_1} \leq T(X_1, \dots, X_n, \mu) \leq z_{1-\alpha_2}) \\ &= \mathbb{P}\left(z_{\alpha_1} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\alpha_2}\right) \\ &\stackrel{(-z_{\alpha_1} = z_{1-\alpha_1})}{=} \mathbb{P}\left(\bar{X}_n - \frac{z_{1-\alpha_2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{1-\alpha_1}\sigma}{\sqrt{n}}\right). \end{aligned}$$

Somit ist $[\underline{\theta}(X_1, \dots, X_n), \bar{\theta}(X_1, \dots, X_n)]$ mit $\underline{\theta}(X_1, \dots, X_n) = \bar{X}_n - z_{1-\alpha_2} \frac{\sigma}{\sqrt{n}}$ und $\bar{\theta}(X_1, \dots, X_n) = \bar{X}_n + z_{1-\alpha_1} \frac{\sigma}{\sqrt{n}}$ ein exaktes Konfidenzintervall für μ zum Niveau $1 - \alpha$.

Es hat die Länge $l_\mu(X_1, \dots, X_n) = \frac{\sigma}{\sqrt{n}} (z_{1-\alpha_2} + z_{1-\alpha_1})$. Es gilt $l_\mu(X_1, \dots, X_n) \rightarrow 0$, für $n \rightarrow \infty$ was bedeutet, daß bei wachsendem Informationsumfang ($n \rightarrow \infty$) die Präzision der Schätzung immer besser wird.

Im Symmetriefall ($\alpha_1 = \alpha_2 = \alpha/2$) müssen wir schreiben $\underline{\theta}(X_1, \dots, X_n) = \bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$, $\bar{\theta}(X_1, \dots, X_n) = \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ und $l_\mu(X_1, \dots, X_n) = \frac{2\sigma}{\sqrt{n}} z_{1-\alpha/2}$.

Daraus folgt, daß man bei vorgegebener Länge $\varepsilon > 0$ die Anzahl der Beobachtungen n bestimmen kann, die dann notwendig sind, um die vorgegebene Präzision zu erreichen:

$$\frac{2\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \varepsilon \iff n \geq \left(\frac{2\sigma z_{1-\alpha/2}}{\varepsilon} \right)^2 \quad (4.2.1)$$

Für $\alpha_1 = 0$ bzw. $\alpha_2 = 0$ kann man einseitige Intervalle $\left(-\infty, \bar{X}_n + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right]$ und $\left[\bar{X}_n - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}, +\infty\right)$ genauso angeben.

- **bei unbekannter Varianz σ^2 :** siehe Bemerkung 3.3.4.

Dort wurde das Konfidenzintervall $\left[\bar{X}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} S_n, \bar{X}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} S_n\right]$ für μ zum Konfidenzniveau $1 - \alpha$ konstruiert, wobei $t_{n-1,1-\alpha/2}$ das $(1 - \frac{\alpha}{2})$ -Quantil der t_{n-1} -Verteilung ist.

Wie man sieht, ist die Länge des Konfidenzintervalls zufällig: $l_\mu(X_1, \dots, X_n) = \frac{2S_n}{\sqrt{n}} t_{n-1,1-\alpha/2}$, somit macht es Sinn, mit erwarteter Länge

$$\mathbb{E} l_\mu(X_1, \dots, X_n) = \frac{2}{\sqrt{n}} \mathbb{E} S_n t_{n-1,1-\alpha/2}$$

zu arbeiten, um zum Beispiel die Frage nach der notwendigen Anzahl n von Beobachtungen bei vorgegebener Genauigkeit $\varepsilon > 0$ (vergleiche Gleichung (4.2.1)) zu beantworten.

Konfidenzintervalle für die Varianz σ^2

- **bei bekanntem Erwartungswert μ :**

Betrachten wir den Schätzer $\tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ für σ^2 . Aus Satz 3.3.5, 2. folgt $\frac{n\tilde{S}_n^2}{\sigma^2} \sim \chi_n^2$.

Wir setzen $T(X_1, \dots, X_n, \sigma^2) = \frac{n\tilde{S}_n^2}{\sigma^2}$ und bekommen

$$\mathbb{P} \left(\chi_{n,\alpha_2}^2 \leq \frac{n\tilde{S}_n^2}{\sigma^2} \leq \chi_{n,1-\alpha_1}^2 \right) = \mathbb{P} \left(\frac{n\tilde{S}_n^2}{\chi_{n,1-\alpha_1}^2} \leq \sigma^2 \leq \frac{n\tilde{S}_n^2}{\chi_{n,\alpha_2}^2} \right) = 1 - \alpha.$$

Somit ist $\left[\frac{n\tilde{S}_n^2}{\chi_{n,1-\alpha_1}^2}, \frac{n\tilde{S}_n^2}{\chi_{n,\alpha_2}^2} \right]$ ein Konfidenzintervall für σ^2 zum Niveau $1 - \alpha$, $\alpha = \alpha_1 + \alpha_2$ mit der mittleren Länge $\mathbb{E} l_{\sigma^2} = n\sigma^2 \left(\frac{1}{\chi_{n,\alpha_2}^2} - \frac{1}{\chi_{n,1-\alpha_1}^2} \right)$. Da die χ^2 -Verteilung nicht symmetrisch ist, ist auch das Konfidenzintervall nicht symmetrisch.

- **bei unbekanntem Erwartungswert μ :**

Ähnlich wie oben beschrieben folgt das Konfidenzintervall $\left[\frac{(n-1)S_n^2}{\chi_{n-1,1-\alpha_1}^2}, \frac{(n-1)S_n^2}{\chi_{n-1,\alpha_2}^2} \right]$ zum Niveau $1 - \alpha$, $\alpha = \alpha_1 + \alpha_2$ aus Satz 3.3.5, 1., weil $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$ für die Stichprobenvarianz $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Die erwartete Länge ist $\mathbb{E} l_{\sigma^2} = (n-1)\sigma^2 \left(\frac{1}{\chi_{n-1,\alpha_2}^2} - \frac{1}{\chi_{n-1,1-\alpha_1}^2} \right)$.

4.2.2 Konfidenzintervalle aus stochastischen Ungleichungen

Eine alternative Methode zur Gewinnung von Konfidenzintervallen besteht in der Anwendung stochastischer Ungleichungen. So kann man zum Beispiel bei einer Stichprobe (X_1, \dots, X_n) von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E} X_i = \mu$, $\text{Var} X_i = \sigma^2 \in (0, \infty)$ die Ungleichung von Tschebyschew benutzen, um ein einfaches, aber grobes Konfidenzintervall für μ zu konstruieren:

$$\begin{aligned} \mathbb{P}\left(|\bar{X}_n - \mu| > \varepsilon\right) &\leq \frac{\text{Var} \bar{X}_n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} = \alpha \\ &\Rightarrow \text{für } \varepsilon = \frac{\sigma}{\sqrt{n\alpha}} \text{ gilt: } 1 - \alpha = \mathbb{P}\left(|\bar{X}_n - \mu| \leq \varepsilon\right) \\ &= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n\alpha}} \leq -\bar{X}_n + \mu \leq \frac{\sigma}{\sqrt{n\alpha}}\right) \\ &= \mathbb{P}\left(\bar{X}_n - \frac{\sigma}{\sqrt{n\alpha}} \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\right). \end{aligned}$$

Das Konfidenzintervall $\left[\bar{X}_n - \frac{\sigma}{\sqrt{n\alpha}}, \bar{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\right]$ für μ bei bekannter Varianz σ^2 ist verteilungsunabhängig, da keinerlei Annahmen über die Verteilung von X_i gemacht wurden.

Präzisere Konfidenzintervalle können bei der Verwendung folgender *Ungleichung von Hoeffding* konstruiert werden:

Satz 4.2.1 (Ungleichung von Hoeffding):

Es seien Y_1, \dots, Y_n unabhängige Zufallsvariablen mit $\mathbb{E} Y_i = 0$, $a_i \leq Y_i \leq b_i$ fast sicher, $i = 1, \dots, n$. Für alle $\varepsilon > 0$ gilt

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

(ohne Beweis).

Diese Ungleichung ist schärfer als die Tschebyschew-Ungleichung. Falls man spezielle Annahmen über die Verteilung von Y_i macht, kann man mit ihrer Hilfe auf gute Konfidenzintervalle unter Verwendung des Satzes 4.2.1 kommen.

Nehmen wir z.B. an, daß X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariablen sind, $X_i \sim \text{Bernoulli}(p)$, $p \in (0, 1)$. Wir wollen ein Konfidenzintervall für p bestimmen.

Folgerung 4.2.1

Es seien X_1, \dots, X_n unabhängige Bernoulli(p)-verteilte Zufallsvariablen. Dann gilt $\mathbb{P}\left(|\bar{X}_n - p| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$, $\varepsilon > 0$.

Beweis Es gilt

$$\bar{X}_n - p = \frac{1}{n} \sum_{i=1}^n \underbrace{(X_i - p)}_{Y_i}, \quad Y_i \in [-p, 1 - p],$$

das heißt $a_i = -p$, $b_i = 1 - p$, $b_i - a_i = 1$, $i = 1, \dots, n$, $\mathbb{E} Y_i = p - p = 0$. Dann gilt:

$$\begin{aligned} \mathbb{P}_p \left(|\bar{X}_n - p| > \varepsilon \right) &\leq \mathbb{P}_p \left(\left| \sum_{i=1}^n Y_i \right| \geq \varepsilon n \right) \\ &= \mathbb{P}_p \left(\sum_{i=1}^n Y_i \geq \varepsilon n \right) + \mathbb{P}_p \left(\sum_{i=1}^n (-Y_i) \geq \varepsilon n \right) \\ &\stackrel{\text{(Satz 4.2.1)}}{\leq} 2e^{-\frac{2\varepsilon^2 n^2}{n}} = 2e^{-2\varepsilon^2 n}, \end{aligned}$$

wobei man den Satz 4.2.1 sowohl für die Folge $\{Y_i\}$ als auch $\{-Y_i\}$ anwendet. Damit ist die Behauptung bewiesen. \square

Bemerkung 4.2.1

Die Form der Ungleichung von Hoeffding ähnelt sehr der von Dvoretzky-Kiefer-Wolfowitz, Satz 3.3.10.

Nun fixieren wir $\alpha > 0$ und wählen $\varepsilon_n = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$. Durch Anwendung von Folgerung 4.2.1 mit diesem ε_n erhalten wir $\mathbb{P}_p \left(|\bar{X}_n - p| > \varepsilon_n \right) \leq \alpha$, somit $\mathbb{P}_p \left(|\bar{X}_n - p| \leq \varepsilon_n \right) \geq 1 - \alpha$ und darum ist $\left[\bar{X}_n - \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}, \bar{X}_n + \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}} \right]$ ein Konfidenzintervall für p zum Niveau $1 - \alpha$.

4.2.3 Asymptotische Konfidenzintervalle

Die Philosophie der Konstruktion von asymptotischen Konfidenzintervallen ist relativ einfach: Wir erläutern sie am Beispiel eines asymptotisch normalverteilten Schätzers $\hat{\theta}$ für einen Parameter θ .

Sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen, $X_i \sim F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ , der asymptotisch normalverteilt ist. Dann gilt für erwartungstreue $\hat{\theta}_n$

$$\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \xrightarrow{d} Y \sim N(0, 1),$$

wobei $\hat{\sigma}_n$ ein konsistenter Schätzer der asymptotischen Varianz von $\hat{\theta}_n$ ist.

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(z_{\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \leq z_{1-\alpha/2} \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\theta \in \left[\hat{\theta}_n - z_{1-\alpha/2} \hat{\sigma}_n, \hat{\theta}_n + z_{1-\alpha/2} \hat{\sigma}_n \right] \right) = 1 - \alpha. \end{aligned}$$

Somit ist $\left[\hat{\theta}_n - z_{1-\alpha/2} \hat{\sigma}_n, \hat{\theta}_n + z_{1-\alpha/2} \hat{\sigma}_n \right]$ ein asymptotisches Konfidenzintervall für θ zum Niveau $1 - \alpha$.

Diese Vorgehensweise werden wir jetzt anhand von zwei Beispielen klar machen:

- **Bernoulli-Verteilung:**

Seien $X_i \sim \text{Bernoulli}(p)$ -verteilt, $i = 1, \dots, n$. Dann gilt $\theta = p$, $\hat{\theta}_n = \hat{p}_n = \bar{X}_n$. $\mathbb{E}_p \hat{p}_n = p$, $\text{Var}_p \hat{p}_n = \frac{p(1-p)}{n}$. Wir wählen $\hat{\sigma}^2 = \frac{1}{n} \hat{p}(1 - \hat{p}_n) = \frac{\bar{X}_n}{n} (1 - \bar{X}_n)$ als Plug-In-Schätzer

für σ^2 . Dann gilt nach dem zentralen Grenzwertsatz (Satz 7.2.1, WR) und dem Satz von Slutsky (Satz 6.4.2, 3. WR):

$$\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1),$$

das heißt $p \in \left[\bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right]$ stellt ein asymptotisches Konfidenzintervall für p zum Niveau $1 - \alpha$ dar. Da aber $p \in [0, 1]$ sein soll, betrachtet man

$$\underline{p}(X_1, \dots, X_n) = \max \left\{ 0, \bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right\}$$

und

$$\bar{p}(X_1, \dots, X_n) = \min \left\{ 1, \bar{X}_n + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right\}.$$

Bemerkung 4.2.2

Ein anderes asymptotisches Konfidenzintervall für den Parameter p der Bernoulli-Verteilung bekommt man, wenn man die Aussage des zentralen Grenzwertsatzes

$\lim_{n \rightarrow \infty} \mathbb{P}_p \left(-z_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq z_{1-\alpha/2} \right) = 1 - \alpha$ nimmt und die quadratische Ungleichung dann bezüglich p auflöst.

Übungsaufgabe 4.2.1

Lösen Sie die Ungleichung auf!

• Poissonverteilung:

Es seien $X_i \sim \text{Poisson}(\lambda)$, $i = 1, \dots, n$, dann gilt $\theta = \lambda$, $\hat{\theta}_n = \hat{\lambda} = \bar{X}_n$. Da $\mathbb{E}_\lambda X_i = \text{Var}_\lambda X_i = \lambda$, kann man den zentralen Grenzwertsatz (Satz 7.2.1, WR) anwenden

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1),$$

Da \bar{X}_n stark konsistent für λ ist, gilt nach dem Satz von Slutsky (Satz 6.4.2, 4, WR)

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1).$$

Daraus folgt ein asymptotisches Konfidenzintervall

$$\left[\bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n}{n}}, \bar{X}_n + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n}{n}} \right]$$

für den Parameter λ zum Konfidenzniveau $1 - \alpha$.

Bemerkung 4.2.3 1. Ähnlich wie in Bemerkung 4.2.2 angegeben, kann man durch Auflösen der quadratischen Ungleichung in

$$\lim_{n \rightarrow \infty} \mathbb{P}_\lambda \left(\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \in [-z_{1-\alpha/2}, z_{1-\alpha/2}] \right) = 1 - \alpha$$

bezüglich λ ein alternatives asymptotisches Konfidenzintervall für λ angeben.

Übungsaufgabe 4.2.2

Bitte führen Sie diese Berechnungen durch.

2. Da $\lambda > 0$ ist, kann man die untere Schranke diesbezüglich korrigieren:

$$\underline{\lambda}(X_1, \dots, X_n) = \max \left\{ 0, \bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n}{n}} \right\}$$

4.3 Zwei-Stichproben-Probleme

In diesem Abschnitt werden Charakteristiken bzw. Parameter von zwei unterschiedlichen Stichproben miteinander verglichen, indem man Konfidenzintervalle für einfache Funktionen dieser Parameter konstruiert.

Betrachten wir zwei Zufallsstichproben $Y_1 = (X_{11}, \dots, X_{1n_1})$, $Y_2 = (X_{21}, \dots, X_{2n_2})$ von Zufallsvariablen X_{i1}, \dots, X_{in_i} , $i = 1, 2$, die innerhalb der Stichprobe Y_i jeweils unabhängig und identisch verteilt sind, $X_{ij} \stackrel{d}{=} X_i$, $j = 1, \dots, n_i$, $i = 1, 2$ und die Prototyp-Zufallsvariable $X_i \sim F_{\theta_i}$, $\theta_i \in \Theta \subset \mathbb{R}^m$. Es wird im Allgemeinen nicht gefordert, daß Y_1 und Y_2 unabhängig sind. Falls sie voneinander abhängen, spricht man von *verbundenen Stichproben* Y_1 und Y_2 . Betrachten wir eine Funktion $g : \mathbb{R}^{2m} \rightarrow \mathbb{R}$ von den Parametervektoren θ_1 und θ_2 . In diesem Skript werden dabei meistens die Fälle $m = 1, 2$, $g(\theta_1, \theta_2) = \theta_{1j} - \theta_{2j}$, $g(\theta_1, \theta_2) = \frac{\theta_{1j}}{\theta_{2j}}$ untersucht, wobei $\theta_i = (\theta_{i1}, \dots, \theta_{im})$, $i = 1, 2$.

Unsere Zielstellung wird sein, ein (möglicherweise asymptotisches) Konfidenzintervall für $g(\theta_1, \theta_2)$ mit Hilfe der Stichprobe (Y_1, Y_2) zu gewinnen.

Dabei wird die selbe Philosophie wie in Abschnitt 4.1 beschrieben verfolgt. Es wird eine Statistik $T(Y_1, Y_2, g(\theta_1, \theta_2))$ gesucht, die eine (möglicherweise asymptotische) Prüfverteilung F besitzt und von $g(\theta_1, \theta_2)$ explizit abhängt.

Durch das Auflösen der Ungleichung $F_{\alpha_1}^{-1} \leq T(Y_1, Y_2, g(\theta_1, \theta_2)) \leq F_{1-\alpha_2}^{-1}$ bzgl. $g(\theta_1, \theta_2)$ bekommt man dann ein (möglicherweise asymptotisches) Konfidenzintervall zum Niveau $1 - \alpha$, $\alpha = \alpha_1 + \alpha_2$.

4.3.1 Normalverteilte Stichproben

Hier wird angenommen, daß $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$.

Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ bei bekannten Varianzen σ_1^2 und σ_2^2 und unabhängigen Stichproben

Seien Y_1 und Y_2 voneinander unabhängig und σ_1^2, σ_2^2 bekannt. Wir betrachten die Parameterfunktion $g(\mu_1, \mu_2) = \mu_1 - \mu_2$. Es seien $\bar{X}_{in_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$, $i = 1, 2$ die Stichprobenmittel der Stichproben Y_1 und Y_2 . Es gilt $\bar{X}_{in_i} \sim N(\mu_i, \frac{\sigma_i^2}{n_i})$, $i = 1, 2$. Nach Satz 3.3.3, 4)

sind \bar{X}_{1n_1} und \bar{X}_{2n_2} unabhängig. Dann ist wegen der Faltungstabilität der Normalverteilung $X_{1n_1} - X_{2n_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$. Nach dem Normieren erhält man die Statistik $T(Y_1, Y_2, \mu_1 - \mu_2) = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$. Daraus bekommt man das Konfidenzintervall

$$\left[\bar{X}_{1n_1} - \bar{X}_{2n_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X}_{1n_1} - \bar{X}_{2n_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

für $\mu_1 - \mu_2$ zum Niveau $1 - \alpha$.

Konfidenzintervall für den Quotienten $\frac{\sigma_1^2}{\sigma_2^2}$ bei unbekanntem Erwartungswerten μ_1 und μ_2 und unabhängigen Stichproben

Seien Y_1 und Y_2 voneinander unabhängig. Sei $g(\sigma_1, \sigma_2) = \frac{\sigma_1^2}{\sigma_2^2}$. Wir konstruieren die Statistik $T(Y_1, Y_2, \frac{\sigma_1^2}{\sigma_2^2})$ folgendermaßen: Seien $S_{in_i}^2 = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{in_i})^2$, $i = 1, 2$ die Stichprobenvarianzen der Stichproben Y_1 und Y_2 . Dann gilt $\frac{(n_i-1)S_{in_i}^2}{\sigma_i^2} \sim \chi_{n_i-1}^2$, $i = 1, 2$ nach Satz 3.3.5.

Da die $S_{in_i}^2$ voneinander unabhängig sind, gilt

$$T\left(Y_1, Y_2, \frac{\sigma_1^2}{\sigma_2^2}\right) = \frac{\frac{(n_2-1)S_{2n_2}^2}{(n_2-1)\sigma_2^2}}{\frac{(n_1-1)S_{1n_1}^2}{(n_1-1)\sigma_1^2}} = \frac{S_{2n_2}^2}{S_{1n_1}^2} \cdot \frac{\sigma_1^2}{\sigma_2^2} \sim F_{n_2-1, n_1-1}$$

nach der Definition der F -Verteilung. Daraus ergibt sich das Konfidenzintervall

$$\left[\frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, \alpha_1}, \frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, 1-\alpha_2} \right]$$

für $\frac{\sigma_1^2}{\sigma_2^2}$ zum Niveau $1 - \alpha$.

Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ der Erwartungswerte bei verbundenen Stichproben

Dieses Mal seien Y_1 und Y_2 verbunden, $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma^2)$ für ein unbekanntes $\sigma^2 > 0$, $n_1 = n_2 = n$. Da $X_{ij}, j = 1, \dots, n$ unabhängig und identisch verteilt sind, gilt $Z_j = X_{1j} - X_{2j} \sim N(\mu_1 - \mu_2, \sigma^2)$, $j = 1, \dots, n$.

Unser Ziel ist es, ein Konfidenzintervall für $\mu_1 - \mu_2$ zu bekommen. Wenn wir die Stichprobe (Z_1, \dots, Z_n) betrachten, und Ergebnisse des Abschnittes 4.2.1, 2. anwenden, so erhalten wir sofort folgendes Konfidenzintervall:

$$\left[\bar{Z}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \bar{Z}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right]$$

für $\mu_1 - \mu_2$ zum Niveau $1 - \frac{\alpha}{2}$, wobei $\bar{Z}_n = \frac{1}{n} \sum_{j=1}^n Z_j = \frac{1}{n} \sum_{j=1}^n (X_{1j} - X_{2j}) = \bar{X}_{1n} - \bar{X}_{2n}$,

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (Z_j - \bar{Z}_n)^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{1j} - X_{2j} - \bar{X}_{1n} + \bar{X}_{2n})^2$$

4.3.2 Poissonverteilte Stichproben

Wir nehmen jetzt an, daß die Stichproben Y_1 und Y_2 unabhängig sind, und $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2$. Konstruieren wir asymptotische Konfidenzintervalle für $g(\lambda_1, \lambda_2) = \lambda_1 - \lambda_2$ und $g(\lambda_1, \lambda_2) = \frac{n_2 \lambda_2}{n_1 \lambda_1 + n_2 \lambda_2} = \frac{\lambda_2}{\rho \lambda_1 + \lambda_2}$, $\rho = \frac{n_1}{n_2} = \text{const}$, wobei $n_1, n_2 \rightarrow \infty$.

Asymptotisches Konfidenzintervall für $\lambda_1 - \lambda_2$

Um zu einer Statistik $T(Y_1, Y_2, \lambda_1 - \lambda_2)$ zu kommen, die asymptotisch (für $n_1, n_2 \rightarrow \infty$) $N(0, 1)$ -verteilt ist, verwenden wir den zentralen Grenzwertsatz von Ljapunow (vergleiche Satz 7.2.6, WR).

Lemma 4.3.1

Es gilt

$$\frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}} \xrightarrow[n_1 \rightarrow \infty, n_2 \rightarrow \infty]{d} Y \sim N(0, 1)$$

Beweis Führen wir die Zufallsvariable

$$Z_{nk} = \begin{cases} \frac{X_{1k} - \lambda_1}{n_1 \sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}}, & k = 1, \dots, n_1 \\ -\frac{X_{2k-n_1} - \lambda_2}{n_2 \sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}}, & k = n_1 + 1, \dots, n_1 + n_2 \end{cases}$$

ein, wobei $n = n_1 + n_2$. Es gilt: $\mathbb{E} Z_{nk} = 0$ für alle $k = 1, \dots, n$, und

$$0 < \sigma_{nk}^2 = \text{Var} Z_{nk} = \begin{cases} \frac{\text{Var} X_{1k}}{n_1^2 \left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)} = \frac{\lambda_1}{n_1 \left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)}, & k = 1, \dots, n_1, \\ \frac{\lambda_2}{n_2^2 \left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)}, & k = n_1 + 1, \dots, n, \end{cases}$$

somit

$$\sum_{k=1}^n \sigma_{nk}^2 = \left(\frac{\lambda_1}{n_1^2} n_1 + \frac{\lambda_2}{n_2^2} n_2 \right) \frac{1}{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}} = 1.$$

Außerdem gilt für $\delta > 0$ und $n_1, n_2 \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E} (|Z_{nk}|)^{2+\delta} = \lim_{n_1, n_2 \rightarrow \infty} \left(\frac{\mathbb{E} (|X_{11} - \lambda_1|^{2+\delta})}{n_1^{1+\delta} \left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} + \frac{\mathbb{E} (|X_{21} - \lambda_2|^{2+\delta})}{n_2^{1+\delta} \left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} \right) = 0$$

Somit ist die Ljapunow-Bedingung erfüllt und nach Satz 7.2.6 (WR) gilt

$$\sum_{k=1}^n Z_{nk} \xrightarrow[n_1 \rightarrow \infty, n_2 \rightarrow \infty]{d} Y \sim N(0, 1).$$

Es gilt aber auch $\sum_{n=1}^n Z_{nk} = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}}$, somit ist das Lemma bewiesen. \square

Da $\bar{X}_{in_i} \xrightarrow{f.s.} \lambda_i$, $i = 1, 2$ nach dem starken Gesetz der großen Zahlen, gilt mit Hilfe des Satzes von Slutsky

$$T(Y_1, Y_2, \lambda_1 - \lambda_2) = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\bar{X}_{1n_1}/n_1 + \bar{X}_{2n_2}/n_2}} \xrightarrow[n_1, n_2 \rightarrow \infty]{d} Y \sim N(0, 1)$$

Daraus läßt sich sofort das asymptotische Konfidenzintervall für $\lambda_1 - \lambda_2$ zum Niveau $1 - \alpha$ ableiten:

$$\left[\bar{X}_{1n_1} - \bar{X}_{2n_2} - z_{1-\alpha/2} \sqrt{\frac{\bar{X}_{1n_1}}{n_1} + \frac{\bar{X}_{2n_2}}{n_2}}, \bar{X}_{1n_1} - \bar{X}_{2n_2} + z_{1-\alpha/2} \sqrt{\frac{\bar{X}_{1n_1}}{n_1} + \frac{\bar{X}_{2n_2}}{n_2}} \right]$$

Asymptotisches Konfidenzintervall für $\frac{n_2 \lambda_2}{n_1 \lambda_1 + n_2 \lambda_2}$

Es sei $n_1/n_2 = \rho = \text{const}$ und $g(\lambda_1, \lambda_2) = \frac{n_2 \lambda_2}{n_1 \lambda_1 + n_2 \lambda_2} = \frac{\lambda_2}{\beta \lambda_1 + \lambda_2} \stackrel{\text{Def.}}{=} p$. Es wird ein asymptotisches Konfidenzintervall für p gesucht. Wir führen die Statistik

$$T(Y_1, Y_2, p) = \frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{\hat{p}(1 - \hat{p})(S_{1n_1} + S_{2n_2})}}$$

ein, wobei $S_{in_i} = \sum_{j=1}^{n_i} X_{ij}$, $i = 1, 2$ und

$$\hat{p} = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} = \frac{n_2 \bar{X}_{2n_2}}{n_1 \bar{X}_{1n_1} + n_2 \bar{X}_{2n_2}} \xrightarrow[n_1, n_2 \rightarrow \infty]{f.s.} p$$

ein konsistenter Schätzer für p (wegen des starken Gesetzes der großen Zahlen) ist. Falls wir zeigen können, daß $T(Y_1, Y_2, p) \xrightarrow[n_1, n_2 \rightarrow \infty]{d} Y \sim N(0, 1)$, so wird daraus folgendes Konfidenzintervall ableitbar: Aus

$$\lim_{\substack{n_1 \rightarrow \infty \\ n_2 \rightarrow \infty}} \mathbb{P} \left(-z_{1-\alpha/2} \leq \frac{\frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} - p}{\sqrt{\frac{S_{1n_1} \cdot S_{2n_2}}{(S_{1n_1} + S_{2n_2})^3}}} \cdot (S_{1n_1} + S_{2n_2})^{3/2} \leq z_{1-\alpha/2} \right) = 1 - \alpha$$

folgt, daß

$$\left[\underline{\theta}(Y_1, Y_2), \bar{\theta}(Y_1, Y_2) \right]$$

mit

$$\underline{\theta}(\lambda_1, \lambda_2) = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} - z_{1-\alpha/2} \cdot \sqrt{\frac{S_{1n_1} \cdot S_{2n_2}}{(S_{1n_1} + S_{2n_2})^3}}$$

$$\bar{\theta}(\lambda_1, \lambda_2) = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} + z_{1-\alpha/2} \cdot \sqrt{\frac{S_{1n_1} \cdot S_{2n_2}}{(S_{1n_1} + S_{2n_2})^3}}$$

ein asymptotisches Konfidenzintervall für p zum Niveau $1 - \alpha$ ist.

Da $0 < p < 1$ sein soll, können die Schranken des Intervalls diesbezüglich korrigiert werden:

$$\begin{aligned}\underline{\theta}^*(Y_1, Y_2) &= \max\{0, \underline{\theta}(Y_1, Y_2)\}, \\ \bar{\theta}^*(Y_1, Y_2) &= \min\{1, \bar{\theta}(Y_1, Y_2)\}.\end{aligned}$$

Nun soll die asymptotische Normalverteiltheit von $T(Y_1, Y_2, p)$ gezeigt werden. Sie folgt aus dem Satz von Slutsky und folgendem Lemma:

Lemma 4.3.2

Es gilt:

$$\frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{p(1-p)(S_{1n_1} + S_{2n_2})}} \xrightarrow[n_1 \rightarrow \infty]{d} Y \sim N(0, 1)$$

Beweis Um die Aussage des Lemmas zu zeigen, verwenden wir einen zentralen Grenzwertsatz für Summen von Zufallsvariablen in zufälliger Anzahl (vgl. Satz 7.2.2 (WR)). Führen wir die Folge $N_n = S_{1n_1} + S_{2n_2}$ von nichtnegativen Zufallsvariablen ein. Die Summe ist monoton wachsend. Gleichzeitig setzen wir $a_{n_2} = n_1\lambda_1 + n_2\lambda_2$. Offensichtlich gilt

$$\begin{aligned}\frac{N_n}{a_{n_2}} &= \frac{S_{1n_1}}{n_1\lambda_1 + n_2\lambda_2} + \frac{S_{2n_2}}{n_1\lambda_1 + n_2\lambda_2} \\ &= \frac{\bar{X}_{1n_1}}{\lambda_1 + \rho^{-1}\lambda_2} + \frac{\bar{X}_{2n_2}}{\rho\lambda_1 + \lambda_2} \\ &\xrightarrow[n_1, n_2 \rightarrow \infty]{f.s.} \frac{\lambda_1}{\lambda_1 + \rho^{-1}\lambda_2} + \frac{\lambda_2}{\rho\lambda_1 + \lambda_2} \\ &= \frac{\rho\lambda_1}{\rho\lambda_1 + \lambda_2} + \frac{\lambda_2}{\rho\lambda_1 + \lambda_2} = 1\end{aligned}$$

Außerdem gilt:

$$\begin{aligned}\mathbb{P}(S_{2n_2} = k \mid N_n = m) &= \frac{\mathbb{P}(S_{2n_2} = k, S_{1n_1} + S_{2n_2} = m)}{\mathbb{P}(S_{1n_1} + S_{2n_2} = m)} \\ &= \frac{\mathbb{P}(S_{2n_2} = k, S_{1n_1} = m - k)}{\mathbb{P}(S_{1n_1} + S_{2n_2} = m)} \\ &= \frac{e^{-n_2\lambda_2} \frac{(\lambda_2 n_2)^k}{k!} \cdot e^{-n_1\lambda_1} \frac{(n_1\lambda_1)^{m-k}}{(m-k)!}}{e^{-n_1\lambda_1 - n_2\lambda_2} \frac{(n_1\lambda_1 + n_2\lambda_2)^m}{m!}} \\ &= \frac{m!}{(m-k)!k!} \left(\frac{n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2} \right)^m \left(\frac{n_1\lambda_1}{n_1\lambda_1 + n_2\lambda_2} \right)^{m-k} \\ &= \binom{m}{k} p^k (1-p)^{m-k}\end{aligned}$$

was bedeutet, daß $S_{2n_2} \mid N_n = m \sim \text{Bin}(m, p)$. Dann gilt $\frac{S_{2n_2} - mp}{\sqrt{mp(1-p)}} \mid N_n = m \stackrel{d}{=} \frac{S_m - mp}{\sqrt{mp(1-p)}}$,

wobei $S_m = \sum_{i=1}^m Z_i$ eine Summe von unabhängigen, identisch verteilten Zufallsvariablen $Z_i \sim \text{Bernoulli}(p)$ ist. Nach Satz 7.2.2 (WR) gilt dann

$$\frac{S_{N_n} - N_n p}{\sqrt{N_n p(1-p)}} \xrightarrow{d} Y \sim N(0, 1) \iff \frac{S_{2n_2} - N_n p}{\sqrt{N_n p(1-p)}} \xrightarrow{d} Y \sim N(0, 1).$$

□

Literaturverzeichnis

- [1] BICKEL, P. ; DOKSUM, K.: *Mathematical Statistics: Basic Ideas and Selected Topics*. London : Prentice Hall, 2001. – 2nd ed., Vol. 1
- [2] BOROVKOV, A. A.: *Mathematical Statistics*. Gordon & Breach, 1998
- [3] BURKSCHAT, M. ; CRAMER, E. ; KAMPS, U.: *Beschreibende Statistik, Grundlegende Methoden*. Berlin : Springer, 2004
- [4] CASELLA, G. ; BERGER, R. L.: *Statistical Inference*. Duxbury : Pacific Grove (CA), 2002
- [5] CRAMER, E. ; CRAMER, K. ; KAMPS, U. ; ZUCKSCHWERDT: *Beschreibende Statistik, Interaktive Grafiken*. Berlin : Springer, 2004
- [6] CRAMER, E. ; KAMPS, U.: *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik*. Berlin : Springer, 2007
- [7] DALGAARD, P.: *Introductory Statistics with R*. Berlin : Springer, 2002
- [8] FAHRMEIR, L. ; KNEIB, T. ; LANG, S.: *Regression. Modelle, Methoden und Anwendungen*. Berlin : Springer, 2007
- [9] FAHRMEIR, L. ; KÜNSTLER, R. ; PIGEOT, I. ; TUTZ, G.: *Statistik. Der Weg zur Datenanalyse*. Berlin : Springer, 2001
- [10] GEORGH, H. O.: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Berlin : de Gruyter, 2002
- [11] HARTUNG, J. ; ELPERT, B. ; KLÖSENER, K. H.: *Statistik*. München : R. Oldenbourg Verlag, 1993. – 9. Auflage
- [12] HEYDE, C. C. ; SENETA, E.: *Statisticians of the Centuries*. Berlin : Springer, 2001
- [13] IRLE, A.: *Wahrscheinlichkeitstheorie und Statistik, Grundlagen – Resultate – Anwendungen*. Teubner, 2001
- [14] KAZMIR, L. J.: *Wirtschaftsstatistik*. McGraw-Hill, 1996
- [15] KOCH, K. R.: *Parameter Estimation and Hypothesis Testing in Linear Models*. Berlin : Springer, 1999
- [16] KRAUSE, A. ; OLSON, M.: *The Basics of S-PLUS*. Berlin : Springer, 2002. – Third Ed.
- [17] KRENGEL, U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig : Vieweg, 2002. – 6. Auflage
- [18] LEHMANN, E. L.: *Elements of Large-Sample Theory*. New York : Springer, 1999

- [19] LEHN, J. ; WEGMANN, H.: *Einführung in die Statistik*. Stuttgart : Teubner, 2000. – 3. Auflage
- [20] MAINDONALD, J. ; BRAUN, J.: *Data Analysis and Graphics Using R*. Cambridge University Press, 2003
- [21] OVERBECK-LARISCH, M. ; DOLEJSKY, W.: *Stochastik mit Mathematica*. Braunschweig : Vieweg, 1998
- [22] PRUSCHA, H.: *Angewandte Methoden der Mathematischen Statistik*. Stuttgart : Teubner, 1996
- [23] PRUSCHA, H.: *Vorlesungen über Mathematische Statistik*. Stuttgart : Teubner, 2000
- [24] SACHS, L.: *Angewandte Statistik*. Springer, 1992
- [25] SACHS, L. ; HEDDERICH, J.: *Angewandte Statistik, Methodensammlung mit R*. Berlin : Springer, 2006
- [26] SHIRYAEV, A. N.: *Probability*. New York : Springer, 1996
- [27] SPIEGEL, M. R. ; STEPHENS, L. J.: *Statistik*. McGraw-Hill, 1999
- [28] STAHEL, W. A.: *Statistische Datenanalyse*. Vieweg, 1999
- [29] VENABLES, W. ; RIPLEY, D.: *Modern applied statistics with S-PLUS*. Springer, 1999. – 3rd ed
- [30] WASSERMAN, L.: *All of Statistics. A Concise Course in Statistical Inference*. Springer, 2004

Index

- a-posteriori-Verteilung, 82
- a-priori-Verteilung, 82
- absolute Häufigkeit, 7
- Abweichung, mittlere quadratische, 16
- arithmetisches Mittel, 12
- asymptotisch erwartungstreu, 50
- asymptotisch normalverteilt, 51
- Ausgangsvariable, 33

- Balkendiagramm, 8
- Bandbreite, 25
- Bayes-Schätzer, 83
- Bayesche Formel, 82
- bedingte Wahrscheinlichkeit, 95
- bedingter Erwartungswert, 94
- Bernoulli-Verteilung
 - asymptotisches Konfidenzintervall, 111
- besserer Schätzer, 51
- besten erwartungstreuer Schätzer, 51
- Bestimmtheitsmaß, 37
- Bias, 50
- bimodal, 10
- Blackwell-Rao, Ungleichung von, 104
- Bootstrap
 - Konfidenzintervall, 88
 - Schätzer, 87
- Bootstrap-Schätzer
 - Monte-Carlo-Methoden, 87
- Box-Plot, 14
 - modifizierter, 14
- Bravais-Pearson-Koeffizient, 31
- Bravais-Pearson-Korrelationskoeffizient, 29
- Brownsche Brücke, 69

- χ^2 -Verteilung, 43
- Cramér-Rao, Ungleichung von, 89

- Daten-Stichproben, 1
- Datenbereinigung, 1
- Datenerhebung, 1

- Dichteschätzung, 25
- Dichtetransformationssatz für Zufallsvektoren, 47
- gleichmäßiger Abstand D_n , 64
- Dvoretzky-Kiefer-Wolfowitz, Ungleichung von, 66

- Eindeutigkeit der besten erwartungstreuen Schätzer, 102
- Einfache lineare Regression, 33
- Einflussfaktor, 33
- empirische(r)
 - Kovarianz, 29
 - Median, 14
 - Standardabweichung, 15
 - Varianz, 15
 - Variationskoeffizient, 15, 17
 - Verteilungsfunktion, 10
- Erlangverteilung, 43
- erwartungstreu, 49
- Erwartungstreue, 16
- Erwartungswert, bedingter, 94
- Explorative Datenanalyse, 1

- Faktorisierungssatz von Neyman-Fisher, 98
- Fehler 1. Art, 69
- Fisher
 - Fisher-Information, 78
 - Fisher-Snedecor-Verteilung, F-Verteilung, 47
 - Wölbungsmaß von Fisher, 21

- Gammaverteilung
 - Faltungstabilität, 43
 - Momenterzeugende und charakteristische Funktion, 42
- geometrisches Mittel, 12, 13
- Gesamtstreuung, 37
- getrimmtes Mittel, 12
- Gini-Koeffizient, 17, 18

Gini-Koeffizient, Darstellung von, 18
 Gliwenco-Cantelli, Satz von, 64
 Grundgesamtheit, 1
 Häufigkeit
 absolute, 7
 relative, 7
 harmonisches Mittel, 13
 Herfindahl-Index, 17
 Histogramm, 7
 eindimensionales Histogramm, 7
 zweidimensionales Histogramm, 27
 Hoeffding-Ungleichung, 109
 identifizierbar, 40
 Information von Kullback-Leibler, 75
 Invarianzeigenschaften, 32
 Irrtumswahrscheinlichkeit, 105
 Jackknife-Schätzer für die/den
 Erwartungswert, 85
 Varianz, 86
 Verzerrung (Bias), 86
 Kerndichteschätzer
 eindimensionaler Kerndichteschätzer, 25
 zweidimensionaler Kerndichteschätzer, 29
 Kolmogorow, Satz von, 68
 Kolmogorow-Abstand D_n , 64
 Kolmogorow-Verteilung, 68
 Konfidenzband, 66
 Konfidenzintervall, 57, 105
 asymptotisches, 105, 110
 für die Bernoulli-Verteilung, 111
 für die Poissonverteilung, 111
 Bootstrap, 88
 Lange, 105
 minimales, 105
 Konfidenzniveau, 105
 konsistenter Schätzer, 50
 Konstanzbereich, 67
 Konzentrationsrate, 17, 20
 Korrelationskoeffizient, 29
 Spearman's, 31
 Kovarianz, empirische, 29
 Kreisdiagramm, 7, 8
 Kullback-Leibler, Information von, 75
 Kurtosis, 21
 Lagemaß, 12
 Lehmann-Scheffé, Satz von, 103
 Lernstichprobe, 3
 Likelihood-Funktion, 73
 linksschief, 10
 linkssteil, 10
 Lorenz-Münzner-Koeffizient, 20
 Lorenzkurve, 17
 maximale Streuung, 16
 Maximum-Likelihood-Schätzer, 73
 schwache Konsistenz, 76
 Median, 12, 15
 empirischer, 14
 Mittel
 arithmetisches, 12
 geometrisches, 12, 13
 getrimmtes, 12
 harmonisches, 12, 13
 Mittelwert, 12, 15
 mittlere quadratische Abweichung, 16
 mittlerer quadratischer Fehler, 50
 Modalität, 10
 Modellierung von Daten, 1
 Modellvalidierung, 1
 modifizierter Box-Plot, 14
 Modus, 12, 15
 Momentenmethode, 71
 Momentenschätzer, 71
 multimodal, 10
 Neyman-Fisher, Faktorisierungssatz, 98
 Normalverteilung
 Konfidenzintervall
 für eine Stichprobe, 107
 für zwei Stichproben, 112
 Ordnungsstatistik, 6, 12, 13
 Parameterraum, 40
 Parametervektor, 40
 Plug-in-Methode, 69
 Plug-in-Schätzer, 69, 70
 Poissonverteilung, 114
 asymptotisches Konfidenzintervall, 111
 Polynomiale Regression, 33
 Punktschätzer, 40

Quantil, 12, 13
 Quantilplot, 21
 Quartil, 12, 14

 Rangkorrelationskoeffizient, 31
 Realisierung, 5, 7
 rechtsschief, 10
 rechtssteil, 10
 Regressand, 33
 Regression, 33
 einfache lineare, 33
 polynomiale, 33
 Regressionsgerade, 34
 Regressionsgerade, Eigenschaften von, 37
 Regressionskoeffizient, 34
 Regressionskonstante, 34
 Regressionsvarianz, 34
 Regressor, 33
 relative Häufigkeit, 7
 Resampling-Methode, 84
 Residualplot, 39
 Residuen, 35

 Säulendiagramm, 8
 Satz
 χ^2 -Verteilung, Spezialfall, 43
 Darstellung des Gini-Koeffizient, 18
 Dichte der t -Verteilung, 45
 Dichtetransformationssatz für Zufallsvektoren, 47
 Eigenschaften der empirischen Momente, 51
 Eigenschaften des bedingten Erwartungswertes, 94
 Faktorisierungssatz von Neyman-Fisher, 98
 Gliwienko-Cantelli, 64
 Invarianzeigenschaften, 32
 Kolmogorow, 68
 Lehmann-Scheffé, 103
 Momenterzeugende und charakteristische Funktion der Gamma-Verteilung, 42
 Schwache Konsistenz von ML-Schätzern, 76
 Ungleichung von Cramér-Rao, 89
 Ungleichung von Dvoretzky-Kiefer-Wolfowitz, 66

 Schätzer, 49
 besserer, 51
 bester erwartungstreuer, 51
 konsistenter, 50
 suffizienter, 96
 Vergleich von, 51
 Schiefe, 12, 20
 Schließende Datenanalyse, 2
 Spannweite, 15
 Spearmans Korrelationskoeffizient, 31
 Stabdiagramm, 7
 Stamm-Blatt-Diagramm, 9
 Standardabweichung, 17
 Statistische Merkmale, 4
 stem-leaf display, 9
 Stichproben, 5
 Stichprobenfunktion, 6
 Stichprobenmittel, 6, 12
 Stichprobenvarianz, 6, 15
 Streudiagramm, 27, 39
 suffizienter Schätzer, 96
 sum of squared residuals, 37
 sum of squares explained, 37
 sum of squares total, 37
 Symmetriekoeffizient, 20
 symmetrisch, 10

 t -Verteilung, 45
 Teststatistik, 106
 Tortendiagramm, 8
 Transformationsregel, 16

 unimodal, 10, 76
 unverzerrt, 49

 Varianz, empirische, 15
 Verlustfunktion, 82
 verteilungsfrei, 66
 Verteilungsfunktion, empirische, 10
 Vertrauensintervall, 57
 Verzerrung, 50
 Visualisierung, 1
 Vollständigkeit, 100

 Wölbung, 12
 Wölbungsmaß von Fisher, 21
 Zielgröße, 33

Zufallsstichprobe, [5](#)

Zufallsvektoren

Dichtetransformationssatz, [47](#)
zweidimensionaler Kerndichteschätzer, [29](#)
zweidimensionales Histogramm, [27](#)



ulm university universität
uulm

Stochastik II

Vorlesungsskript
(Arbeitsversion)

Prof. Dr. Evgeny Spodarev

Dieses Exemplar wurde aus
Studiengebühren finanziert.

ULM
2010

Inhaltsverzeichnis

1	Allgemeine Theorie der zufälligen Funktionen	1
1.1	Zufällige Funktionen	1
1.2	Elementare Beispiele	5
1.3	Regularitätseigenschaften von Trajektorien	6
1.4	Differenzierbarkeit von Trajektorien	11
1.5	Momente und Kovarianz	12
1.6	Stationarität und Unabhängigkeit	14
1.7	Prozesse mit unabhängigen Zuwächsen	15
1.8	Ergänzende Aufgaben	17
2	Zählprozesse	20
2.1	Erneuerungsprozesse	20
2.2	Poisson-artige Prozesse	29
2.2.1	Poisson-Prozesse	29
2.2.2	Zusammengesetzter Poisson-Prozess	34
2.2.3	Cox-Prozess	36
2.3	Ergänzende Aufgaben	36
3	Wiener-Prozess	39
3.1	Elementare Eigenschaften	39
3.2	Explizite Konstruktion des Wiener-Prozesses	40
3.2.1	Haar- und Schauder-Funktionen	40
3.2.2	Wiener-Prozess mit f.s. stetigen Pfaden	43
3.3	Verteilungs- und Pfadeseigenschaften vom Wiener-Prozess	46
3.3.1	Verteilung des Maximums	46
3.3.2	Invarianzeigenschaften	48
3.4	Ergänzende Aufgaben	51
4	Lèvy Prozesse	53
4.1	Lèvy-Prozesse	53
4.1.1	Unbegrenzte Teilbarkeit	53
4.1.2	Lèvy-Chintschin-Darstellung	56
4.1.3	Beispiele	59
4.1.4	Subordinatoren	62
4.2	Ergänzende Aufgaben	65
5	Martingale	67
5.1	Grundbegriffe	67
5.2	(Sub-, Super-)Martingale	69
5.3	Gleichgradige Integrierbarkeit	71

5.4	Gestoppte Martingale	73
5.5	Lèvy-Prozesse und Martingale	77
5.6	Martingale und Wiener-Prozesse	78
5.7	Ergänzende Aufgaben	82
6	Stationäre Folgen von Zufallsvariablen	86
6.1	Reihen von unabhängigen Zufallsvariablen	86
6.2	Stationarität im engeren Sinne und Ergodentheorie	87
6.2.1	Grundbegriffe	87
6.2.2	Mischungseigenschaften und Ergodizität	90
6.2.3	Ergodensatz	93
6.3	Stationarität im weiteren Sinne	96
6.3.1	Korrelationstheorie	96
6.3.2	Orthogonale Zufallsmaße	97
6.3.3	Integral bezüglich eines orthogonalen Zufallsmaßes	98
6.3.4	Spektraldarstellung	99
6.4	Ergänzende Aufgaben	100
	Literaturverzeichnis	101

1 Allgemeine Theorie der zufälligen Funktionen

1.1 Zufällige Funktionen

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{S}, \mathcal{B})$ ein meßbarer Raum, $\Omega, \mathcal{S} \neq \emptyset$.

Definition 1.1.1

Ein *zufälliges Element* $X : \Omega \rightarrow \mathcal{S}$ ist eine $\mathcal{A}|\mathcal{B}$ -meßbare Abbildung (Bezeichnung: $X \in \mathcal{A}|\mathcal{B}$), d.h.,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \quad B \in \mathcal{B}.$$

Falls X ein zufälliges Element ist, dann ist $X(\omega)$ eine *Realisierung von X* für beliebige $\omega \in \Omega$.

Wir sagen, dass die σ -Algebra \mathcal{B} von Teilmengen von \mathcal{S} durch das Mengensystem \mathcal{M} erzeugt wird (\mathcal{M} enthält ebenso Teilmengen von \mathcal{S} als seine Elemente), wenn

$$\mathcal{B} = \bigcap_{\substack{\mathcal{F} \supset \mathcal{M} \\ \mathcal{F}\text{-}\sigma\text{-Algebra auf } \mathcal{S}}} \mathcal{F}$$

(Bezeichnung: $\mathcal{B} = \sigma(\mathcal{M})$).

Falls \mathcal{S} ein topologischer oder metrischer Raum ist, dann wählt man oft \mathcal{M} als Klasse aller offenen Mengen von \mathcal{S} und nennt $\sigma(\mathcal{M})$ Borelsche σ -Algebra (Bezeichnung: $\mathcal{B} = \mathcal{B}(\mathcal{S})$).

Beispiel 1.1.1 1. Falls $\mathcal{S} = \mathbb{R}$, $\mathcal{B} = \mathcal{B}(\mathbb{R})$, dann heißt ein zufälliges Element X eine *Zufallsvariable*.

2. Falls $\mathcal{S} = \mathbb{R}^m$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^m)$, $m > 1$, dann heißt X *Zufallsvektor*. Zufallsvariablen und Zufallsvektoren betrachtet man oft in den Vorlesungen „Elementare Wahrscheinlichkeitsrechnung und Statistik“ und „Stochastik I“.

3. Sei \mathcal{S} die Klasse aller abgeschlossenen Mengen von \mathbb{R}^m . Sei

$$\mathcal{M} = \{\{A \in \mathcal{S} : A \cap B \neq \emptyset\}, \quad B - \text{beliebiges Kompaktum aus } \mathbb{R}^m\}.$$

Dann ist $X : \Omega \rightarrow \mathcal{S}$ eine *zufällige abgeschlossene Menge*.

Als Beispiel betrachten wir n unabhängige gleichverteilte Punkte $Y_1, \dots, Y_n \in [0, 1]^m$ und $R_1, \dots, R_n > 0$ fast sicher unabhängige Zufallsvariablen, die auf dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ wie Y_1, \dots, Y_n definiert sind. Betrachten wir $X = \cup_{i=1}^n B_{R_i}(Y_i)$. Dies ist offensichtlich eine zufällige Menge. Eine beispielhafte Realisierung liefert Abbildung 1.1.

Aufgabe 1.1.1

Seien (Ω, \mathcal{A}) und $(\mathcal{S}, \mathcal{B})$ meßbare Räume, $\mathcal{B} = \sigma(\mathcal{M})$, wobei \mathcal{M} eine Klasse von Teilmengen von \mathcal{S} ist. Zeigen Sie, dass $X : \Omega \rightarrow \mathcal{S}$ genau dann $\mathcal{A}|\mathcal{B}$ -meßbar ist, wenn $X^{-1}(C) \in \mathcal{A}$, $C \in \mathcal{M}$.

Definition 1.1.2

Sei T eine beliebige Indexmenge und $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$ eine Familie von meßbaren Räumen. Eine Familie $X = \{X(t), t \in T\}$ von Zufallselementen $X(t) : \Omega \rightarrow \mathcal{S}_t$ definiert auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{A}|\mathcal{B}_t$ -meßbar für alle $t \in T$ heißt *zufällige Funktion* (assoziiert mit $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$).

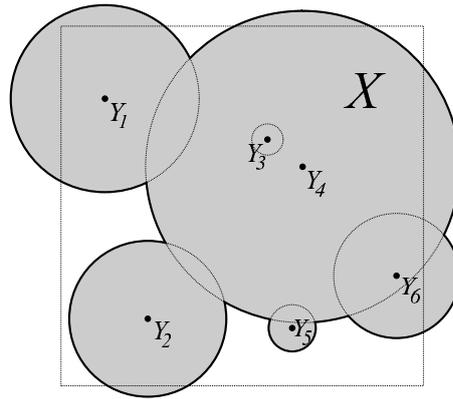


Abbildung 1.1: Beispiel einer zufälligen Menge $X = \cup_{i=1}^6 B_{R_i}(Y_i)$

Es gilt also $X : \Omega \times T \rightarrow (\mathcal{S}_t, t \in T)$, d.h. $X(\omega, t) \in \mathcal{S}_t$ für alle $\omega \in \Omega, t \in T$ und $X(\cdot, t) \in \mathcal{A} | \mathcal{B}_t, t \in T$. Sehr oft wird ω in der Bezeichnung unterlassen und man schreibt $X(t)$ an Stelle von $X(\omega, t)$. In den meisten Fällen hängt auch $(\mathcal{S}_t, \mathcal{B}_t)$ nicht von $t \in T$ ab: $(\mathcal{S}_t, \mathcal{B}_t) = (\mathcal{S}, \mathcal{B})$ für alle $t \in T$.

Spezialfälle zufälliger Funktionen:

1. $T \subseteq \mathbb{Z}$: X heißt dann *zufällige Folge* oder *stochastischer Prozess in diskreter Zeit*.
Beispiel: $T = \mathbb{Z}, \mathbb{N}$.
2. $T \subseteq \mathbb{R}$: X heißt *stochastischer Prozess in stetiger Zeit*.
Beispiel: $T = \mathbb{R}_+, [a, b], -\infty < a < b < \infty, \mathbb{R}$.
3. $T \subseteq \mathbb{R}^d, d \geq 2$: X heißt *zufälliges Feld*.
Beispiel: $T = \mathbb{Z}^d, \mathbb{R}_+^d, \mathbb{R}^d, [a, b]^d$.
4. $T \subseteq \mathcal{B}(\mathbb{R}^d)$: X heißt *Mengen-indizierter Prozess*.
Falls $X(t)$ fast sicher nichtnegativ und σ -additiv auf der σ -Algebra T ist, dann wird X *zufälliges Maß* genannt.

Die Tradition, die Indexmenge durch T zu bezeichnen, kommt von der Interpretation von $t \in T$ in den Fällen 1 und 2 als *Zeitparameter*.

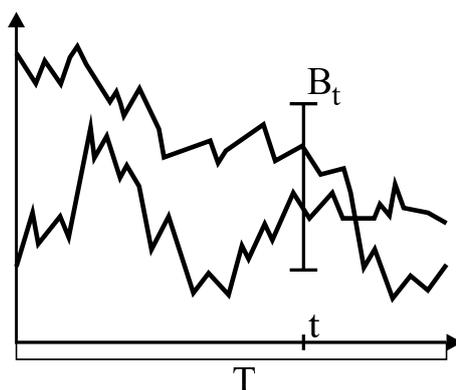
Für jedes $\omega \in \Omega$ heißt $\{X(\omega, t), t \in T\}$ eine *Trajektorie* bzw. ein *Pfad* der zufälligen Funktion X .

Wir möchten zeigen, dass die zufällige Funktion $X = \{X(t), t \in T\}$ ein zufälliges Element im entsprechenden Funktionsraum ist, welcher mit einer σ -Algebra ausgestattet ist, die jetzt spezifiziert wird.

Sei $\mathcal{S}_T = \prod_{t \in T} \mathcal{S}_t$ das kartesische Produkt von $\mathcal{S}_t, t \in T$, d.h., $X \in \mathcal{S}_T$ falls $X(t) \in \mathcal{S}_t, t \in T$. Die *elementare Zylindermenge* in \mathcal{S}_T wird definiert als

$$C_T(B_t) = \{X \in \mathcal{S}_T : X(t) \in B_t\},$$

wobei $t \in T$ ein ausgewählter Punkt aus T und $B_t \in \mathcal{B}_t$ eine Teilmenge in \mathcal{B}_t ist. $C_T(B_t)$ enthält also alle Trajektorien X , die durch das „Tor“ B_t gehen, siehe Abbildung 1.2.

Abbildung 1.2: Trajektorien, die ein „Tor“ B_t passieren.**Definition 1.1.3**

Die *zylindrische σ -Algebra* \mathcal{B}_T wird eingeführt als eine σ -Algebra erzeugt in \mathcal{S}_T durch die Familie von allen Elementarzylindern. Man bezeichnet sie durch $\mathcal{B}_T = \otimes_{t \in T} \mathcal{B}_t$. Falls $\mathcal{B}_t = \mathcal{B}$ für alle $t \in T$, dann schreibt man \mathcal{B}^T an Stelle von \mathcal{B}_T .

Lemma 1.1.1

Die Familie $\{X = X(t), t \in T\}$ ist eine zufällige Funktion auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Phasenräumen $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$ genau dann, wenn für jedes $\omega \in \Omega$ die Abbildung $\omega \mapsto X(\omega, \cdot)$ $\mathcal{A}|\mathcal{B}_T$ -meßbar ist.

Aufgabe 1.1.2

Beweisen Sie Lemma 1.1.1.

Definition 1.1.4

Sei X ein zufälliges Element: $X : \Omega \rightarrow \mathcal{S}$, d.h. X sei $\mathcal{A}|\mathcal{B}$ -meßbar. Die *Verteilung von X* ist das Wahrscheinlichkeitsmaß \mathbb{P}_X auf $(\mathcal{S}, \mathcal{B})$, so dass $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$, $B \in \mathcal{B}$.

Lemma 1.1.2

Ein beliebiges Wahrscheinlichkeitsmaß μ auf $(\mathcal{S}, \mathcal{B})$ kann als die Verteilung eines Zufallselementes X betrachtet werden.

Beweis Setze $\Omega = \mathcal{S}$, $\mathcal{A} = \mathcal{B}$, $\mathbb{P} = \mu$ und $X(\omega) = \omega$, $\omega \in \Omega$. □

Wann existiert eine zufällige Funktion mit vorgegebenen Eigenschaften? Eine zufällige Funktion, die aus unabhängigen Zufallselementen besteht, existiert immer. Diese Behauptung ist bekannt.

Theorem 1.1.1 (Lomnicki, Ulam):

Sei $(\mathcal{S}_t, \mathcal{B}_t, \mu_t)_{t \in T}$ eine Folge von Wahrscheinlichkeitsräumen. Es existiert eine zufällige Folge $X = \{X(t), t \in T\}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ (assoziiert mit $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$), so dass

1. $X(t)$, $t \in T$ unabhängige Zufallselemente sind.
2. $\mathbb{P}_{X(t)} = \mu_t$ auf $(\mathcal{S}_t, \mathcal{B}_t)$, $t \in T$.

Viele wichtige Zufallsprozesse sind auf Basis von unabhängigen zufälligen Elementen konstruiert; vgl. Beispiele im Abschnitt 1.2.

Definition 1.1.5

Sei $X = \{X(t), t \in T\}$ eine zufällige Funktion auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Phasenraum $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$. Die endlich-dimensionalen Verteilungen von X werden definiert als das Verteilungsgesetz $\mathbb{P}_{t_1, \dots, t_n}$ von $(X(t_1), \dots, X(t_n))^T$ auf $(\mathcal{S}_{t_1, \dots, t_n}, \mathcal{B}_{t_1, \dots, t_n})$, für beliebige $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, wobei $\mathcal{S}_{t_1, \dots, t_n} = \mathcal{S}_{t_1} \times \dots \times \mathcal{S}_{t_n}$ und $\mathcal{B}_{t_1, \dots, t_n} = \mathcal{B}_{t_1} \otimes \dots \otimes \mathcal{B}_{t_n}$, die σ -Algebra in $\mathcal{S}_{t_1, \dots, t_n}$ ist, die von allen Mengen $B_{t_1} \times \dots \times B_{t_n}$, $B_{t_i} \in \mathcal{B}_{t_i}$, $i = 1, \dots, n$, erzeugt wird, d.h., $\mathbb{P}_{t_1, \dots, t_n}(C) = \mathbb{P}((X(t_1), \dots, X(t_n))^T \in C)$, $C \in \mathcal{B}_{t_1, \dots, t_n}$. Insbesondere für $C = B_1 \times \dots \times B_n$, $B_k \in \mathcal{B}_{t_k}$:

$$\mathbb{P}_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = \mathbb{P}(X(t_1) \in B_1, \dots, X(t_n) \in B_n).$$

Aufgabe 1.1.3

Zeigen Sie, dass $X_{t_1, \dots, t_n} = (X(t_1), \dots, X(t_n))^T$ ein $\mathcal{A}|\mathcal{B}_{t_1, \dots, t_n}$ -meßbares Zufallselement ist.

Definition 1.1.6

Sei $\mathcal{S}_t = \mathbb{R}$ für alle $t \in T$. Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt *symmetrisch*, falls alle ihre endlich-dimensionalen Verteilungen symmetrische Wahrscheinlichkeitsmaße sind, d.h., $\mathbb{P}_{t_1, \dots, t_n}(A) = \mathbb{P}_{t_1, \dots, t_n}(-A)$ für $A \in \mathcal{B}_{t_1, \dots, t_n}$ und alle $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$. Dabei bedeutet $\mathbb{P}_{t_1, \dots, t_n}(-A) = \mathbb{P}((-X(t_1), \dots, -X(t_n))^T \in A)$.

Aufgabe 1.1.4

Zeigen Sie, dass die endlich-dimensionalen Verteilungen einer zufälligen Funktion X folgende Eigenschaften besitzen: für beliebiges $n \in \mathbb{N}$, $n \geq 2$, $\{t_1, \dots, t_n\} \subset T$, $B_k \in \mathcal{S}_{t_k}$, $k = 1, \dots, n$ und eine beliebige Permutation (i_1, \dots, i_n) von $(1, \dots, n)$ gilt:

1. *Symmetrie*: $\mathbb{P}_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = \mathbb{P}_{t_{i_1}, \dots, t_{i_n}}(B_{i_1} \times \dots \times B_{i_n})$
2. *Konsistenz*: $\mathbb{P}_{t_1, \dots, t_n}(B_1 \times \dots \times B_{n-1} \times \mathcal{S}_{t_n}) = \mathbb{P}_{t_1, \dots, t_{n-1}}(B_1 \times \dots \times B_{n-1})$

Folgender Satz zeigt, dass diese Eigenschaften hinreichend sind, um die Existenz einer zufälligen Funktion X mit vorgegebenen endlich-dimensionalen Verteilungen zu beweisen.

Theorem 1.1.2 (Kolmogorov):

Sei $\{\mathbb{P}_{t_1, \dots, t_n}, n \in \mathbb{N}, \{t_1, \dots, t_n\} \subset T\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^m \times \dots \times \mathbb{R}^m, \mathcal{B}(\mathbb{R}^m) \otimes \dots \otimes \mathcal{B}(\mathbb{R}^m))$, welche die Bedingungen 1 und 2 von Aufgabe 1.1.4 erfüllen. Dann existiert eine zufällige Funktion $X = \{X(t), t \in T\}$ definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit endlich-dimensionalen Verteilungen $\mathbb{P}_{t_1, \dots, t_n}$.

Beweis Siehe [13], Abschnitt II.9. □

Dieser Satz gilt auch auf allgemeineren (jedoch nicht beliebigen!) Räumen als \mathbb{R}^m , auf sog. *Borel-Räumen*, die in einem gewissen Sinne isomorph zu $([0, 1], \mathcal{B}[0, 1])$ oder einem Teilraum davon sind.

Definition 1.1.7

Sei $X = \{X(t), t \in T\}$ eine zufällige Funktion mit Werten in $(\mathcal{S}, \mathcal{B})$, d.h., $X(t) \in \mathcal{S}$ fast sicher für beliebige $t \in T$. X heißt *meßbar*, falls die Abbildung $X : (\omega, t) \mapsto X(\omega, t) \in \mathcal{S}$, $(\omega, t) \in \Omega \times T$, $\mathcal{A} \otimes \mathcal{C}|\mathcal{B}$ -meßbar ist.

Somit liefert die Definition 1.1.7 nicht nur die Meßbarkeit von X bzgl. $\omega \in \Omega$: $X(\cdot, t) \in \mathcal{A}|\mathcal{B}$ für alle $t \in T$, sondern $X(\cdot, \cdot) \in \mathcal{A} \otimes \mathcal{C}|\mathcal{B}$ als Funktion von (ω, t) . Die Meßbarkeit von X ist dann von Bedeutung, wenn $X(\omega, t)$ zu zufälligen Zeitpunkten $\tau : \Omega \rightarrow T$ betrachtet wird: $X(\omega, \tau(\omega))$. Dies ist insbesondere in der Martingaltheorie der Fall, wenn τ eine sog. Stoppzeit für X ist. Denn die Verteilung von $X(\omega, \tau(\omega))$ kann stark von der Verteilung von $X(\omega, t)$, $t \in T$, abweichen.

1.2 Elementare Beispiele

Für die explizite Konstruktion kann der Satz von Kolmogorov nur in wenigen Fällen direkt benutzt werden, da bei vielen zufälligen Funktionen ihre endlich-dimensionalen Verteilungen nicht explizit angegeben werden können. In diesen Fällen konstruiert man eine neue zufällige Funktion $X = \{X(t), t \in T\}$ als $X(t) = g(t, Y_1, Y_2, \dots)$, $t \in T$, wobei g eine meßbare Funktion ist und $\{Y_n\}$ eine Folge von Zufallselementen (auch zufälligen Funktionen) ist, deren Existenz bereits sichergestellt wurde. Hier geben wir einige Beispiele dafür.

Sei $X = \{X(t), t \in T\}$ eine reellwertige zufällige Funktion mit einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Weißes Rauschen:

Definition 1.2.1

Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt *weißes Rauschen*, falls alle $X(t)$, $t \in T$, unabhängige und identisch verteilte (u.i.v.) Zufallsvariablen sind.

Weißes Rauschen existiert nach dem Satz 1.1.1. Es wird verwendet um das Rauschen in (elektromagnetischen oder akustischen) Signalen darzustellen. Falls $X(t) \sim \text{Ber}(p)$, $p \in (0, 1)$, $t \in T$, so spricht man von *Salt-and-pepper Rauschen*, also vom binären Rauschen, das bei Übertragung von binären Daten in Computer-Netzwerken auftritt. Falls $X(t) \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$, $t \in T$, so wird X *Gauß'sches weißes Rauschen* genannt. Es tritt z.B. in akustischen Signalen auf.

2. Gauß'sche zufällige Funktion:

Definition 1.2.2

Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt *Gauß'sch*, falls alle ihre endlich-dimensionalen Verteilungen Gauß'sch sind, d.h. für alle $n \in \mathbb{N}$, $t_1, \dots, t_n \subset T$ gilt

$$X_{t_1, \dots, t_n} = ((X(t_1), \dots, X(t_n)))^\top \sim \mathcal{N}(\mu_{t_1, \dots, t_n}, \sum_{t_1, \dots, t_n}),$$

wobei der Mittelwert durch $\mu_{t_1, \dots, t_n} = (\mathbf{E}X(t_1), \dots, \mathbf{E}X(t_n))^\top$ und die Kovarianzmatrix durch $\sum_{t_1, \dots, t_n} = ((\text{cov}(X(t_i), X(t_j)))_{i,j=1}^n)$ gegeben ist.

Aufgabe 1.2.1

Zeigen Sie, dass die Verteilung einer Gauß'schen zufälligen Funktion X eindeutig durch ihre Mittelwertfunktion $\mu(t) = \mathbf{E}X(t)$, $t \in T$, bzw. Kovarianzfunktion $C(s, t) = \mathbf{E}[X(s)X(t)]$, $s, t \in T$, bestimmt wird.

Als Beispiel eines Gauß'schen Prozesses kann der sog. *Wiener-Prozess* (oder *Brown'sche Bewegung*) $X = \{X(t), t \geq 0\}$ dienen, der den Erwartungswert Null ($\mu(t) \equiv 0, t \geq 0$) und die Kovarianzfunktion $C(s, t) = \min\{s, t\}$, $s, t \geq 0$ hat. Normalerweise fordert man zusätzlich, dass die Pfade von X stetige Funktionen sind.

Die Regularitätseigenschaften der Pfade von zufälligen Funktionen werden wir detaillierter im Abschnitt 1.3 erforschen. Jetzt können wir sagen, dass ein solcher Prozess mit Wahrscheinlichkeit 1 (mit fast sicher stetigen Trajektorien) existiert.

Aufgabe 1.2.2

Zeigen Sie, dass Gauß'sches Weißes Rauschen eine Gauß'sche Zufallsfunktion ist.

3. Lognormal- und χ^2 -Funktionen:

Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt *lognormal*, falls $X(t) = e^{Y(t)}$, wobei $Y = \{Y(t), t \in T\}$ eine Gauß'sche zufällige Funktion ist. X heißt χ^2 -Funktion, falls $X(t) = \|Y(t)\|^2$, wobei $Y = \{Y(t), t \in T\}$ eine Gauß'sche zufällige Funktion mit Werten in \mathbb{R}^n ist, für die $Y(t) \sim \mathcal{N}(0, I)$, $t \in T$; hier ist I die $(n \times n)$ -Einheitsmatrix. Es gilt dann $X(t) \sim \chi_n^2$, $t \in T$.

4. Kosinus-Welle:

$X = \{X(t), t \in \mathbb{R}\}$ wird definiert durch $X(t) = \sqrt{2} \cos(2\pi Y + tZ)$, wobei $Y \sim \mathcal{U}([0, 1])$ und Z eine Zufallsvariable ist, die von Y unabhängig ist.

Aufgabe 1.2.3

Seien X_1, X_2, \dots u.i.v. Kosinus-Wellen. Bestimmen Sie den schwachen Grenzwert der endlich-dimensionalen Verteilungen der zufälligen Funktion $\left\{ \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k(t), t \in \mathbb{R} \right\}$ für $n \rightarrow \infty$.

5. Poisson-Prozess:

Sei $\{Y_n\}_{n \in \mathbb{N}}$ eine Folge von u.i.v. Zufallsvariablen $Y_n \sim \text{Exp}(\lambda)$, $\lambda > 0$. Der stochastische Prozess $X = \{X(t), t \geq 0\}$ definiert als $X(t) = \max\{n \in \mathbb{N} : \sum_{k=1}^n Y_k \leq t\}$ heißt *Poisson-Prozess* mit Intensität $\lambda > 0$. $X(t)$ zählt die Anzahl gewisser Ereignisse bis zum Zeitpunkt $t > 0$, wobei das typische Intervall zwischen zwei solchen Ereignissen eine $\text{Exp}(\lambda)$ -Verteilung besitzt. Diese Ereignisse können z.B. eine Schadensmeldung eines Versicherers, das Registrieren eines Elementarteilchens im Geigerzähler, usw. sein. Dann ist $X(t)$ die Schaden- bzw. Teilchenanzahl im Zeitintervall $[0, t]$.

1.3 Regularitätseigenschaften von Trajektorien

Der Satz von Kolmogorov gibt die Existenz der Verteilung einer zufälligen Funktion mit vorgegebenen endlich-dimensionalen Verteilungen an. Jedoch er sagt nichts über die Pfadeneigenschaften von X aus. Dies ist auch verständlich, denn alle zufälligen Objekte sind in der Wahrscheinlichkeitstheorie im fast sicheren Sinne (f.s.) definiert, also bis auf eine Menge $A \subset \Omega$ mit $\mathbb{P}(A) = 0$.

Beispiel 1.3.1

Sei $(\Omega, \mathcal{A}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \nu_1)$, wobei ν_1 das Lebesgue-Maß auf $[0, 1]$ ist. Definieren wir $\{X = X(t), t \in [0, 1]\}$ durch $X(t) \equiv 0$, $t \in [0, 1]$ und $Y = \{Y(t), t \in [0, 1]\}$ durch

$$Y(t) = \begin{cases} 1, & t = U, \\ 0, & \text{sonst,} \end{cases}$$

wobei $U(\omega) = \omega$, $\omega \in [0, 1]$, eine $\mathcal{U}([0, 1])$ -verteilte Zufallsvariable definiert auf $(\Omega, \mathcal{A}, \mathbb{P})$ ist. Da $\mathbb{P}(Y(t) = 0) = 1$, $t \in T$, ist, weil $\mathbb{P}(U = t) = 0$, $t \in T$, ist es klar, dass $X \stackrel{d}{=} Y$. Dennoch besitzen X und Y unterschiedliche Pfadeneigenschaften, da X stetige und Y sprunghafte Trajektorien hat, und $\mathbb{P}(X(t) = 0, \forall t \in T) = 1$, wobei $\mathbb{P}(Y(t) = 0, \forall t \in T) = 0$.

Es kann sein, dass die „Ausnahmemenge“ A (siehe oben) für $X(t)$ für jedes $t \in T$ sehr unterschiedlich ist. Deshalb fordert man, dass alle $X(t)$, $t \in T$, simultan auf einer Teilmenge $\Omega_0 \subseteq \Omega$ mit $\mathbb{P}(\Omega_0) = 1$ definiert sind. Die so definierte zufällige Funktion $\tilde{X} : \Omega_0 \times T \rightarrow \mathbb{R}$ heißt *Modifikation* von $X : \Omega \times T \rightarrow \mathbb{R}$. X und \tilde{X} unterscheiden sich auf einer Menge Ω/Ω_0 von Wahrscheinlichkeit Null. Deshalb meinen wir später, wenn wir sagen, dass „Zufällige Funktion X eine Eigenschaft C besitzt“ dass eine Modifikation von X mit dieser Eigenschaft C existiert.

Definition 1.3.1

Die zufälligen Funktionen $X = \{X(t), t \in T\}$ und $Y = \{Y(t), t \in T\}$ definiert auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ heißen (*stochastisch äquivalent*), falls

$$B_t = \{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t)\} \in \mathcal{A}, \quad t \in T,$$

und $\mathbb{P}(B_t) = 0, t \in T$.

Man sagt auch, dass X und Y Versionen einer und derselben zufälligen Funktion sind. Es ist klar, dass alle Modifikationen (oder Versionen) von X äquivalent zu X sind.

Aufgabe 1.3.1

Beweisen Sie, dass die zufälligen Funktionen X und Y im Beispiel 1.3.1 stochastisch äquivalent sind.

Definition 1.3.2

Die zufälligen Funktionen $X = \{X(t), t \in T\}$ und $Y = \{Y(t), t \in T\}$ (nicht unbedingt auf demselben Wahrscheinlichkeitsraum definiert) heißen *äquivalent in Verteilung*, falls $\mathbb{P}_X = \mathbb{P}_Y$ auf $(\mathcal{S}_t, \mathcal{B}_t)$. Bezeichnung: $X \stackrel{d}{=} Y$.

Nach dem Satz 1.1.2 ist es ausreichend für die Äquivalenz in Verteilung von X und Y , wenn sie dieselben endlich-dimensionalen Verteilungen besitzen. Es ist klar, dass die stochastische Äquivalenz die Äquivalenz in Verteilung impliziert, jedoch nicht umgekehrt.

Definition 1.3.3

Die zufälligen Funktionen $X = \{X(t), t \in T\}$ und $Y = \{Y(t), t \in T\}$ definiert auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ assoziiert mit $(\mathcal{S}_t, \mathcal{B}_t)_{t \in T}$ haben *äquivalente Trajektorien* (oder heißen auch *stochastisch ununterscheidbar*), falls

$$A = \{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t) \text{ für ein } t \in T\} \in \mathcal{A}$$

und $\mathbb{P}(A) = 0$.

Dieser Begriff bedeutet, dass X und Y Pfade haben, die mit Wahrscheinlichkeit 1 übereinstimmen. Falls der Raum $(\Omega, \mathcal{A}, \mathbb{P})$ *vollständig* ist (d.h. aus $A \in \mathcal{A} : \mathbb{P}(A) = 0$ folgt für alle $B \subset A : B \in \mathcal{A}$ (und dann $\mathbb{P}(B) = 0$)), dann sind ununterscheidbare Prozesse stochastisch äquivalent.

Seien nun T und \mathcal{S} *Banach-Räume* mit den Normen $|\cdot|_T$ bzw. $|\cdot|_{\mathcal{S}}$. Die zufällige Funktion $X = \{X(t), t \in T\}$ sei nun auf $(\Omega, \mathcal{A}, \mathbb{P})$ definiert mit Werten in $(\mathcal{S}, \mathcal{B})$.

Definition 1.3.4

Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt

- a) *stochastisch stetig auf T* , falls $X(s) \xrightarrow[s \rightarrow t]{\mathbb{P}} X(t)$, für beliebige $t \in T$, d.h.

$$\mathbb{P}(|X(s) - X(t)|_{\mathcal{S}} > \varepsilon) \xrightarrow[s \rightarrow t]{} 0, \text{ für alle } \varepsilon > 0.$$

- b) *L^p -stetig auf T* , $p \geq 1$, falls $X(s) \xrightarrow[s \rightarrow t]{L^p} X(t)$, $t \in T$, d.h. $\mathbb{E}|X(s) - X(t)|^p \xrightarrow[s \rightarrow t]{} 0$. Für $p = 2$ benutzt man die spezielle Bezeichnung „Stetigkeit im quadratischen Mittel“.

- c) *f.s. stetig auf T* , falls $X(s) \xrightarrow[s \rightarrow t]{f.s.} X(t)$, $t \in T$, d.h., $\mathbb{P}(X(s) \xrightarrow[s \rightarrow t]{} X(t)) = 1, t \in T$.

d) *stetig*, falls alle Trajektorien von X stetige Funktionen sind.

In Anwendungen interessiert man sich für die Fälle c) und d), obwohl die schwächste Form der Stetigkeit die stochastische Stetigkeit ist.

$$\boxed{L^p\text{-Stetigkeit}} \implies \boxed{\text{Stochastische Stetigkeit}} \iff \boxed{\text{f.s. Stetigkeit}} \iff \boxed{\text{Stetigkeit aller Pfade}}$$

Warum sind Fälle c) und d) wichtig? Betrachten wir ein Beispiel.

Beispiel 1.3.2

Sei $T = [0, 1]$ und $(\Omega, \mathcal{A}, \mathbb{P})$ sei ein kanonischer Wahrscheinlichkeitsraum mit $\Omega = \mathbb{R}^{[0,1]}$, d.h. $\Omega = \prod_{t \in [0,1]} \mathbb{R}$. Sei $X = \{X(t), t \in [0, 1]\}$ ein stochastischer Prozess auf $(\Omega, \mathcal{A}, \mathbb{P})$. Nicht alle Ereignisse sind aber Elemente von \mathcal{A} , wie z.B. $A = \{\omega \in \Omega : X(\omega, t) = 0 \text{ für alle } t \in [0, 1]\} = \bigcap_{t \in [0,1]} \{X(\omega, t) = 0\}$, weil dies ein Schnitt von messbaren Ereignissen aus \mathcal{A} in überzählbarer Anzahl ist. Falls allerdings X stetig ist, dann sind auch alle seine Pfade stetige Funktionen und man kann $A = \bigcap_{t \in D} \{X(\omega, t) = 0\}$ darstellen lassen, wobei D eine dichte abzählbare Teilmenge von $[0, 1]$ ist, z.B., $D = \mathbb{Q} \cap [0, 1]$. Dann gilt aber $A \in \mathcal{A}$.

Es ist allerdings in vielen Anwendungen (wie z.B. in der Finanzmathematik) nicht realistisch, stochastische Prozesse mit stetigen Pfaden als Modelle für reale Phänomene zu betrachten. Deshalb wird eine größere Klasse von möglichen Trajektorien von X erlaubt: die sog. *càdlàg-Klasse* (*càdlàg* = *continue à droite, limitée à gauche* (fr.)).

Definition 1.3.5

Ein stochastischer Prozess $X = \{X(t), t \in \mathbb{R}\}$ heißt *càdlàg*, wenn alle seine Trajektorien rechtsseitig stetige Funktionen sind, die linksseitige Grenzwerte besitzen.

Jetzt wollen wir die Eigenschaften der oben eingeführten Stetigkeitsbegriffen näher betrachten. Es stellt sich z.B. fest, dass die stochastische Stetigkeit eine Eigenschaft der zweidimensionalen Verteilung $\mathbb{P}_{s,t}$ von X ist, wie folgendes Lemma zeigt.

Lemma 1.3.1

Sei $X = \{X(t), t \in T\}$ eine zufällige Funktion assoziiert mit $(\mathcal{S}, \mathcal{B})$, wobei \mathcal{S} und T Banach-Räume sind. Folgende Aussagen sind äquivalent:

- a) $X(s) \xrightarrow[s \rightarrow t_0]{\mathbb{P}} Y$,
- b) $\mathbb{P}_{s,t} \xrightarrow[s,t \rightarrow t_0]{d} \mathbb{P}_{(Y,Y)}$,

wobei $t_0 \in T$ und Y ein \mathcal{B} -Zufallselement ist. Für die stochastische Stetigkeit von X sollen $t_0 \in T$ beliebig und $Y = X(t_0)$ gewählt werden.

Beweis a) \Rightarrow b)

$X(s) \xrightarrow[s \rightarrow t_0]{\mathbb{P}} Y$ bedeutet $(X(s), X(t))^\top \xrightarrow[s,t \rightarrow t_0]{\mathbb{P}} (Y, Y)^\top$. Daraus folgt $\mathbb{P}_{s,t} \xrightarrow{d} \mathbb{P}_{(Y,Y)}$, weil $\xrightarrow{\mathbb{P}}$ -

Konvergenz strenger als \xrightarrow{d} -Konvergenz ist.

b) \Rightarrow a)

Für beliebiges $\varepsilon > 0$ betrachten wir eine stetige Funktion $g_\varepsilon : \mathbb{R} \rightarrow [0, 1]$ mit $g_\varepsilon(0) = 0$, $g_\varepsilon(x) = 1$, $x \notin B_\varepsilon(0)$. Es gilt für alle $s, t \in T$, dass

$$\mathbb{E}g_\varepsilon(|X(s) - X(t)|_{\mathcal{S}}) = \mathbb{P}(|X(s) - X(t)|_{\mathcal{S}} > \varepsilon) + \mathbb{E}(g_\varepsilon(|X(s) - X(t)|_{\mathcal{S}}) \mathbb{E}(|X(s) - X(t)|_{\mathcal{S}} | \leq \varepsilon)),$$

daher $\mathbb{P}(|X(s) - X(t)|_{\mathcal{S}} > \varepsilon) \leq \mathbb{E}g_\varepsilon(|X(s) - X(t)|_{\mathcal{S}}) = \int_{\mathcal{S}} \int_{\mathcal{S}} g_\varepsilon(|x - y|_{\mathcal{S}}) \mathbb{P}_{s,t}(d(x, y)) \xrightarrow[s \rightarrow t_0]{t \rightarrow t_0} \int_{\mathcal{S}} \int_{\mathcal{S}} g_\varepsilon(|x - y|_{\mathcal{S}}) \mathbb{P}_{(Y,Y)}(d(x, y)) = 0$, weil $\mathbb{P}_{(Y,Y)}$ auf $\{(x, y) \in \mathcal{S}^2 : x = y\}$ konzentriert ist und $g_\varepsilon(0) = 0$. Daher ist $\{X(s)\}_{s \rightarrow t_0}$ eine fundamentale Folge (in Wahrscheinlichkeit), weshalb $X(s) \xrightarrow[s \rightarrow t_0]{\mathbb{P}} Y$. \square

Es kann sein, dass X stochastisch stetig ist, obwohl alle Pfade von X Sprünge haben, d.h. X kann keine f.s. stetige Modifikation besitzen. Die anschauliche Erklärung dessen ist, dass solche X mit Wahrscheinlichkeit Null einen Sprung für konkretes $t \in T$ haben können. Deshalb treten Sprünge der Pfade von X immer an anderen Stellen $t \in T$ auf.

Aufgabe 1.3.2

Zeigen Sie, dass der Poisson-Prozess stochastisch stetig ist, obwohl er keine f.s. stetige Modifikation besitzt.

Aufgabe 1.3.3

Sei T kompakt. Zeigen Sie, dass falls X stochastisch stetig auf T ist, dann ist es auch gleichmäßig stochastisch stetig, d.h., für alle $\varepsilon, \eta > 0 \exists \delta > 0$, so dass für alle $s, t \in T$ mit $|s - t|_T < \delta$ gilt: $\mathbb{P}(|X(s) - X(t)|_{\mathcal{S}} > \varepsilon) < \eta$.

Nun sei $\mathcal{S} = \mathbb{R}$, $\mathbb{E}X^2(t) < \infty$, $t \in T$, $\mathbb{E}X(t) = 0$, $t \in T$. Sei $C(s, t) = \mathbb{E}[X(s)X(t)]$ die Kovarianzfunktion von X .

Lemma 1.3.2

Für alle $t_0 \in T$ und eine Zufallsvariable Y mit $\mathbb{E}Y^2 < \infty$ sind folgende Behauptungen äquivalent:

- a) $X(s) \xrightarrow[s \rightarrow t_0]{L^2} Y$
- b) $C(s, t) \xrightarrow[s, t \rightarrow t_0]{} \mathbb{E}Y^2$

Beweis a) \Rightarrow b)

Die Behauptung folgt aus der Ungleichung von Cauchy-Schwarz:

$$\begin{aligned} |C(s, t) - \mathbb{E}Y^2| &= |\mathbb{E}(X(s)X(t)) - \mathbb{E}Y^2| = |\mathbb{E}[(X(s) - Y + Y)(X(t) - Y + Y)] - \mathbb{E}Y^2| \\ &\leq \mathbb{E}|(X(s) - Y)(X(t) - Y)| + \mathbb{E}|(X(s) - Y)Y| + \mathbb{E}|(X(t) - Y)Y| \\ &\leq \sqrt{\underbrace{\mathbb{E}(X(s) - Y)^2}_{\|X(s)-Y\|_{L^2}^2} \mathbb{E}(X(t) - Y)^2} \\ &\quad + \sqrt{\underbrace{\mathbb{E}Y^2 \mathbb{E}(X(s) - Y)^2}_{\|X(s)-Y\|_{L^2}^2}} + \sqrt{\underbrace{\mathbb{E}Y^2 \mathbb{E}(X(t) - Y)^2}_{\|X(t)-Y\|_{L^2}^2}} \xrightarrow[s, t \rightarrow t_0]{} 0 \end{aligned}$$

nach Voraussetzung a).

b) \Rightarrow a)

$$\begin{aligned} \mathbb{E}(X(s) - X(t))^2 &= \mathbb{E}(X(s))^2 - 2\mathbb{E}[X(s)X(t)] + \mathbb{E}(X(t))^2 \\ &= C(s, s) + C(t, t) - 2C(s, t) \xrightarrow[s, t \rightarrow t_0]{} 2\mathbb{E}Y^2 - 2\mathbb{E}Y^2 = 0. \end{aligned}$$

Daher ist $\{X(s), s \rightarrow t_0\}$ eine fundamentale Folge im L^2 -Sinne, und es folgt $X(s) \xrightarrow[s \rightarrow t_0]{L^2} Y$. \square

Eine zufällige Funktion X , die stetig im mittleren quadratischen Sinne ist, kann immer noch unstetige Trajektorien besitzen. In den meisten Fällen, die praktische Relevanz besitzen, hat X jedoch eine f.s. stetige Modifikation. Dies werden wir später in Form eines Satzes präziser machen.

Folgerung 1.3.1

Die zufällige Funktion X , die den Voraussetzungen des Lemmas 1.3.2 genügt, ist stetig auf T im mittleren quadratischen Sinne genau dann, wenn ihre Kovarianzfunktion $C : T^2 \rightarrow \mathbb{R}$ stetig auf der Diagonalen $\text{diag } T^2 = \{(s, t) \in T^2 : s = t\}$ ist, d.h., $\lim_{s,t \rightarrow t_0} C(s, t) = C(t)$ für alle $t_0 \in T$.

Beweis Wähle $Y = X(t_0)$ in Lemma 1.3.2. □

Bemerkung 1.3.1

Falls X nicht zentriert ist, dann fordert man die Stetigkeit von $\mu(\cdot)$ zusammen mit der Stetigkeit von C auf $\text{diag } T^2$, um die L^2 -Stetigkeit von X auf T zu gewährleisten.

Aufgabe 1.3.4

Geben Sie ein Beispiel eines stochastischen Prozesses mit f.s. unstetigen Trajektorien, der L^2 -stetig ist.

Nun betrachten wir die Eigenschaft der (f.s.) Stetigkeit etwas näher. Wie vorher erwähnt, können wir lediglich von einer stetigen Modifikation oder Version eines Prozesses sprechen. Die Möglichkeit, eine solche Version zu besitzen, hängt ebenso von den Eigenschaften der zweidimensionalen Verteilungen des Prozesses ab, wie folgender Satz (ursprünglich bewiesen von A. Kolmogorov) zeigt.

Theorem 1.3.1

Sei $X = \{X(t), t \in [a, b]\}$, $-\infty < a < b \leq +\infty$, ein reellwertiger stochastischer Prozess X hat eine stetige Version, falls es Konstanten $\alpha, c, \delta > 0$ gibt, so dass

$$\mathbb{E}|X(t+h) - X(t)|^\alpha < c|h|^{1+\delta}, \quad t \in (a, b), \quad (1.3.1)$$

für ausreichend kleine $|h|$.

Beweis Siehe, z.B. [7], Theorem 2.23. □

Nun wenden wir uns den Prozessen mit càdlàg-Trajektorien zu. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein vollständiger Wahrscheinlichkeitsraum.

Theorem 1.3.2

Sei $X = \{X(t), t \geq 0\}$ ein reellwertiger stochastischer Prozess und D eine abzählbare dichte Teilmenge von $[0, \infty)$. Falls

- a) X stochastisch rechtsseitig stetig ist, d.h., $X(t+h) \xrightarrow[h \rightarrow +0]{\mathbb{P}} X(t)$, $t \in [0, +\infty)$,
- b) die Trajektorien von X für jedes $t \in D$ endliche rechts- und linksseitige Grenzwerte haben, d.h., $\exists \lim_{h \rightarrow \pm 0} X(t+h)$, $t \in D$ f.s.,

dann hat X eine Version mit f.s. càdlàg-Pfaden.

Ohne Beweis.

Lemma 1.3.3

Seien $X = \{X(t), t \geq 0\}$ und $\{Y = Y(t), t \geq 0\}$ zwei Versionen einer zufälligen Funktion, beide definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, mit der Eigenschaft, dass X und Y f.s. rechtsseitig stetige Trajektorien haben. Dann sind X und Y ununterscheidbar.

Beweis Seien Ω_X, Ω_Y „Ausnahmemengen“, für die die Trajektorien von X bzw. von Y nicht rechtsseitig stetig sind. Es gilt $\mathbb{P}(\Omega_X) = \mathbb{P}(\Omega_Y) = 0$. Betrachte $A_t = \{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t)\}$, $t \in [0, +\infty)$ und $A = \cup_{t \in \mathbb{Q}_+} A_t$, wobei $\mathbb{Q}_+ = \mathbb{Q} \cap [0, +\infty)$. Da X und Y stochastisch äquivalent sind, gilt $\mathbb{P}(A) = 0$ und deshalb $\mathbb{P}(\tilde{A}) = \mathbb{P}(A \cup \Omega_X \cup \Omega_Y) \leq \mathbb{P}(A) + \mathbb{P}(\Omega_X) + \mathbb{P}(\Omega_Y) = 0$, wobei $\tilde{A} = A \cup \Omega_X \cup \Omega_Y$. Somit gilt $X(\omega, t) = Y(\omega, t)$ für $t \in \mathbb{Q}_+$ und $\omega \in \Omega \setminus \tilde{A}$. Wir beweisen dies nun für alle $t \geq 0$. Für beliebiges $t \geq 0$ existiert eine Folge $\{t_n\} \subset \mathbb{Q}_+$, so dass $t_n \downarrow t$. Da $X(\omega, t_n) = Y(\omega, t_n)$ für alle $n \in \mathbb{N}$ und $\omega \in \Omega \setminus \tilde{A}$, gilt $X(\omega, t) = \lim_{n \rightarrow \infty} X(\omega, t_n) = \lim_{n \rightarrow \infty} Y(\omega, t_n) = Y(\omega, t)$ für $t \geq 0$ und $\omega \in \Omega \setminus \tilde{A}$. Deshalb sind X und Y ununterscheidbar. \square

Folgerung 1.3.2

Falls càdlàg-Prozesse $X = \{X(t), t \geq 0\}$ und $Y = \{Y(t), t \geq 0\}$ Versionen einer zufälligen Funktion sind, dann sind sie ununterscheidbar.

1.4 Differenzierbarkeit von Trajektorien

Sei T ein linearer normierter Raum.

Definition 1.4.1

Eine reellwertige zufällige Funktion $X = \{X(t), t \in T\}$ ist *differenzierbar auf T in Richtung $h \in T$ stochastisch, im L^p -Sinne, $p \geq 1$, oder f.s., falls es*

$$\lim_{l \rightarrow 0} \frac{X(t+hl) - X(t)}{l} = X'_h(t), \quad t \in T$$

existiert im entsprechenden Sinne, also stochastisch, im L^p -Raum oder f.s..

Die Lemmata 1.3.2 - 1.3.3 zeigen, dass die stochastische Differenzierbarkeit eine Eigenschaft ist, die durch dreidimensionale Verteilungen von X bestimmt ist (weil die gemeinsame Verteilung von $\frac{X(t+hl) - X(t)}{l}$ und $\frac{X(t+hl') - X(t)}{l'}$ schwach konvergieren soll), wobei die Differenzierbarkeit im mittleren quadratischen Sinne durch die Glattheit der Kovarianzfunktion $C(s, t)$ bestimmt wird.

Aufgabe 1.4.1

Zeigen Sie, dass

1. der Wiener-Prozess nicht stochastisch differenzierbar auf $[0, \infty)$ ist.
2. der Poisson-Prozess stochastisch differenzierbar auf $[0, \infty)$ ist, jedoch nicht im L^p -Mittel, $p \geq 1$.

Lemma 1.4.1

Eine zentrierte zufällige Funktion $X = \{X(t), t \in T\}$ (d.h., $\mathbb{E}X(t) \equiv 0$, $t \in T$), ist L^2 -differenzierbar in $t \in T$ in Richtung $h \in T$, falls ihre Kovarianzfunktion C zweimal differenzierbar in (t, t) in Richtung h ist, d.h., falls $C''_{hh}(t, t) = \frac{\partial^2 C(s, t)}{\partial s_h \partial t_h} \Big|_{s=t}$. $X'_h(t)$ ist L^2 -stetig in

$t \in T$, falls $C''_{hh}(s, t) = \frac{\partial^2 C(s, t)}{\partial s_h \partial t_h}$ stetig in $s = t$ ist. Daher ist $C''_{hh}(s, t)$ die Kovarianzfunktion von $X'_h = \{X'_h(t), t \in T\}$.

Beweis Nach Lemma 1.3.3 reicht es zu zeigen, dass

$$I = \lim_{l, l' \rightarrow 0} \mathbb{E} \left(\frac{X(t+lh) - X(t)}{l} \cdot \frac{X(s+l'h) - X(s)}{l'} \right)$$

existiert für $s = t$. In der Tat erhalten wir

$$\begin{aligned} I &= \frac{1}{ll'} \left(C(t+lh, s+l'h) - C(t+lh, s) - C(t, s+l'h) + C(t, s) \right) \\ &= \frac{1}{l} \left(\frac{C(t+lh, s+l'h) - C(t+lh, s)}{l'} - \frac{C(t, s+l'h) - C(t, s)}{l'} \right) \xrightarrow{l, l' \rightarrow 0} C''_{hh}(s, t). \end{aligned}$$

Alle anderen Aussagen des Lemmas folgen aus dieser Relation. \square

Bemerkung 1.4.1

Die Eigenschaften der L^2 -Differenzierbarkeit und der f.s.-Differenzierbarkeit von zufälligen Funktionen sind definiert im folgenden Sinne: es gibt stochastische Prozesse, die L^2 -differenzierbare Pfade haben, obwohl sie f.s. unstetig sind, und umgekehrt sind Prozesse mit f.s. differenzierbaren Pfaden nicht immer L^2 -differenzierbar, weil z.B. die erste Ableitung ihrer Kovarianzfunktion nicht stetig ist.

Aufgabe 1.4.2

Geben Sie entsprechende Beispiele an!

1.5 Momente und Kovarianz

Sei $X = \{X(t), t \in T\}$ eine zufällige Funktion, die reellwertig ist, und sei T ein beliebiger Indexraum.

Definition 1.5.1

Das *gemischte Moment* $\mu^{(j_1, \dots, j_n)}(t_1, \dots, t_n)$ von X der Ordnung $(j_1, \dots, j_n) \in \mathbb{N}^n$, $t_1, \dots, t_n \in T$ ist gegeben durch $\mu^{(j_1, \dots, j_n)}(t_1, \dots, t_n) = \mathbb{E}[X^{j_1}(t_1) \cdot \dots \cdot X^{j_n}(t_n)]$, vorausgesetzt, dass dieser Erwartungswert existiert und endlich ist. Dann ist es ausreichend vorauszusetzen, dass $\mathbb{E}|X(t)|^j < \infty$ für alle $t \in T$ und $j = j_1 + \dots + j_n$.

Wichtige Spezialfälle:

1. $\mu(t) = \mu^{(1)}(t) = \mathbb{E}X(t)$, $t \in T$ – *Mittelwertfunktion von X .*
2. $\mu^{(1,1)}(s, t) = \mathbb{E}[X(s)X(t)] = C(s, t)$ – *(nicht-zentrierte) Kovarianzfunktion von X .* Sie ist zu unterscheiden von der *zentrierten Kovarianzfunktion*: $K(s, t) = \text{cov}(X(s), X(t)) = \mu^{(1,1)}(s, t) - \mu(s)\mu(t)$, $s, t \in T$.

Aufgabe 1.5.1

Zeigen Sie, dass die zentrierte Kovarianzfunktion einer reellwertigen zufälligen Funktion X

1. *symmetrisch* ist, d.h., $K(s, t) = K(t, s)$, $s, t \in T$.

2. *positiv semidefinit* ist, d.h., für alle $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $z_1, \dots, z_n \in \mathbb{R}$ gilt

$$\sum_{i,j=1}^n K(t_i, t_j) z_i z_j \geq 0.$$

3. $K(t, t) = \text{var } X(t)$ erfüllt, $t \in T$.

Die Eigenschaft 2) gilt auch für die nicht-zentrierte Kovarianzfunktion $C(s, t)$.

Die Mittelwertfunktion $\mu(t)$ zeigt einen (nicht zufälligen) Trend dar. Falls sie bekannt ist, kann die zufällige Funktion X zentriert werden, indem man eine zufällige Funktion $Y = \{Y(t), t \in T\}$ mit $Y(t) = X(t) - \mu(t)$, $t \in T$ betrachtet.

Die Kovarianzfunktion $K(s, t)$ bzw. $C(s, t)$ enthält Informationen über die Abhängigkeitsstruktur von X . Manchmal wird statt K bzw. C die Korrelationsfunktion $R(s, t) = \frac{K(s, t)}{\sqrt{K(s, s)K(t, t)}}$ verwendet, für alle $s, t \in T$: $K(s, s) = \text{var } X(s) > 0$, $K(t, t) = \text{var } X(t) > 0$. Durch die Ungleichung von Cauchy-Schwarz gilt $|R(s, t)| \leq 1$, $s, t \in T$. Die Menge aller gemischten Momente legt die Verteilung einer zufälligen Funktion im Allgemeinen nicht (eindeutig) fest.

Aufgabe 1.5.2

Geben Sie Beispiele von verschiedenen zufälligen Funktionen $X = \{X(t), t \in T\}$ und $Y = \{Y(t), t \in T\}$, für die gilt $\mathbf{E}X(t) = \mathbf{E}Y(t)$, $t \in T$ und $\mathbf{E}(X(s)X(t)) = \mathbf{E}(Y(s)Y(t))$, $s, t \in T$.

Aufgabe 1.5.3

Sei $\mu : T \rightarrow \mathbb{R}$ eine beliebige Funktion und $K : T \times T \rightarrow \mathbb{R}$ eine positiv semidefinite symmetrische Funktion. Zeigen Sie, dass eine zufällige Funktion $X = \{X(t), t \in T\}$ existiert mit $\mathbf{E}X(t) = \mu(t)$, $\text{cov}(X(s), X(t)) = C(s, t)$, $s, t \in T$.

Sei nun $X = \{X(t), t \in T\}$ eine reellwertige zufällige Funktion mit $\mathbf{E}|X(t)|^k < \infty$, $t \in T$, für ein $k \in \mathbb{N}$.

Definition 1.5.2

Der *mittlere Zuwachs der Ordnung k* von X ist gegeben durch $\gamma_k(s, t) = \mathbf{E}(X(s) - X(t))^k$, $s, t \in T$.

Besondere Aufmerksamkeit gilt der Funktion $\gamma(s, t) = \frac{1}{2}\gamma_2(s, t) = \frac{1}{2}\mathbf{E}(X(s) - X(t))^2$, $s, t \in T$, die *Variogramm von X* genannt wird. Das Variogramm wird in Geostatistik oft an Stelle der Kovarianzfunktion benutzt. Oft wird dafür die Bedingung $\mathbf{E}X^2(t) < \infty$, $t \in T$ nicht gestellt, sondern es wird vorausgesetzt, dass $\gamma(s, t) < \infty$ für alle $s, t \in T$.

Aufgabe 1.5.4

Zeigen Sie, dass es zufällige Funktion ohne endlichen 2. Momenten mit $\gamma(s, t) < \infty$, $s, t \in T$ gibt.

Aufgabe 1.5.5

Zeigen Sie, dass für eine zufällige Funktion $X = \{X(t), t \in T\}$ mit Mittelwertfunktion μ und Kovarianzfunktion K gilt:

$$\gamma(s, t) = \frac{K(s, s) + K(t, t)}{2} - K(s, t) + \frac{1}{2}(\mu(s) - \mu(t))^2, \quad s, t \in T.$$

Falls die zufällige Funktion X *komplexwertig* ist, d.h., $X : \Omega \times T \rightarrow \mathbb{C}$, mit $\mathbf{E}|X(t)|^2 < \infty$, $t \in T$, dann wird die Kovarianzfunktion von X als $K(s, t) = \mathbf{E}(X(s) - \mathbf{E}X(s))(\overline{X(t) - \mathbf{E}X(t)})$,

$s, t \in T$, eingeführt, wobei \bar{z} das Komplex-konjugierte von $z \in \mathbb{C}$ ist. Es gilt dann $K(s, t) = \overline{K(t, s)}$, $s, t \in T$, und K ist positiv semidefinit, d.h., für alle $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $z_1, \dots, z_n \in \mathbb{C}$ gilt $\sum_{i,j=1}^n K(t_i, t_j) z_i \bar{z}_j \geq 0$.

1.6 Stationarität und Unabhängigkeit

Sei T eine Teilmenge vom linearen Vektorraum mit Operationen $+$, $-$ über den Raum \mathbb{R} .

Definition 1.6.1

Die zufällige Funktion $X = \{X(t), t \in T\}$ heißt *stationär* (im engen Sinne), falls für alle $n \in \mathbb{N}$, $h, t_1, \dots, t_n \in T$ mit $t_1 + h, \dots, t_n + h \in T$ gilt:

$$P_{(X(t_1), \dots, X(t_n))} = P_{(X(t_1+h), \dots, X(t_n+h))},$$

d.h., alle endlich-dimensionalen Verteilungen von X sind invariant gegenüber Verschiebungen in T .

Definition 1.6.2

Eine (komplexwertige) zufällige Funktion $X = \{X(t), t \in T\}$ heißt *stationär 2. Ordnung* (oder *im weiten Sinne*), falls $E|X(t)|^2 < \infty$, $t \in T$, und $\mu(t) \equiv EX(t) \equiv \mu$, $t \in T$, $K(s, t) = \text{cov}(X(s), X(t)) = K(s+h, t+h)$ für alle $h, s, t \in T : s+h, t+h \in T$.

Falls X stationär 2. Ordnung ist, ist es günstig eine Funktion $K(t) := K(0, t)$, $t \in T$, einzuführen, wobei $0 \in T$ ist.

Stationarität im engen Sinne und 2. Ordnung folgen nicht aus einander. Es ist jedoch klar, dass, wenn eine komplexwertige zufällige Funktion stationär im engen Sinne ist und endliche 2. Momente besitzt, dann ist sie auch stationär 2. Ordnung.

Definition 1.6.3

Eine reellwertige zufällige Funktion $X = \{X(t), t \in T\}$ ist *intrinsisch stationär 2. Ordnung*, falls $\gamma_k(s, t)$, $s, t \in T$ existieren für $k \leq 2$, und es gilt für alle $s, t, h \in T$, $s+h, t+h \in T$, dass $\gamma_1(s, t) = 0$, $\gamma_2(s, t) = \gamma_2(s+h, t+h)$.

Die intrinsische Stationarität 2. Ordnung ist für reellwertige zufällige Funktionen etwas allgemeiner als Stationarität 2. Ordnung, da die Existenz von $E|X(t)|^2$, $t \in T$, nicht gefordert wird.

Es gibt aber auch das Analogon der Stationarität der Zuwächse von X im engen Sinne.

Definition 1.6.4

Sei $X = \{X(t), t \in T\}$ ein reellwertiger stochastischer Prozess, $T \subset \mathbb{R}$. Man sagt, dass X

1. *stationäre Zuwächse* besitzt, falls für alle $n \in \mathbb{N}$, $h, t_0, t_1, t_2, \dots, t_n \in T$, mit $t_0 < t_1 < t_2 < \dots < t_n$, $t_i + h \in T$, $i = 0, \dots, n$ die Verteilung von $(X(t_1 + h) - X(t_0 + h), \dots, X(t_n + h) - X(t_{n-1} + h))^T$ nicht von h abhängt.
2. *unabhängige Zuwächse* besitzt, falls für alle $n \in \mathbb{N}$, $t_0, t_1, \dots, t_n \in T$ mit $t_0 < t_1 < \dots < t_n$ die Zufallsvariablen $X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1})$ paarweise unabhängig.

Seien $(\mathcal{S}_1, \mathcal{B}_1)$ und $(\mathcal{S}_2, \mathcal{B}_2)$ meßbare Räume. Generell sagt man, dass zwei zufällige Elemente $X : \Omega \rightarrow \mathcal{S}_1$ und $X : \Omega \rightarrow \mathcal{S}_2$ auf dem selben Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) *unabhängig* sind, wenn $P(X \in A_1, Y \in A_2) = P(X \in A_1)P(Y \in A_2)$ für alle $A_1 \in \mathcal{B}_1$, $A_2 \in \mathcal{B}_2$.

Diese Definition läßt sich übertragen auf die Unabhängigkeit von zufälligen Funktionen X und Y mit dem Phasenraum $(\mathcal{S}_T, \mathcal{B}_T)$, da sie als zufällige Elemente angesehen werden können, mit $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}_T$, $\mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}_T$ (vgl. Lemma 1.1.1). Dasselbe gilt für die Unabhängigkeit eines zufälligen Elementes (bzw. einer zufälligen Funktion) X und einer Teil- σ -Algebra $\mathcal{G} \in \mathcal{A}$: dies ist der Fall, wenn $P(\{X \in A\} \cap G) = P(X \in A)P(G)$, für alle $A \in \mathcal{B}_1$, $G \in \mathcal{G}$ (bzw. $A \in \mathcal{B}_T$, $G \in \mathcal{G}$).

1.7 Prozesse mit unabhängigen Zuwächsen

In diesem Abschnitt wollen wir auf die Eigenschaften und Existenz der Prozesse mit unabhängigen Zuwächsen eingehen.

Sei $\{\varphi_{s,t}, s, t \geq 0\}$ eine Familie von charakteristischen Funktionen der Wahrscheinlichkeitsmaße $Q_{s,t}$, $s, t \geq 0$ auf $\mathcal{B}(\mathbb{R})$, d.h., für $z \in \mathbb{R}$, $s, t \geq 0$ gilt $\varphi_{s,t}(z) = \int_{\mathbb{R}} e^{izx} Q_{s,t}(dx)$.

Theorem 1.7.1

Es existiert ein stochastischer Prozess $X = \{X(t), t \geq 0\}$ mit unabhängigen Zuwächsen mit der Eigenschaft, dass für alle $s, t \geq 0$ die charakteristische Funktion von $X(t) - X(s)$ gleich $\varphi_{s,t}$ ist, genau dann, wenn

$$\varphi_{s,t} = \varphi_{s,u} \varphi_{u,t} \tag{1.7.1}$$

für alle $0 \leq s < u < t < \infty$. Dabei kann die Verteilung von $X(0)$ beliebig gewählt werden.

Beweis Die Notwendigkeit der Bedingung 1.7.1 ist klar, weil für alle $s \in (0, \infty) : s < u < t$ gilt: $X(t) - X(s) = \underbrace{X(t) - X(u)}_{Y_1} + \underbrace{X(u) - X(s)}_{Y_2}$ und $X(t) - X(u)$ und $X(u) - X(s)$ sind paarweise

unabhängig. Dann gilt $\varphi_{s,t} = \varphi_{Y_1+Y_2} = \varphi_{Y_1} \varphi_{Y_2} = \varphi_{s,u} \varphi_{u,t}$.

Nun beweisen wir die Suffizienz.

Falls die Existenz eines Prozesses X mit unabhängigen Zuwächsen und Eigenschaft $\varphi_{X(t)-X(s)} = \varphi_{s,t}$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) bereits bewiesen wäre, könnte man die charakteristischen Funktionen aller seiner endlich-dimensionalen Verteilungen wie folgt durch $\{\varphi_{s,t}\}$ angeben.

Sei $n \in \mathbb{N}$, $0 = t_0 < t_1 < \dots < t_n < \infty$ und $Y = (X(t_0), X(t_1) - X(t_0), \dots, X(t_n) - X(t_{n-1}))^\top$. Aus der Unabhängigkeit der Zuwächse folgt

$$\varphi_Y(\underbrace{z_0, z_1, \dots, z_n}_z) = \mathbb{E} e^{i\langle z, Y \rangle} = \varphi_{X(t_0)}(z_0) \varphi_{t_0, t_1}(z_1) \dots \varphi_{t_{n-1}, t_n}(z_n), \quad z \in \mathbb{R}^{n+1},$$

wobei die Verteilung von $X(t_0)$ ein beliebiges Wahrscheinlichkeitsmaß Q_0 auf $\mathcal{B}(\mathbb{R})$ ist. Für $X_{t_0, \dots, t_n} = (X(t_0), X(t_1), \dots, X(t_n))^\top$ gilt allerdings $X_{t_0, \dots, t_n} = AY$, wobei

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

Dann gilt $\varphi_{X_{t_0, \dots, t_n}}(z) = \varphi_{AY}(z) = \mathbb{E} e^{i\langle z, AY \rangle} = \mathbb{E} e^{i\langle A^\top z, Y \rangle} = \varphi_Y(A^\top z)$. Deshalb hat die endlich-dimensionale Verteilung von X_{t_0, \dots, t_n} die charakteristische Funktion $\varphi_{X_{t_0, \dots, t_n}}(z) =$

$\varphi_{Q_0}(l_0)\varphi_{t_0,t_1}(l_1)\dots\varphi_{t_{n-1},t_n}(l_n)$, wobei $l = (l_1, l_1, \dots, l_n)^\top = A^\top z$, also

$$\begin{cases} l_0 &= z_0 + \dots + z_n \\ l_1 &= z_1 + \dots + z_n \\ &\vdots \\ l_n &= z_n \end{cases}$$

Dabei gilt $\varphi_{X(t_0)} = \varphi_{Q_0}$ und $\varphi_{X_{t_1, \dots, t_n}}(z_1, \dots, z_n) = \varphi_{X_{t_0, \dots, t_n}}(0, z_1, \dots, z_n)$ für alle $z_i \in \mathbb{R}$.
Nun beweisen wir die Existenz eines solchen Prozesses X .

Dabei konstruieren wir die Familie der charakteristischen Funktionen

$$\{\varphi_{t_0}, \varphi_{t_0, t_1, \dots, t_n}, \varphi_{t_1, \dots, t_n}, \quad 0 = t_0 < t_1 < \dots < t_n < \infty, \quad n \in \mathbb{N}\}$$

aus φ_{Q_0} und $\{\varphi_{s,t}, 0 \leq s < t\}$ wie oben, also

$$\begin{aligned} \varphi_{t_0} &= \varphi_{Q_0}, \quad \varphi_{t_1, \dots, t_n}(0, z_1, \dots, z_n) = \varphi_{t_0, t_1, \dots, t_n}(0, z_1, \dots, z_n), \quad z_i \in \mathbb{R}, \\ \varphi_{t_0, \dots, t_n}(z) &= \varphi_{t_0}(z_1 + \dots + z_n)\varphi_{t_0, t_1}(z_1 + \dots + z_n)\dots\varphi_{t_{n-1}, t_n}(z_n). \end{aligned}$$

Nun sollten wir prüfen, dass die Wahrscheinlichkeitsmaße, denen diese charakteristische Funktionen entsprechen, die Bedingungen des Theorems 1.1.2 erfüllen. Dies werden wir in äquivalenter Form tun, denn nach Aufgabe ... des Übungsblattes ... sind die Bedingungen der Symmetrie und der Konsistenz im Theorem 1.1.2 äquivalent zu:

- $\varphi_{t_{i_0}, \dots, t_{i_n}}(z_{i_0}, \dots, z_{i_n}) = \varphi_{t_0, \dots, t_n}(z_0, \dots, z_n)$ für eine beliebige Permutation $(0, 1, \dots, n) \mapsto (i_0, i_1, \dots, i_n)$,
- $\varphi_{t_0, \dots, t_{m-1}, t_{m+1}, \dots, t_n}(z_0, \dots, z_{m-1}, z_{m+1}, \dots, z_n) = \varphi_{t_0, \dots, t_n}(z_0, \dots, 0, \dots, z_n)$, für alle $z_0, \dots, z_n \in \mathbb{R}, m \in \{1, \dots, n\}$.

Die erste Bedingung a) ist offensichtlich. Es gilt b), weil

$$\varphi_{t_{m-1}, t_m}(0 + z_{m+1} + \dots + z_n)\varphi_{t_m, t_{m+1}}(z_{m+1} + \dots + z_n) = \varphi_{t_{m-1}, t_{m+1}}(z_{m+1}, \dots, z_n)$$

für alle $m \in \{1, \dots, n\}$. Damit ist die Existenz von X bewiesen. \square

Beispiel 1.7.1 1. Falls $T = \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, dann hat $X = \{X(t), t \in \mathbb{N}_0\}$ unabhängige Zuwächse genau dann, wenn $X(n) \stackrel{d}{=} \sum_{i=0}^n Y_i$, wobei $\{Y_i\}$ unabhängige Zufallsvariablen sind und $Y_n \stackrel{d}{=} X(n) - X(n-1)$, $n \in \mathbb{N}$. Ein solcher Prozess X heißt *zufällige Irrfahrt*. Er kann auch für Y_i mit Werten in \mathbb{R}^m definiert werden.

2. Der Poisson-Prozess mit Intensität λ hat unabhängige Zuwächse, wie wir es später zeigen werden.

3. Der Wiener-Prozess besitzt unabhängige Zuwächse.

Aufgabe 1.7.1

Beweisen Sie es!

Aufgabe 1.7.2

Sei $X = \{X(t), t \geq 0\}$ ein Prozess mit unabhängigen Zuwächsen und $g : [0, \infty) \rightarrow \mathbb{R}$ eine beliebige (deterministische) Funktion. Zeigen Sie, dass der Prozess $Y = \{Y(t), t \geq 0\}$ mit $Y(t) = X(t) + g(t)$, $t \geq 0$, ebenso unabhängige Zuwächse besitzt.

1.8 Ergänzende Aufgaben

Aufgabe 1.8.1

Beweisen Sie folgende Behauptung: Die Familie von Wahrscheinlichkeitsmaßen $\mathbb{P}_{t_1, \dots, t_n}$ auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \geq 1$, $t = (t_1, \dots, t_n)^\top \in T^n$ erfüllt die Bedingungen des Theorems von Kolmogorov genau dann, wenn für alle $n \geq 2$ und für alle $s = (s_1, \dots, s_n)^\top \in \mathbb{R}^n$ die folgenden Bedingungen erfüllt sind:

- $\varphi_{\mathbb{P}_{t_1, \dots, t_n}}((s_1, \dots, s_n)^\top) = \varphi_{\mathbb{P}_{t_{\pi(1)}, \dots, t_{\pi(n)}}}((s_{\pi(1)}, \dots, s_{\pi(n)})^\top)$ für alle $\pi \in \mathcal{S}_n$.
- $\varphi_{\mathbb{P}_{t_1, \dots, t_{n-1}}}((s_1, \dots, s_{n-1})^\top) = \varphi_{\mathbb{P}_{t_1, \dots, t_n}}((s_1, \dots, s_{n-1}, 0)^\top)$.

Bemerkung: $\varphi(\cdot)$ bezeichnet die charakteristische Funktion des jeweiligen Maßes. \mathcal{S}_n bezeichnet die Gruppe aller Permutationen $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$.

Aufgabe 1.8.2

Zeigen Sie die Existenz einer zufälligen Funktion, deren endlich-dimensionale Verteilungen multivariat normalverteilt sind, und geben Sie die messbaren Räume $(E_{t_1, \dots, t_n}, \mathcal{E}_{t_1, \dots, t_n})$ explizit an.

Aufgabe 1.8.3

Geben Sie ein Beispiel für eine Familie von Wahrscheinlichkeitsmaßen $\mathbb{P}_{t_1, \dots, t_n}$, welche nicht die Bedingungen des Theorems von Kolmogorov erfüllt.

Aufgabe 1.8.4

Seien $X = \{X(t), t \in T\}$ und $Y = \{Y(t), t \in T\}$ zwei stochastische Prozesse, die auf dem selben vollständigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert sind und Werte in einem messbaren Raum (S, \mathcal{B}) annehmen.

- Beweisen Sie: X und Y sind stochastisch äquivalent $\implies \mathbb{P}_X = \mathbb{P}_Y$.
- Geben Sie ein Beispiel zweier Prozesse X und Y an, für die gilt: $\mathbb{P}_X = \mathbb{P}_Y$, aber X und Y sind nicht stochastisch äquivalent.
- Beweisen Sie: X und Y sind stochastisch ununterscheidbar $\implies X$ und Y sind stochastisch äquivalent.
- Beweisen Sie im Falle der Abzählbarkeit von T : X und Y sind stochastisch äquivalent $\implies X$ und Y sind stochastisch ununterscheidbar.
- Geben Sie im Falle der Überzählbarkeit von T ein Beispiel zweier Prozesse X und Y an, für die gilt: X und Y sind stochastisch äquivalent, aber nicht stochastisch ununterscheidbar.

Aufgabe 1.8.5

Sei $W = \{W(t), t \in \mathbb{R}\}$ ein Wiener-Prozess. Welche der folgenden Prozesse sind ebenfalls Wiener-Prozesse?

- $W_1 = \{W_1(t) := -W(t), t \in \mathbb{R}\}$,
- $W_2 = \{W_2(t) := \sqrt{t}W(1), t \in \mathbb{R}\}$,
- $W_3 = \{W_3(t) := W(2t) - W(t), t \in \mathbb{R}\}$.

Aufgabe 1.8.6

Es sei der stochastische Prozess $X = \{X(t), t \in [0, 1]\}$ gegeben, welcher aus identischen und unabhängig verteilten Zufallsvariablen mit einer Dichte $f(x)$, $x \in \mathbb{R}$, besteht. Zeigen Sie, dass ein solcher Prozess nicht stochastisch stetig in $t \in [0, 1]$ sein kann.

Aufgabe 1.8.7

Geben Sie ein Beispiel eines stochastischen Prozesses $X = \{X(t), t \in T\}$ an, welcher stochastisch stetig auf T ist, aber nicht fast sicher stetig auf T , und beweisen Sie, warum dies so ist.

Aufgabe 1.8.8

Im Zusammenhang mit der Stetigkeit von stochastischen Prozessen spielt das sogenannte *Kriterium von Kolmogorov* eine zentrale Rolle (siehe auch Satz 1.3.1 im Skript): Sei $X = \{X(t), t \in [a, b]\}$ ein reellwertiger stochastischer Prozess. Falls Konstanten $\alpha, \varepsilon > 0$ und $C := C(\alpha, \varepsilon) > 0$ existieren, so dass

$$\mathbb{E}|X(t+h) - X(t)|^\alpha \leq C|h|^{1+\varepsilon} \quad (1.8.1)$$

für ausreichend kleines h , dann besitzt der Prozess X eine stetige Modifikation. Zeigen Sie:

- Falls man in Bedingung (1.8.1) die Variable $\varepsilon = 0$ fixiert, dann reicht diese Bedingung im Allgemeinen nicht zur Existenz einer stetigen Modifikation aus. *Tipp: Betrachten Sie den Poisson-Prozess.*
- Der Wiener-Prozess $W = \{W(t), t \in [0, \infty)\}$ besitzt eine stetige Modifikation. *Tipp: Betrachten Sie den Fall $\alpha = 4$.*

Aufgabe 1.8.9

Zeigen Sie, dass der Wiener-Prozesses W an keiner Stelle $t \in [0, \infty)$ stochastisch differenzierbar ist.

Aufgabe 1.8.10

Zeigen Sie, dass die Kovarianzfunktion $C(s, t)$ eines komplexwertigen stochastischen Prozesses $X = \{X(t), t \in T\}$

- symmetrisch ist, d.h. $C(s, t) = \overline{C(t, s)}$, $s, t \in T$,
- die Identität $C(t, t) = \text{var } X(t)$, $t \in T$, erfüllt,
- positiv semidefinit* ist, d.h. für alle $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $z_1, \dots, z_n \in \mathbb{C}$ gilt:

$$\sum_{i=1}^n \sum_{j=1}^n C(t_i, t_j) z_i \bar{z}_j \geq 0.$$

Aufgabe 1.8.11

Zeigen Sie, dass es eine zufällige Funktion $X = \{X(t), t \in T\}$ gibt, die gleichzeitig folgende Bedingungen erfüllt:

- Das zweite Moment $\mathbb{E}X^2$ existiert nicht.
- Das Variogramm $\gamma(s, t)$ ist endlich für alle $s, t \in T$.

Aufgabe 1.8.12

Geben Sie ein Beispiel für einen stochastischen Prozess $X = \{X(t), t \in T\}$ an, dessen Pfade gleichzeitig L^2 -differenzierbar, aber nicht fast sicher differenzierbar sind, und beweisen Sie, warum dies so ist.

Aufgabe 1.8.13

Geben Sie ein Beispiel für einen stochastischen Prozess $X = \{X(t), t \in T\}$ an, dessen Pfade gleichzeitig fast sicher differenzierbar, aber nicht L^1 -differenzierbar sind, und beweisen Sie, warum dies so ist.

Aufgabe 1.8.14

Beweisen Sie, dass der Wiener-Prozess unabhängige Zuwächse besitzt.

Aufgabe 1.8.15

Beweisen Sie: Ein (reellwertiger) stochastischer Prozess $X = \{X(t), t \in [0, \infty)\}$ mit unabhängigen Zuwächsen hat bereits dann stationäre Zuwächse, wenn die Verteilung der Zufallsvariablen $X(t+h) - X(h)$ unabhängig von h ist.

2 Zählprozesse

Hier werden einige Beispiele von stochastischen Prozessen betrachtet, die das Zählen von Ereignissen modellieren und daher stückweise konstante Pfade besitzen.

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und sei $\{S_n\}_{n \in \mathbb{N}}$ eine nichtfallende Folge von f.s. nicht-negativen Zufallsvariablen, d.h. $0 \leq S_1 \leq S_2 \leq \dots \leq S_n \leq \dots$

Definition 2.0.1

Der stochastische Prozess $N = \{N(t), t \geq 0\}$ wird *Zählprozess* genannt, falls

$$N(t) = \sum_{n=1}^{\infty} \mathbf{1}(S_n \leq t),$$

wobei $\mathbf{1}(A)$ die Indikatorfunktion eines Ereignisses $A \in \mathcal{A}$ ist.

$N(t)$ zählt die Ereignisse, die zu Zeitpunkten S_n bis zur Zeit t eintreten. S_n können z.B. Zeitpunkte des Eintretens

1. des n -ten Elementarteilchens im Geigerzähler sein, oder
2. eines Schadens in der Sachschadenversicherung, oder
3. eines Datenpakets beim Server in einem Computernetzwerk, usw.

Einen Spezialfall der Zählprozesse bilden die sog. *Erneuerungsprozesse*.

2.1 Erneuerungsprozesse

Definition 2.1.1

Sei $\{T_n\}_{n \in \mathbb{N}}$ eine Folge von u.i.v. nicht-negativen Zufallsvariablen mit $P(T_1 > 0) > 0$. Ein Zählprozess $N = \{N(t), t \geq 0\}$ mit $N(0) = 0$ f.s., $S_n = \sum_{k=1}^n T_k$, $n \in \mathbb{N}$, wird *Erneuerungsprozess* genannt. Dabei heißt S_n der n -te *Erneuerungszeitpunkt*, $n \in \mathbb{N}$.

Den Namen „Erneuerungsprozess“ leitet man von folgender Interpretation ab. Die „Zwischenankunftszeiten“ T_n werden als Lebensdauer eines technischen Ersatzteils bzw. Mechanismus in einem System interpretiert, somit sind S_n die Zeitpunkte des n -ten Versagens des Systems. Das defekte Teil wird sofort durch ein neues baugleiches Teil ersetzt (wie z.B. beim Auswechseln einer kaputten Glühbirne). Somit ist $N(t)$ die Anzahl der Reparaturen (die sog. „Erneuerungen“) des Systems bis zur Zeit t .

Bemerkung 2.1.1 1. Man setzt $N(t) = \infty$, falls $S_n \leq t$ für alle $n \in \mathbb{N}$.

2. Oft wird vorausgesetzt, dass nur T_2, T_3, \dots identisch verteilt sind mit $ET_n < \infty$. Die Verteilung von T_1 ist dann beliebig wählbar. Ein solcher Prozess $N = \{N(t), t \geq 0\}$ wird *verzögerter Erneuerungsprozess* (mit Verzögerung T_1) genannt.
3. Manchmal wird die Forderung $T_n \geq 0$ weggelassen.

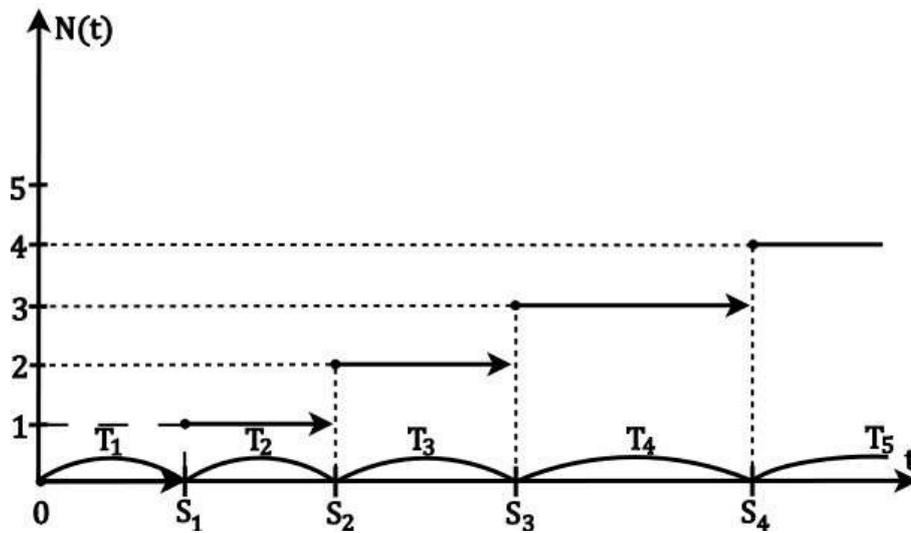


Abbildung 2.1: Konstruktion und Trajektorien eines Erneuerungsprozesses

4. Es ist klar, dass $\{S_n\}_{n \in \mathbb{N}_0}$ mit $S_0 = 0$ f.s., $S_n = \sum_{k=1}^n T_k$, $n \in \mathbb{N}$ eine *zufällige Irrfahrt* ist.
5. Wenn man voraussetzt, dass das n -te Auswechseln des defekten Teils im System eine Zeit T'_n dauert, so wird durch $\tilde{T}_n = T_n + T'_n$, $n \in \mathbb{N}$ ein anderer Erneuerungsprozess gegeben, der von seiner stochastischen Beschaffenheit sich nicht von dem in der Definition 2.1.1 gegebenen Prozess unterscheidet.

Im weiteren Verlauf der Vorlesung wird vorausgesetzt, dass $\mu = ET_n \in (0, \infty)$, $n \in \mathbb{N}$.

Theorem 2.1.1 (Individueller Ergodensatz):

Sei $N = \{N(t), t \geq 0\}$ ein Erneuerungsprozess. Dann gilt:

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad \text{f.s.}$$

Beweis Für alle $t \geq 0$ und $n \in \mathbb{N}$ gilt $\{N(t) = n\} = \{S_n \leq t < S_{n+1}\}$, deshalb $S_{N(t)} \leq t < S_{N(t)+1}$ und

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}.$$

Wenn wir zeigen könnten, dass $\frac{S_{N(t)}}{N(t)} \xrightarrow[t \rightarrow \infty]{f.s.} \mu$ und $N(t) \xrightarrow[t \rightarrow \infty]{f.s.} \infty$, dann gilt $\frac{t}{N(t)} \xrightarrow[t \rightarrow \infty]{f.s.} \mu$ und deshalb gilt die Aussage des Theorems.

Nach dem Starken Gesetz der Großen Zahlen von Kolmogorov (vgl. Skript „Wahrscheinlichkeitsrechnung“ (WR), Satz 7.4) gilt $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \mu$, also $S_n \xrightarrow[n \rightarrow \infty]{f.s.} \infty$ und daher $P(N(t) < \infty) = 1$, weil $P(N(t) = \infty) = P(S_n \leq t, \forall n) = 1 - \underbrace{P(\exists n : \forall m \in \mathbb{N}_0 S_{n+m} > t)}_{=1, \text{ falls } S_n \xrightarrow[n \rightarrow \infty]{f.s.} \infty} = 1 - 1 = 0$. Dann ist

$N(t), t \geq 0$, eine echte Zufallsvariable.

Zeigen wir, dass $N(t) \xrightarrow[t \rightarrow \infty]{f.s.} \infty$. Alle Trajektorien von $N(t)$ sind monoton nichtfallend in $t \geq 0$,

also $\exists \lim_{t \rightarrow \infty} N(\omega, t)$ für alle $\omega \in \Omega$. Außerdem gilt

$$\begin{aligned} \mathbb{P}(\lim_{t \rightarrow \infty} N(t) < \infty) &= \lim_{n \rightarrow \infty} \mathbb{P}(\lim_{t \rightarrow \infty} N(t) < n) \stackrel{(*)}{=} \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}(N(t) < n) \\ &= \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}(S_n > t) = \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \mathbb{P}\left(\sum_{k=1}^n T_k > t\right) \\ &\leq \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} \underbrace{\sum_{k=1}^n \mathbb{P}(T_k > \frac{t}{n})}_{\xrightarrow{t \rightarrow \infty} 0} = 0. \end{aligned}$$

Der Übergang (*) gilt, weil $\{\lim_{t \rightarrow \infty} N(t) < n\} = \{\exists t_0 \in \mathbb{Q}_+ : \forall t \geq t_0 \ N(t) < n\} = \cup_{t_0 \in \mathbb{Q}_+} \cap_{t \geq t_0} \{N(t) < n\} = \liminf_{t \rightarrow \infty} \{N(t) < n\}$, und dann benutzt man die Stetigkeit des Wahrscheinlichkeitsmaßes, wobei $\mathbb{Q}_+ = \mathbb{Q} \cap \mathbb{R}_+ = \{q \in \mathbb{Q} : q \geq 0\}$. Da für jedes $\omega \in \Omega$ gilt $\lim_{n \rightarrow \infty} \frac{S_n}{n} = \lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)}$ (der Wertebereich einer Realisierung von $N(\cdot)$ ist ja eine Teilfolge von \mathbb{N}), gilt $\lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)} \stackrel{f.s.}{=} \mu$. \square

Bemerkung 2.1.2

Der Ergodensatz lässt sich verallgemeinern auf den Fall von nicht identisch verteilten T_n . Dabei wird gefordert, dass $\mu_n = \mathbb{E}T_n$, $\{T_n - \mu_n\}_{n \in \mathbb{N}}$ gleichmäßig integrierbar sind und $\frac{1}{n} \sum_{k=1}^n \mu_k \xrightarrow{n \rightarrow \infty} \mu > 0$. Dann kann bewiesen werden, dass $\frac{N(t)}{t} \xrightarrow[t \rightarrow \infty]{\mathbb{P}} \frac{1}{\mu}$ (vgl. [2], S. 276).

Theorem 2.1.2 (Zentraler Grenzwertsatz):

Falls $\mu \in (0, \infty)$, $\sigma^2 = \text{var } T_1 \in (0, \infty)$, dann gilt

$$\mu^{\frac{3}{2}} \cdot \frac{N(t) - \frac{t}{\mu}}{\sigma \sqrt{t}} \xrightarrow[t \rightarrow \infty]{d} Y,$$

wobei $Y \sim \mathcal{N}(0, 1)$.

Beweis Nach dem zentralen Grenzwertsatz für Summen von u.i.v. Zufallsvariablen (vgl. Satz 7.5, WR) gilt

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n \rightarrow \infty]{d} Y. \quad (2.1.1)$$

Sei $[x]$ der ganze Teil von $x \in \mathbb{R}$. Es gilt für $a = \frac{\sigma^2}{\mu^3}$, dass

$$\mathbb{P}\left(\frac{N(t) - \frac{t}{\mu}}{\sqrt{at}} \leq x\right) = \mathbb{P}\left(N(t) \leq x\sqrt{at} + \frac{t}{\mu}\right) = \mathbb{P}\left(S_{m(t)} > t\right),$$

wobei $m(t) = \left\lceil x\sqrt{at} + \frac{t}{\mu} \right\rceil + 1$, $t \geq 0$, und $\lim_{t \rightarrow \infty} m(t) = \infty$. Deshalb folgt, dass

$$\begin{aligned} \left| \mathbb{P}\left(\frac{N(t) - \frac{t}{\mu}}{\sqrt{at}} \leq x\right) - \varphi(x) \right| &= \left| \mathbb{P}\left(S_{m(t)} > t\right) - \varphi(x) \right| \\ &= \left| \mathbb{P}\left(\frac{S_{m(t)} - \mu m(t)}{\sigma \sqrt{m(t)}} > \frac{t - \mu m(t)}{\sigma \sqrt{m(t)}}\right) - \varphi(x) \right| := I_t(x) \end{aligned}$$

für beliebiges $t \geq 0$ und $x \in \mathbb{R}$, wobei φ die Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung ist. Für festes $x \in \mathbb{R}$ führen wir $Z_t = -\frac{t - \mu m(t)}{\sigma \sqrt{m(t)}} - x$, $t \geq 0$, ein. Es gilt dann

$$I_t(x) = \left| \mathbb{P} \left(\frac{S_{m(t)} - \mu m(t)}{\sigma \sqrt{m(t)}} + Z_t > -x \right) - \varphi(x) \right|.$$

Wenn wir zeigen könnten, dass $Z_t \xrightarrow[t \rightarrow \infty]{} 0$, dann würde nach (2.1.1) und dem Satz von Slutsky (Satz 6.4.1, WR) folgen, dass $\frac{S_{m(t)} - \mu m(t)}{\sigma \sqrt{m(t)}} + Z_t \xrightarrow[t \rightarrow \infty]{d} Y \sim \mathcal{N}(0, 1)$, denn aus $Z_t \xrightarrow[t \rightarrow \infty]{} 0$ f.s. folgt $Z_t \xrightarrow[t \rightarrow \infty]{d} 0$. Deshalb könnte man schreiben $I_t(x) \xrightarrow[t \rightarrow \infty]{} |\bar{\varphi}(-x) - \varphi(x)| = |\varphi(x) - \varphi(x)| = 0$, wobei $\bar{\varphi}(x) = 1 - \varphi(x)$ die Tail-Funktion der $\mathcal{N}(0, 1)$ -Verteilung ist, und man hier die Symmetrie-Eigenschaft von $\mathcal{N}(0, 1)$: $\bar{\varphi}(-x) = \varphi(x)$, $x \in \mathbb{R}$ benutzt hat.

Zeigen wir nun, dass $Z_t \xrightarrow[t \rightarrow \infty]{} 0$, also $\frac{t - \mu m(t)}{\sigma \sqrt{m(t)}} \xrightarrow[t \rightarrow \infty]{} -x$. Es gilt $m(t) = x\sqrt{at} + \frac{t}{\mu} + \varepsilon(t)$, wobei $\varepsilon(t) \in [0, 1)$. Dann gilt

$$\begin{aligned} \frac{t - \mu m(t)}{\sigma \sqrt{m(t)}} &= \frac{t - \mu x\sqrt{at} - t - \mu\varepsilon(t)}{\sigma \sqrt{m(t)}} = -x \frac{\sqrt{at} - \mu}{\sigma \sqrt{x\sqrt{at} + \frac{t}{\mu} + \varepsilon(t)}} - \frac{\mu\varepsilon(t)}{\sigma \sqrt{m(t)}} \\ &= -\frac{x\mu}{\sigma \sqrt{\frac{x}{\sqrt{at}} + \frac{1}{\mu a} + \frac{\varepsilon(t)}{at}}} - \frac{\mu - \varepsilon(t)}{\sigma \sqrt{m(t)}} \\ &= \underbrace{-\frac{x\frac{\mu}{\sigma}}{\sqrt{\frac{\mu^2}{\sigma^2} + \frac{x}{\sqrt{at}} + \frac{\varepsilon(t)}{at}}}}_{\xrightarrow[t \rightarrow \infty]{} -x} - \underbrace{\frac{\mu\varepsilon(t)}{\sigma \sqrt{m(t)}}}_{\xrightarrow[t \rightarrow \infty]{} 0} \xrightarrow[t \rightarrow \infty]{} -x. \end{aligned}$$

□

Bemerkung 2.1.3

Der zentrale Grenzwertsatz läßt sich in Lindeberg-Form auch für nicht identisch verteilte T_n beweisen, vgl. [2], S. 276 - 277.

Definition 2.1.2

Die Funktion $H(t) = \mathbb{E}N(t)$, $t \geq 0$ heißt *Erneuerungsfunktion* des Prozesses N (oder der Folge $\{S_n\}_{n \in \mathbb{N}}$).

Sei $F_T(x) = \mathbb{P}(T_1 \leq x)$, $x \in \mathbb{R}$ die Verteilungsfunktion von T_1 . Für beliebige Verteilungsfunktionen $F, G: \mathbb{R} \rightarrow [0, 1]$ sei die *Faltung* $F * G$ definiert als $F * G(x) = \int_{-\infty}^x F(x-y)dG(y)$. Die *k-fache Faltung* F^{*k} der Verteilungsfunktion F mit sich selbst, $k \in \mathbb{N}_0$, wird induktiv definiert:

$$\begin{aligned} F^{*0}(x) &= 1(x \in [0, \infty)), \quad x \in \mathbb{R}, \\ F^{*1}(x) &= F(x), \quad x \in \mathbb{R}, \\ F^{*(k+1)}(x) &= F^{*k} * F(x), \quad x \in \mathbb{R}. \end{aligned}$$

Lemma 2.1.1

Die Erneuerungsfunktion H eines Erneuerungsprozesses N ist monoton nichtfallend und rechtsseitig stetig auf \mathbb{R}_+ . Außerdem gilt

$$H(t) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \leq t) = \sum_{n=1}^{\infty} F_T^{*n}(t), \quad t \geq 0. \quad (2.1.2)$$

Beweis Die Monotonie und rechtsseitige Stetigkeit von H folgt aus der fast sicheren Monotonie und rechtsseitigen Stetigkeit der Trajektorien von N . Nun beweisen wir (2.1.2):

$$H(t) = \mathbb{E}N(t) = \mathbb{E} \sum_{n=1}^{\infty} \mathbf{1}(S_n \leq t) \stackrel{(*)}{=} \sum_{n=1}^{\infty} \mathbb{E} \mathbf{1}(S_n \leq t) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \leq t) = \sum_{n=1}^{\infty} F_T^{*n}(t),$$

weil $\mathbb{P}(S_n \leq t) = \mathbb{P}(T_1 + \dots + T_n \leq t) = F_T^{*n}(t)$, $t \geq 0$. Die Gleichung (*) gilt für alle partiellen Summen auf beiden Seiten, also auch im Grenzwert. \square

Bis auf Ausnahmefälle ist es unmöglich, die Erneuerungsfunktion H durch die Formel (2.1.2) analytisch zu berechnen. Deshalb benutzt man oft in Berechnungen die *Laplace-Transformierte* von H .

Für eine monotone (z.B. monoton nichtfallende) rechtsseitig stetige Funktion $G : [0, \infty) \rightarrow \mathbb{R}$ ist ihre *Laplace-Transformierte* definiert als $\hat{l}_G(s) = \int_0^{\infty} e^{-sx} dG(x)$, $s \geq 0$. Hier ist das Integral als Lebesgue-Stieltjes-Integral zu verstehen, also als ein Lebesgue-Integral bzgl. des Maßes μ_G auf $\mathcal{B}_{\mathbb{R}_+}$ definiert durch $\mu_G((x, y]) = G(y) - G(x)$, $0 \leq x < y < \infty$, falls G monoton nichtfallend ist.

Zur Erinnerung, für eine Zufallsvariable $X \geq 0$ ist ihre Laplace-Transformierte \hat{l}_X definiert durch $\hat{l}_X(s) = \int_0^{\infty} e^{-sx} dF_X(x)$, $s \geq 0$.

Lemma 2.1.2

Für $s > 0$ gilt:

$$\hat{l}_H(s) = \frac{\hat{l}_{T_1}(s)}{1 - \hat{l}_{T_1}(s)}.$$

Beweis Es gilt:

$$\begin{aligned} \hat{l}_H(s) &= \int_0^{\infty} e^{-sx} dH(x) \stackrel{(2.1.2)}{=} \int_0^{\infty} e^{-sx} d \left(\sum_{n=1}^{\infty} F_T^{*n}(x) \right) = \sum_{n=1}^{\infty} \int_0^{\infty} e^{-sx} dF_T^{*n}(x) \\ &= \sum_{n=1}^{\infty} \hat{l}_{T_1 + \dots + T_n}(s) = \sum_{n=1}^{\infty} \left(\hat{l}_{T_1}(s) \right)^n = \frac{\hat{l}_{T_1}(s)}{1 - \hat{l}_{T_1}(s)}, \end{aligned}$$

wobei für $s > 0$ gilt $\hat{l}_{T_1}(s) < 1$ und somit konvergiert die geometrische Reihe $\sum_{n=1}^{\infty} \left(\hat{l}_{T_1}(s) \right)^n$. \square

Bemerkung 2.1.4

Falls $N = \{N(t), t \geq 0\}$ ein verzögerter Erneuerungsprozess (mit Verzögerung T_1) ist, dann gelten die Aussagen der Lemmas 2.1.1 - 2.1.2 in folgender Form:

1.

$$H(t) = \sum_{n=0}^{\infty} (F_{T_1} * F_{T_2}^{*n})(t), \quad t \geq 0,$$

wobei F_{T_1} bzw. F_{T_2} die Verteilungsfunktionen von T_1 bzw. T_n , $n \geq 2$ sind.

2.

$$\hat{l}_H(s) = \frac{\hat{l}_{T_1}(s)}{1 - \hat{l}_{T_2}(s)}, \quad s \geq 0, \quad (2.1.3)$$

wobei \hat{l}_{T_1} und \hat{l}_{T_2} die Laplace-Transformierten der Verteilung von T_1 bzw. T_n , $n \geq 2$ sind.

Für weitere Betrachtungen brauchen wir einen Satz (von Wald) über den Erwartungswert einer Summe (in zufälliger Anzahl) von unabhängigen Zufallsvariablen.

Definition 2.1.3

Sei ν eine \mathbb{N} -wertige Zufallsvariable und sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen, definiert auf demselben Wahrscheinlichkeitsraum. ν heißt *unabhängig von der Zukunft*, falls für alle $n \in \mathbb{N}$ das Ereignis $\{\nu \leq n\}$ nicht von der σ -Algebra $\sigma(\{X_k, k > n\})$ abhängt.

Theorem 2.1.3 (Waldsche Identität):

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit $\sup E|X_n| < \infty$, $EX_n = a$, $n \in \mathbb{N}$ und sei ν eine \mathbb{N} -wertige Zufallsvariable, die von der Zukunft unabhängig ist, mit $E\nu < \infty$. Dann gilt

$$E\left(\sum_{n=1}^{\nu} X_n\right) = a \cdot E\nu.$$

Beweis Berechne $S_n = \sum_{k=1}^n X_k$, $n \in \mathbb{N}$. Da $E\nu = \sum_{n=1}^{\infty} P(\nu \geq n)$, so folgt die Aussage aus dem Lemma 2.1.3. \square

Lemma 2.1.3 (Kolmogorov-Prokhorov):

Sei ν eine \mathbb{N} -wertige Zufallsvariable, die nicht von der Zukunft abhängt, und es gelte

$$\sum_{n=1}^{\infty} P(\nu \geq n)E|X_n| < \infty. \quad (2.1.4)$$

Dann gilt $ES_{\nu} = \sum_{n=1}^{\infty} P(\nu \geq n)EX_n$. Falls $X_n \geq 0$ f.s., dann braucht man die Bedingung (2.1.4) nicht.

Beweis Es gilt $S_{\nu} = \sum_{n=1}^{\nu} X_n = \sum_{n=1}^{\infty} X_n \mathbf{1}(\nu \geq n)$. Führen wir die Bezeichnung $S_{\nu,n} = \sum_{k=1}^n X_k \mathbf{1}(\nu \geq k)$, $n \in \mathbb{N}$, ein. Beweisen wir das Lemma zunächst für $X_n \geq 0$ f.s., $n \in \mathbb{N}$. Es gilt $S_{\nu,n} \uparrow S_{\nu}$, $n \rightarrow \infty$ für jedes $\omega \in \Omega$, und so gilt nach dem Satz über die monotone Konvergenz: $ES_{\nu} = \lim_{n \rightarrow \infty} ES_{\nu,n} = \lim \sum_{k=1}^n E(X_k \mathbf{1}(\nu \geq k))$. Da allerdings $\{\nu \geq k\} = \{\nu \leq k-1\}^c$ nicht von $\sigma(X_k) \subset \sigma(\{X_n, n \geq k\})$ abhängt, gilt $E(X_k \mathbf{1}(\nu \geq k)) = EX_k P(\nu \geq k)$, $k \in \mathbb{N}$, und daher $ES_{\nu} = \sum_{n=1}^{\infty} P(\nu \geq n)EX_n$.

Sei nun X_n beliebig. Setze $Y_n = |X_n|$, $Z_n = \sum_{k=1}^n Y_k$, $Z_{\nu,n} = \sum_{k=1}^n Y_k \mathbf{1}(\nu \geq k)$, $n \in \mathbb{N}$. Da $Y_n \geq 0$, $n \in \mathbb{N}$, gilt $EZ_{\nu} = \sum_{n=1}^{\infty} E(X_n | P(\nu \geq k)) < \infty$ aus (2.1.4). Da allerdings $|S_{\nu,n}| \leq Z_{\nu,n} \leq Z_{\nu}$, $n \in \mathbb{N}$, dann gilt nach dem Satz von Lebesgue über die dominierte Konvergenz, dass $ES_{\nu} = \lim_{n \rightarrow \infty} ES_{\nu,n} = \sum_{n=1}^{\infty} EX_n P(\nu \geq n)$, wobei diese Reihe absolut konvergiert. \square

Folgerung 2.1.1 1. Für eine beliebige Borel-messbare Funktion $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ und den Erneuerungsprozess $N = \{N(t), t \geq 0\}$ mit Zwischenankunftszeiten $\{T_n\}$, T_n u.i.v., $\mu = ET_n \in (0, \infty)$ gilt

$$E\left(\sum_{k=1}^{N(t)+1} g(T_k)\right) = (1 + H(t))Eg(T_1), \quad t \geq 0.$$

2. $H(t) < \infty$, $t \geq 0$.

Beweis 1. Für jedes $t \geq 0$ hängt $\nu = 1 + H(t)$ offensichtlich nicht von der Zukunft von $\{T_n\}_{n \in \mathbb{N}}$ ab, und der Rest folgt aus dem Theorem 2.1.3 mit $X_n = g(T_n)$, $n \in \mathbb{N}$.

2. Für $s > 0$ betrachte $T_n^{(s)} = \min\{T_n, s\}$, $n \in \mathbb{N}$. Wähle $s > 0$ so, dass für beliebig gewähltes (aber festes) $\varepsilon > 0$: $\mu^{(s)} = \mathbb{E}T_1^{(s)} \geq \mu - \varepsilon > 0$. Sei $N^{(s)}$ der Erneuerungsprozess, der auf der Folge $\{T_n^{(s)}\}_{n \in \mathbb{N}}$ von Zwischenankunftszeiten aufgebaut wird: $N^{(s)}(t) = \sum_{n=1}^{\infty} \mathbf{1}(T_n^{(s)} \leq t)$, $t \geq 0$. Es gilt $N(t) \leq N^{(s)}(t)$, $t \geq 0$, f.s., dabei nach der Folgerung 2.1.1:

$$(\mu - \varepsilon)(\mathbb{E}N^{(s)}(t) + 1) \leq \mu^{(s)}(\mathbb{E}N^{(s)}(t) + 1) = \mathbb{E}S_{N^{(s)}(t)+1}^{(s)} = \mathbb{E}\underbrace{(S_{N^{(s)}(t)}^{(s)})}_{\leq t} + \underbrace{T_{N^{(s)}(t)+1}^{(s)}}_{\leq s} \leq t + s,$$

$t \geq 0$, wobei $S_n^{(s)} = T_1^{(s)} + \dots + T_n^{(s)}$, $n \in \mathbb{N}$. Daher $H(t) = \mathbb{E}N(t) \leq \mathbb{E}N^{(s)}(t) \leq \frac{t+s}{\mu-\varepsilon}$, $t \geq 0$. Da $\varepsilon > 0$ beliebig ist, gilt $\limsup_{t \rightarrow \infty} \frac{H(t)}{t} \leq \frac{1}{\mu}$, und unsere Aussage $H(t) < \infty$, $t \geq 0$. □

Folgerung 2.1.2 (Elementarer Erneuerungssatz):

Für einen Erneuerungsprozess N wie in Folgerung 2.1.1, 1) gilt:

$$\lim_{t \rightarrow \infty} \frac{H(t)}{t} = \frac{1}{\mu}.$$

Beweis In der Folgerung 2.1.1, Teil 2) ist bereits bewiesen worden, dass $\limsup_{t \rightarrow \infty} \frac{H(t)}{t} \leq \frac{1}{\mu}$. Zeigen wir, dass $\liminf_{t \rightarrow \infty} \frac{H(t)}{t} \geq \frac{1}{\mu}$, dann ist unsere Aussage bewiesen. Nach Theorem 2.1.1 gilt $\frac{N(t)}{t} \xrightarrow[t \rightarrow \infty]{} \frac{1}{\mu}$ f.s., daher nach Fatou's Lemma

$$\frac{1}{\mu} = \mathbb{E} \liminf_{t \rightarrow \infty} \frac{N(t)}{t} \leq \liminf_{t \rightarrow \infty} \frac{\mathbb{E}N(t)}{t} = \liminf_{t \rightarrow \infty} \frac{H(t)}{t}.$$

□

Bemerkung 2.1.5 1. Man kann auch zeigen, dass es sich im Falle des endlichen 2. Momentes von T_n ($\mu_2 = \mathbb{E}T_1^2 < \infty$) eine genauere Asymptotik für $H(t)$, $t \rightarrow \infty$ herleiten läßt:

$$H(t) = \frac{t}{\mu} + \frac{\mu_2}{2\mu^2} + o(1), \quad t \rightarrow \infty.$$

2. Der elementare Erneuerungssatz gilt auch für verzögerte Erneuerungsprozesse, wobei $\mu = \mathbb{E}T_2$. Definieren wir das *Erneuerungsmaß* H auf $\mathcal{B}(\mathbb{R}_+)$ durch $H(B) = \sum_{n=1}^{\infty} \int_B dF_T^{*n}(x)$, $B \in \mathcal{B}(\mathbb{R}_+)$. Es gilt $H((-\infty, t]) = H(t)$, $H((s, t]) = H(t) - H(s)$, $s, t \geq 0$, wenn man durch H sowohl die Erneuerungsfunktion als auch das Erneuerungsmaß bezeichnet.

Theorem 2.1.4 (Hauptsatz der Erneuerungstheorie):

Sei $N = \{N(t), t \geq 0\}$ ein (verzögerter) Erneuerungsprozess assoziiert mit der Folge $\{T_n\}_{n \in \mathbb{N}}$, wobei T_n , $n \in \mathbb{N}$ unabhängig sind, $\{T_n, n \geq 2\}$ identisch verteilt, und die Verteilung von T_2 nicht arithmetisch ist, also nicht auf einem regelmäßigen Gitter mit Wahrscheinlichkeit 1 konzentriert ist. Die Verteilung von T_1 sei beliebig. Sei $\mathbb{E}T_2 = \mu \in (0, \infty)$. Dann gilt

$$\int_0^t g(t-x) dH(x) \xrightarrow[t \rightarrow \infty]{} \frac{1}{\mu} \int_0^{\infty} g(x) dx,$$

wobei $g: \mathbb{R}_+ \rightarrow \mathbb{R}$ Riemann-integrierbar auf $[0, n]$ ist, für alle $n \in \mathbb{N}$, und $\sum_{n=0}^{\infty} \max_{n \leq x \leq n+1} |g(x)| < \infty$.

Ohne Beweis.

Insbesondere gilt $H((t-u, t]) \xrightarrow[t \rightarrow \infty]{} \frac{u}{\mu}$, für ein beliebiges $u \in \mathbb{R}_+$, also verhält sich H asymptotisch (für $t \rightarrow \infty$) wie das Lebesgue-Maß.

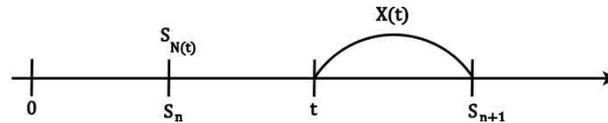


Abbildung 2.2:

Definition 2.1.4

Die Zufallsvariable $\chi(t) = S_{N(t)+1} - t$ heißt *Exzess* von N zum Zeitpunkt $t \geq 0$.

Es gilt offensichtlich $\chi(0) = T_1$. Geben wir nun ein Beispiel eines Erneuerungsprozesses mit stationären Zuwächsen.

Sei $N = \{N(t), t \geq 0\}$ ein verzögerter Erneuerungsprozess assoziiert mit Folge von Zwischenankunftszeiten $\{T_n\}_{n \in \mathbb{N}}$. Sei F_{T_1} bzw. F_{T_2} die Verteilungsfunktion der Verzögerung T_1 bzw. von $T_n, n \geq 2$. Wir nehmen an, dass $\mu = \mathbb{E}T_2 \in (0, \infty), F_{T_2}(0) = 0$, also $T_2 > 0$ f.s. und

$$F_{T_1}(x) = \frac{1}{\mu} \int_0^x \bar{F}_{T_2}(y) dy, \quad x \geq 0. \tag{2.1.5}$$

In diesem Fall sagt man, dass F_{T_1} die *integrierte Tailverteilungsfunktion* von T_2 ist.

Theorem 2.1.5

Unter den obigen Voraussetzungen ist N ein Prozess mit stationären Zuwächsen.

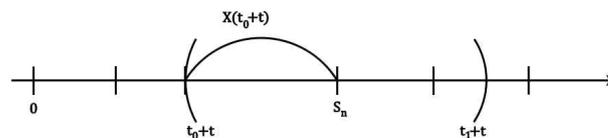


Abbildung 2.3:

Beweis Sei $n \in \mathbb{N}, 0 \leq t_0 < t_1 < \dots < t_n < \infty$. Wegen Unabhängigkeit von $T_n, n \in \mathbb{N}$ hängt die gemeinsame Verteilung von $(N(t_1+t) - N(t_0+t), \dots, N(t_n+t) - N(t_{n-1}+t))^\top$ nicht von t ab, falls die Verteilung von $\chi(t)$ unabhängig von t ist, also $\chi(t) \stackrel{d}{=} \chi(0) = T_1, t \geq 0$, siehe Abbildung

Zeigen wir, dass $F_{T_1} = F_{X(t)}$, $t \geq 0$.

$$\begin{aligned}
F_{X(t)}(x) &= \mathbb{P}(\chi(t) \leq x) = \sum_{n=0}^{\infty} \mathbb{P}(S_n \leq t, t < S_{n+1} \leq t+x) \\
&= \mathbb{P}(S_0 = 0 \leq t, t < S_1 = T_1 \leq t+x) \\
&\quad + \sum_{n=1}^{\infty} \mathbb{E}(\mathbb{E}(\mathbf{1}(S_n \leq t, t < S_n + T_{n+1} \leq t+x) \mid S_n)) \\
&= F_{T_1}(t+x) - F_{T_1}(t) + \sum_{n=1}^{\infty} \int_0^t \mathbb{P}(t-y < T_{n+1} \leq t+x-y) dF_{S_n}(y) \\
&= F_{T_1}(t+x) - F_{T_1}(t) + \int_0^t \mathbb{P}(t-y < T_2 \leq t+x-y) d\underbrace{\left(\sum_{n=1}^{\infty} F_{S_n}(y)\right)}_{H(y)}.
\end{aligned}$$

Falls wir zeigen könnten, dass $H(y) = \frac{y}{\mu}$, $y \geq 0$, dann hätten wir

$$\begin{aligned}
F_{X(t)}(x) &= F_{T_1}(t+x) - F_{T_1}(t) + \frac{1}{\mu} \int_t^0 (F_{T_2}(z+x) - 1 + 1 - F_{T_2}(z)) d(-z) \\
&= F_{T_1}(t+x) - F_{T_1}(t) + \frac{1}{\mu} \int_0^t (\bar{F}_{T_2}(z) - \bar{F}_{T_2}(z+x)) dz \\
&= F_{T_1}(t+x) - F_{T_1}(t) + F_{T_1}(t) - \frac{1}{\mu} \int_x^{t+x} \bar{F}_{T_2}(y) dy \\
&= F_{T_1}(t+x) - F_{T_1}(t+x) + F_{T_1}(x) = F_{T_1}(x), \quad x \geq 0,
\end{aligned}$$

nach der Form (2.1.5) der Verteilung von T_1 .

Nun soll gezeigt werden, dass $H(t) = \frac{t}{\mu}$, $t \geq 0$. Dazu verwenden wir die Formel (2.1.4): es gilt

$$\begin{aligned}
\hat{l}_{T_1}(s) &= \frac{1}{\mu} \int_0^{\infty} e^{-st} (1 - F_{T_2}(t)) dt = \frac{1}{\mu} \underbrace{\int_0^{\infty} e^{-st} dt}_{\frac{1}{s}} - \frac{1}{\mu} \int_0^{\infty} e^{-st} F_{T_2}(t) dt \\
&= \frac{1}{\mu s} \left(1 + \int_0^{\infty} F_{T_2}(t) de^{-st} \right) = \frac{1}{\mu s} \left(1 + \underbrace{e^{-st} F_{T_2}(t)}_{-F_{T_2}(0)=0} \Big|_0^{\infty} - \underbrace{\int_0^{\infty} e^{-st} dF_{T_2}(t)}_{\hat{l}_{T_2}(s)} \right) \\
&= \frac{1}{\mu s} (1 - \hat{l}_{T_2}(s)), \quad s \geq 0.
\end{aligned}$$

Mit Hilfe der Formel (2.1.4) bekommt man

$$\hat{l}_H(s) = \frac{\hat{l}_{T_1}(s)}{1 - \hat{l}_{T_2}(s)} = \frac{1}{\mu s} = \frac{1}{\mu} \int_0^{\infty} e^{-st} dt = \hat{l}_{\frac{t}{\mu}}(s), \quad s \geq 0.$$

Da die Laplace-Transformierte einer Funktion eindeutig diese Funktion bestimmt, gilt $H(t) = \frac{t}{\mu}$, $t \geq 0$. \square

Bemerkung 2.1.6

Im Beweis des Theorems 2.1.5 haben wir gezeigt, dass für den verzögerten Erneuerungsprozess

mit Verzögerung, welche die Verteilung (2.1.5) besitzt, $H(t) \sim \frac{t}{\mu}$ nicht nur asymptotisch für $t \rightarrow \infty$ (wie im elementaren Erneuerungssatz), sondern es gilt $H(t) = \frac{t}{\mu}$, für alle $t \geq 0$. Das bedeutet, es finden im Mittelwert $\frac{1}{\mu}$ Erneuerungen pro Einheitszeitintervall statt. Aus diesem Grund wird ein solcher Prozess N *homogener Erneuerungsprozess* genannt.

Es läßt sich auch folgendes Theorem beweisen:

Theorem 2.1.6

Falls $N = \{N(t), t \geq 0\}$ ein verzögerter Erneuerungsprozess mit beliebiger Verzögerung T_1 und nicht-arithmetischer Verteilung von $T_n, n \geq 2$ ist, $\mu = \mathbb{E}T_2 \in (0, \infty)$, dann gilt

$$\lim_{t \rightarrow \infty} F_{\chi(t)}(x) = \frac{1}{\mu} \int_0^x \bar{F}_{T_2}(y) dy, \quad x \geq 0.$$

Das heißt, die Grenzwertverteilung von Exzess $\chi(t), t \rightarrow \infty$ wird bei der Definition eines homogenen Erneuerungsprozesses als Verteilung von T_1 angenommen.

2.2 Poisson-artige Prozesse

2.2.1 Poisson-Prozesse

In diesem Abschnitt werden wir die Definition eines homogenen Poisson-Prozesses (gegeben im Abschnitt 1.2, Beispiel 5) verallgemeinern.

Definition 2.2.1

Der Zählprozess $N = \{N(t), t \geq 0\}$ heißt Poisson-Prozess mit Intensitätsmaß Λ , falls

1. $N(0) = 0$ f.s.
2. Λ ein lokalendliches Maß auf \mathbb{R}_+ ist, d.h., $\Lambda : \mathcal{B}(\mathbb{R}_+) \rightarrow \mathbb{R}_+$ besitzt die Eigenschaft $\Lambda(B) < \infty$ für jede beschränkte Menge $B \in \mathcal{B}(\mathbb{R}_+)$.
3. N unabhängige Zuwächse besitzt.
4. $N(t) - N(s) \sim \text{Pois}(\Lambda((s, t]))$ für alle $0 \leq s < t < \infty$.

Manchmal wird der Poisson-Prozess $N = \{N(t), t \geq 0\}$ durch das entsprechende zufällige Poissonsche Zählmaß $N = \{N(B), B \in \mathcal{B}(\mathbb{R}_+)\}$ definiert, d.h., $N = ([0, t]), t \geq 0$, wobei ein Zählmaß ein lokalendliches Maß mit Werten aus \mathbb{N}_0 ist.

Definition 2.2.2

Ein zufälliges Zählmaß $N = \{N(B), B \in \mathcal{B}(\mathbb{R}_+)\}$ heißt Poissonsches mit lokalendlichem Intensitätsmaß Λ , falls

1. Für beliebiges $n \in \mathbb{N}$ und für beliebige paarweise disjunkte beschränkte Mengen $B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R}_+)$ die Zufallsvariablen $N(B_1), N(B_2), \dots, N(B_n)$ unabhängig sind.
2. $N(B) \sim \text{Pois}(\Lambda(B)), B \in \mathcal{B}(\mathbb{R}_+), B$ -beschränkt.

Es ist klar, dass die Eigenschaften 3 und 4 der Definition 2.2.1 aus den Eigenschaften 1 und 2 der Definition 2.2.2 folgen. Die Eigenschaft 1 der Definition 2.2.1 ist jedoch eine eigenständige Annahme. $N(B), B \in \mathcal{B}(\mathbb{R}_+)$ wird als die Anzahl der Punkte von N in der Menge B interpretiert.

Bemerkung 2.2.1

Genauso wie in Definition 2.2.2 kann ein Poissonsches Zählmaß auf beliebigem topologischem Raum E , ausgestattet mit der Borel- σ -Algebra $\mathcal{B}(E)$, definiert werden. Sehr häufig wird in Anwendungen $E = \mathbb{R}^d$, $d \geq 1$ gewählt.

Lemma 2.2.1

Für jedes lokalendliche Maß Λ auf \mathbb{R}_+ existiert ein Poisson-Prozess mit Λ als Intensitätsmaß.

Beweis Falls so ein Poisson-Prozess existiert hätte, wäre die charakteristische Funktion $\varphi_{N(t)-N(s)}(\cdot)$ des Zuwachses $N(t) - N(s)$, $0 \leq s < t < \infty$ nach Eigenschaft 4 der Definition 2.2.1 gleich $\varphi_{s,t}(z) = \varphi_{\text{Pois}(\Lambda((s,t]))}(z) = e^{\Lambda((s,t))(e^{iz}-1)}$, $z \in \mathbb{R}$. Zeigen wir, dass die Familie von charakteristischen Funktionen $\{\varphi_{s,t}, 0 \leq s < t < \infty\}$ die Eigenschaft 1.7.1 besitzt: für alle $n : 0 \leq s < u < t$, $\varphi_{s,u}(z)\varphi_{u,t}(z) = e^{\Lambda((s,u))(e^{iz}-1)}e^{\Lambda((u,t))(e^{iz}-1)} = e^{(\Lambda((s,u))+\Lambda((u,t)))(e^{iz}-1)} = e^{\Lambda((s,t))(e^{iz}-1)} = \varphi_{s,t}(z)$, $z \in \mathbb{R}$, weil das Maß Λ additiv ist. Die Existenz des Poisson-Prozesses N folgt daher aus dem Theorem 1.7.1. \square

Bemerkung 2.2.2

Die Existenz eines Poissonschen Zählmaßes kann mit Hilfe des Theorems von Kolmogorov bewiesen werden, allerdings in einer allgemeineren Form wie im Theorem 1.1.2.

Aus den Eigenschaften der Poisson-Verteilung folgt u.A. $\mathbb{E}N(B) = \text{var } N(B) = \Lambda(B)$, $B \in \mathcal{B}(\mathbb{R}_+)$. Daher wird $\Lambda(B)$ als die mittlere Anzahl der Punkte von N in der Menge B , $B \in \mathcal{B}(\mathbb{R}_+)$ interpretiert.

Ein wichtiger Spezialfall liegt vor, wenn $\Lambda(dx) = \lambda dx$ für $\lambda \in (0, \infty)$, d.h., Λ proportional zum Lebesgue-Maß ν_1 auf \mathbb{R}_+ ist. Dann heißt $\lambda = \mathbb{E}N(1)$ die Intensität von N .

Wir werden demnächst zeigen, dass in diesem Fall N ein homogener Poisson-Prozess mit Intensität λ ist. Zur Erinnerung: Im Abschnitt 1.2 wurde der homogene Poisson-Prozess als ein Erneuerungsprozess mit Zwischenankunftszeiten $T_N \sim \text{Exp}(\lambda)$ definiert: $N(t) = \sup\{n \in \mathbb{N} : S_n \leq t\}$, $S_n = T_1 + \dots + T_n$, $n \in \mathbb{N}$, $t \geq 0$.

Aufgabe 2.2.1

Zeigen Sie, dass der homogene Poisson-Prozess ein homogener Erneuerungsprozess mit $T_1 \stackrel{d}{=} T_2 \sim \text{Exp}(\lambda)$ ist. Hinweis: man soll zeigen, dass für eine beliebige Exponentialverteilte Zufallsvariable X die integrierte Tailverteilungsfunktion von X gleich F_X ist.

Theorem 2.2.1

Sei $N = \{N(t), t \geq 0\}$ ein Zählprozess. Folgende Aussagen sind äquivalent.

1. N ist ein homogener Poisson-Prozess mit Intensität $\lambda > 0$.
2. a) $N(t) \sim \text{Pois}(\lambda t)$, $t \geq 0$
b) für beliebiges $n \in \mathbb{N}$, $t \geq 0$, gilt dass der Zufallsvektor (S_1, \dots, S_n) unter der Bedingung $\{N(t) = n\}$ dieselbe Verteilung besitzt, wie die Ordnungsstatistiken von u.i.v. Zufallsvariablen $U_i \in \mathcal{U}([0, t])$, $i = 1, \dots, n$.
3. a) N besitzt unabhängige Zuwächse,
b) $\mathbb{E}N(1) = \lambda$, und
c) es gilt die Eigenschaft 2b).
4. a) N hat stationäre und unabhängige Zuwächse, und

- b) es gilt $P(N(t) = 0) = 1 - \lambda t + o(t)$, $P(N(t) = 1) = \lambda t + o(t)$, $t \downarrow 0$.
5. a) N hat stationäre und unabhängige Zuwächse,
 b) es gilt die Eigenschaft 2a).

Bemerkung 2.2.3 1. Es ist klar, dass die Definition 2.2.1 mit $\Lambda(dx) = \lambda dx$, $\lambda \in (0, \infty)$ nach Theorem 2.2.1 eine äquivalente Definition des homogenen Poisson-Prozesses ist.

2. Der homogene Poisson-Prozess N wurde am Anfang des 20. Jahrhunderts von den Physikern A. Einstein und M. Smoluchovsky eingeführt, um den Zählprozess von Elementarteilchen im Geigerzähler modellieren zu können.
3. Aus 4b) folgt $P(N(t) > 1) = o(t)$, $t \downarrow 0$.
4. Die Intensität von N hat folgende Interpretation: $\lambda = \mathbf{E}N(1) = \frac{1}{\mathbf{E}T_n}$, also die mittlere Anzahl der Erneuerungen von N in einem Zeitintervall der Länge 1.
5. Die Erneuerungsfunktion vom homogenen Poisson-Prozess ist $H(t) = \lambda t$, $t \geq 0$. Dabei gilt offensichtlich $H(t) = \Lambda([0, t])$, $t > 0$.

Beweis Schema des Beweises: 1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 5) \Rightarrow 1)

1) \Rightarrow 2):

Aus 1) folgt $S_n = \sum_{k=1}^n T_k \sim \text{Erl}(n, \lambda)$, weil $T_k \sim \text{Pois}(\lambda)$, $n \in \mathbb{N}$, daher $P(N(t) = 0) = P(T_1 > t) = e^{-\lambda t}$, $t \geq 0$, und für $n \in \mathbb{N}$

$$\begin{aligned} P(N(t) = n) &= P(\{N(t) \geq n\} \setminus \{N(t) \geq n+1\}) = P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) = \int_0^t \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} dx - \int_0^t \frac{\lambda^{n+1} x^n}{n!} e^{-\lambda x} dx \\ &= \int_0^t \frac{d}{dx} \left(\frac{(\lambda x)^n}{n!} e^{-\lambda x} \right) dx = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad t \geq 0. \end{aligned}$$

Daher ist 2a) bewiesen.

Beweisen wir nun 2b). Nach dem Transformationssatz für Zufallsvektoren (vgl. Satz 3.6.1, WR) aus

$$\begin{cases} S_1 &= T_1 \\ S_2 &= T_1 + T_2 \\ &\vdots \\ S_{n+1} &= T_1 + \dots + T_{n+1} \end{cases}$$

folgt, dass die Dichte $f_{(S_1, \dots, S_n)}$ von $(S_1, \dots, S_{n+1})^\top$ durch die Dichte von $(T_1, \dots, T_{n+1})^\top$, $T_i \sim \text{Exp}(\lambda)$, u.i.v., ausgedrückt werden kann:

$$f_{(S_1, \dots, S_{n+1})}(t_1, \dots, t_{n+1}) = \prod_{k=1}^{n+1} f_{T_k}(t_k - t_{k-1}) = \prod_{k=1}^{n+1} \lambda e^{-\lambda(t_k - t_{k-1})} = \lambda^{n+1} e^{-\lambda t_{n+1}}$$

für beliebige $0 \leq t_1 \leq \dots \leq t_{n+1}$, $t_0 = 0$.

Für alle anderen t_1, \dots, t_{n+1} gilt $f_{(S_1, \dots, S_{n+1})}(t_1, \dots, t_{n+1}) = 0$.

Deshalb

$$\begin{aligned}
 f_{(S_1, \dots, S_n)}(t_1, \dots, t_n | N(t) = n) &= f_{(S_1, \dots, S_n)}(t_1, \dots, t_n | S_k \leq t, k \leq n, S_{n+1} > t) \\
 &= \frac{\int_t^\infty f_{(S_1, \dots, S_{n+1})}(t_1, \dots, t_{n+1}) dt_{n+1}}{\int_0^t \int_{t_1}^t \dots \int_{t_{n-1}}^t \int_t^\infty f_{(S_1, \dots, S_{n+1})}(t_1, \dots, t_{n+1}) dt_{n+1} dt_n \dots dt_1} \\
 &= \frac{\int_t^\infty \lambda^{n+1} e^{-\lambda t_{n+1}} dt_{n+1}}{\int_0^t \int_{t_1}^t \dots \int_{t_{n-1}}^t \int_t^\infty \lambda^{n+1} e^{-\lambda t_{n+1}} dt_{n+1} dt_n \dots dt_1} \times \\
 &\quad \times \mathbf{I}(0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t) \\
 &= \frac{n!}{t^n} \mathbf{I}(0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t).
 \end{aligned}$$

Das ist genau die Dichte von n u.i.v $\mathcal{U}([0, t])$ -Zufallsvariablen.

Aufgabe 2.2.2

Zeigen Sie es.

2) \Rightarrow 3)

Aus 2a) folgt offensichtlich 3b). Jetzt soll lediglich die Unabhängigkeit der Zuwächse von N bewiesen werden. Für beliebiges $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathbb{N}$, $t_0 = 0 < t_1 < \dots < t_n$ gilt für $x = x_1 + \dots + x_n$, dass

$$\begin{aligned}
 \mathbf{P}(\cap_{k=1}^n \{N(t_k) - N(t_{k-1}) = x_k\}) &= \underbrace{\mathbf{P}(\cap_{k=1}^n \{N(t_k) - N(t_{k-1}) = x_k\} | N(t_n) = x)}_{\frac{x!}{x_1! \dots x_n!} \prod_{k=1}^n \binom{t_k - t_{k-1}}{t_n}^{x_k} \text{ nach 2c)}} \times \\
 &\quad \times \underbrace{\mathbf{P}(N(t_n) = x)}_{e^{-\lambda t_n} \frac{(\lambda t_n)^x}{x!} \text{ nach 2a)}} \\
 &= \prod_{k=1}^n \frac{(\lambda(t_k - t_{k-1}))^{x_k}}{x_k!} e^{-\lambda(t_k - t_{k-1})},
 \end{aligned}$$

weil die Wahrscheinlichkeit von (*) die Polynomialverteilung mit Parametern n , $\left\{ \frac{t_k - t_{k-1}}{t_n} \right\}_{k=1}^n$ angehört. Denn das Ereignis (*) ist es, beim unabhängigen gleichverteilten Werfen von x Punkte auf $[0, t]$, jeweils x_k Punkte im Korb der Länge $t_k - t_{k-1}$, $k = 1, \dots, n$ zu haben:

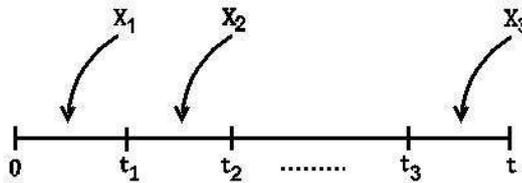


Abbildung 2.4:

Damit ist 3a) bewiesen, weil $\mathbf{P}(\cap_{k=1}^n \{N(t_k) - N(t_{k-1}) = x_k\}) = \prod_{k=1}^n \mathbf{P}(\{N(t_k) - N(t_{k-1}) = x_k\})$.

3) \Rightarrow 4)

Zeigen wir, dass N stationäre Zuwächse besitzt. Für beliebiges $n \in \mathbb{N}_0$, $x_1, \dots, x_n \in \mathbb{N}$, $t_0 = 0 < t_1 < \dots < t_n$ und $h > 0$ betrachten wir $I(h) = \mathbb{P}(\cap_{k=1}^n \{N(t_k + h) - N(t_{k-1} + h) = x_k\})$ und zeigen, dass $I(h)$ nicht von $h \in \mathbb{R}$ abhängt. Nach der Formel der totalen Wahrscheinlichkeit gilt

$$\begin{aligned} I(h) &= \sum_{m=0}^{\infty} \mathbb{P}(\cap_{k=1}^n \{N(t_k + h) - N(t_{k-1} + h) = x_k\} \mid N(t_n + h) = m) \cdot \mathbb{P}(N(t_n + h) = m) \\ &= \sum_{m=0}^{\infty} \frac{m!}{x_1! \dots x_n!} \prod_{k=1}^n \left(\frac{t_k + h - t_{n-1} - h}{t_n + h - h} \right)^{x_k} e^{-\lambda(t_n + h)} \frac{(\lambda(t_n + h))^m}{m!} \\ &= \sum_{m=0}^{\infty} \mathbb{P}(\cap_{k=1}^n \{N(t_k) - N(t_{k-1}) = x_k \mid N(t_n + h) = m\}) \times \mathbb{P}(N(t_n + h) = m) = I(0). \end{aligned}$$

Zeigen wir nun die Eigenschaft 4b) für $h \in (0, 1)$:

$$\begin{aligned} \mathbb{P}(N(h) = 0) &= \sum_{k=0}^{\infty} \mathbb{P}(N(h) = 0, N(1) = k) = \sum_{k=0}^{\infty} \mathbb{P}(N(h) = 0, N(1) - N(h) = k) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N(1) - N(h) = k, N(1) = k) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N(1) = k) \mathbb{P}(N(1) - N(h) = k \mid N(1) = k) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N(1) = k) (1 - h)^k. \end{aligned}$$

Es ist zu zeigen, dass $\mathbb{P}(N(h) = 0) = 1 - \lambda h + o(h)$, d.h., $\lim_{h \rightarrow 0} \frac{1}{h} (1 - \mathbb{P}(N(h) = 0)) = \lambda$. In der Tat es gilt

$$\begin{aligned} \frac{1}{h} (1 - \mathbb{P}(N(h) = 0)) &= \frac{1}{h} \left(1 - \sum_{k=0}^{\infty} \mathbb{P}(N(1) = k) (1 - h)^k \right) = \sum_{k=1}^{\infty} \mathbb{P}(N(1) = k) \cdot \frac{1 - (1 - h)^k}{h} \\ &\xrightarrow{h \rightarrow 0} \sum_{k=1}^{\infty} \mathbb{P}(N(1) = k) \underbrace{\lim_{h \rightarrow 0} \frac{1 - (1 - h)^k}{h}}_k \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N(1) = k) k = \mathbb{E}N(1) = \lambda, \end{aligned}$$

weil diese Reihe gleichmäßig in h konvergiert, da sie durch $\sum_{k=0}^{\infty} \mathbb{P}(N(1) = k) k = \lambda < \infty$ dominiert wird wegen der Ungleichung $(1 - h)^k \geq 1 - kh$, $h \in (0, 1)$, $k \in \mathbb{N}$.

Ähnlich kann man zeigen, dass $\lim_{h \rightarrow 0} \frac{\mathbb{P}(N(h)=1)}{h} = \lim_{h \rightarrow 0} \sum_{k=1}^{\infty} \mathbb{P}(N(1) = k) k (1 - h)^{k-1} = \lambda$.

4) \Rightarrow 5)

Zu zeigen ist es, dass für beliebiges $n \in \mathbb{N}$ und $t \geq 0$

$$p_n(t) = \mathbb{P}(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad (2.2.1)$$

gilt. Wir beweisen dies induktiv bezüglich n . Zunächst zeigen wir, dass $p_0(t) = e^{-\lambda t}$, $h = 0$. Dazu betrachten wir $p_0(t+h) = \mathbb{P}(N(t+h) = 0) = \mathbb{P}(N(t) = 0, N(t+h) - N(t) = 0) = p_0(t)p_0(h) =$

$p_0(t)(1 - \lambda h + o(h))$, $h \rightarrow 0$. Ähnlich kann man zeigen, dass $p_0(t) = p_0(t - h)(1 - \lambda h + o(h))$, $h \rightarrow 0$. Somit gilt $p_0'(t) = \lim_{h \rightarrow 0} \frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t)$, $t > 0$. Da $p_0(0) = \mathbb{P}(N(0) = 0) = 1$, folgt aus

$$\begin{cases} p_0'(t) &= -\lambda p_0(t) \\ p_0(0) &= 0, \end{cases}$$

dass es eine eindeutige Lösung $p_0(t) = e^{-\lambda t}$, $t \geq 0$ existiert. Nun sei für n die Darstellung (2.2.1) bewiesen. Beweisen wir sie für $n + 1$.

$$\begin{aligned} p_{n+1}(t+h) &= \mathbb{P}(N(t+h) = n+1) \\ &= \mathbb{P}(N(t) = n, N(t+h) - N(t) = 1) + \mathbb{P}(N(t) = n+1, N(t+h) - N(t) = 0) \\ &= p_n(t) - p_1(h) + p_{n+1}(t) - p_0(h) \\ &= p_n(t)(\lambda h + o(h)) + p_{n+1}(t)(1 - \lambda h + o(h)), \quad h \rightarrow 0, h > 0. \end{aligned}$$

Daher

$$\begin{cases} p_{n+1}'(t) &= -\lambda p_{n+1}(t) + \lambda p_n(t), \quad t > 0 \\ p_{n+1}(0) &= 0 \end{cases} \tag{2.2.2}$$

Da $p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$, bekommt man $p_{n+1}(t) = e^{-\lambda t} \frac{(\lambda t)^{n+1}}{(n+1)!}$ als Lösung von (2.2.2). (In der Tat $p_{n+1}(t) = C(t)e^{-\lambda t} \Rightarrow C'(t)e^{-\lambda t} = \lambda C(t)e^{-\lambda t} + \lambda p_n(t)$
 $C'(t) = \frac{\lambda^{n+1} t^n}{n!} \Rightarrow C(t) = \frac{\lambda^{n+1} t^{n+1}}{(n+1)!}$, $C(0) = 0$
 5) \Rightarrow 1)

Sei N ein Zählprozess $N(t) = \max\{n : S_n \leq t\}$, $t \geq 0$, der Bedingungen 5a) und 5b) erfüllt. Zeigen wir, dass $S_n = \sum_{k=1}^n T_k$, wobei T_k i.i.d. mit $T_k \sim \text{Exp}(\lambda)$, $k \in \mathbb{N}$. Da $T_k = S_k - S_{k-1}$, $k \in \mathbb{N}$, $S_0 = 0$, betrachten wir für $b_0 = 0 \leq a_1 < b_1 \leq \dots \leq a_n < b_n$

$$\begin{aligned} &\mathbb{P}(\cap_{k=1}^n \{a_k < S_k \leq b_k\}) \\ &= \mathbb{P}(\cap_{k=1}^{n-1} \{N(a_k) - N(b_{k-1}) = 0, N(b_k) - N(a_k) = 1\} \\ &\quad \cap \{N(a_n) - N(b_{n-1}) = 0, N(b_n) - N(a_n) \geq 1\}) \\ &= \prod_{k=1}^{n-1} (\underbrace{\mathbb{P}(N(a_k) - b_{k-1}) = 0}_{e^{-\lambda(a_k - b_{k-1})}} \underbrace{\mathbb{P}(N(b_k) - a_k) = 1}_{\lambda(b_k - a_k)e^{-\lambda(b_k - a_k)}}) \times \\ &\quad \underbrace{\mathbb{P}(N(a_n) - b_{n-1}) = 0}_{e^{-\lambda(a_n - b_{n-1})}} \underbrace{\mathbb{P}(N(b_n) - a_n) \geq 1}_{(1 - e^{-\lambda(b_n - a_n)})} \\ &= e^{-\lambda(a_n - b_{n-1})} (1 - e^{-\lambda(b_n - a_n)}) \prod_{k=1}^{n-1} \lambda(b_k - a_k) e^{-\lambda(b_k - a_k)} \\ &= \lambda^{n-1} (e^{-\lambda a_n} - e^{-\lambda b_n}) \prod_{k=1}^{n-1} (b_k - a_k) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} \lambda^n e^{-\lambda y_n} dy_n \dots y_1. \end{aligned}$$

Die gemeinsame Dichte von $(S_1, \dots, S_n)^\top$ ist also gegeben durch $\lambda^n e^{-\lambda y_n} \mathbf{1}(y_1 \leq y_2 \leq \dots \leq y_n)$.
 \square

2.2.2 Zusammengesetzter Poisson-Prozess

Definition 2.2.3

Sei $N = \{N(t), t \geq 0\}$ ein homogener Poisson-Prozess mit Intensität $\lambda > 0$, konstruiert mit

Hilfe der Folge $\{T_n\}_{n \in \mathbb{N}}$ von Zwischenankunftszeiten. Sei $\{U_n\}_{n \in \mathbb{N}}$ eine Folge von u.i.v. Zufallsvariablen, unabhängigen von $\{T_n\}_{n \in \mathbb{N}}$. Sei F_U die Verteilungsfunktion von U_1 . Für beliebiges $t \geq 0$ setze $X(t) = \sum_{k=1}^{N(t)} U_k$. Der stochastische Prozess $X = \{X(t), t \geq 0\}$ heißt *zusammengesetzter Poisson-Prozess mit Parametern λ, F_U* . Die Verteilung von $X(t)$ heißt dabei *zusammengesetzte Poisson-Verteilung mit Parametern $\lambda t, F_U$* .

Zusammengesetzter Poisson-Prozess $X(t), t \geq 0$ kann als Summe der „Marken“ U_n eines homogenen markierten Poisson-Prozesses (N, U) bis zur Zeit t interpretiert werden.

So wird $X(t)$ als Gesamtarbeitsbelastung eines Servers bis zur Zeit t in der Warteschlangentheorie interpretiert, falls die Aufforderungen zum Service zu Zeitpunkten $S_n = \sum_{k=1}^n T_k, n \in \mathbb{N}$ eingehen und mit sich den Arbeitsaufwand $U_n, n \in \mathbb{N}$ mitbringen.

In der Versicherungsmathematik ist $X(t), t \geq 0$ der Gesamtschaden eines Portfolios bis zum Zeitpunkt $t \geq 0$ mit Schadenanzahl $N(t)$ und Schadenhöhen $U_n, n \in \mathbb{N}$.

Theorem 2.2.2

Sei $X = \{X(t), t \geq 0\}$ ein zusammengesetzter Poisson-Prozess mit Parametern λ, F_U . Es gelten folgende Eigenschaften:

1. X hat unabhängige stationäre Zuwächse.
2. Falls $\hat{m}_U(s) = \mathbb{E}e^{sU_1}, s \in \mathbb{R}$, die momenterzeugende Funktion von U_1 ist, so dass $\hat{m}_U(s) < \infty, s \in \mathbb{R}$, dann gilt

$$\hat{m}_{X(t)}(s) = e^{\lambda t(\hat{m}_U(s)-1)}, s \in \mathbb{R}, t \geq 0, \quad \mathbb{E}X(t) = \lambda t \mathbb{E}U_1, \quad \text{var } X(t) = \lambda t \mathbb{E}U_1^2, t \geq 0.$$

Beweis 1. Zu zeigen ist, dass für beliebige $n \in \mathbb{N}, 0 \leq t_0 < t_1 < \dots < t_n$ und h

$$\mathbb{P} \left(\sum_{i_1=N(t_0+h)+1}^{N(t_1+h)} U_{i_1} \leq x_1, \dots, \sum_{i_n=N(t_{n-1}+h)+1}^{N(t_n+h)} U_{i_n} \leq x_n \right) = \prod_{k=1}^n \mathbb{P} \left(\sum_{i_k=N(t_{k-1}+1)+1}^{N(t_k)} U_{i_k} \leq x_k \right)$$

für beliebige $x_1, \dots, x_n \in \mathbb{R}$. In der Tat, gilt

$$\begin{aligned} & \mathbb{P} \left(\sum_{i_1=N(t_0+h)+1}^{N(t_1+h)} U_{i_1} \leq x_1, \dots, \sum_{i_n=N(t_{n-1}+h)+1}^{N(t_n+h)} U_{i_n} \leq x_n \right) \\ &= \sum_{k_1, \dots, k_n=0}^{\infty} \left(\prod_{j=1}^n F_n^{*k_j}(x_j) \right) \mathbb{P}(\cap_{m=1}^n \{N(t_m+h) - N(t_{m-1}+h) = k_m\}) \\ &= \sum_{k_1, \dots, k_n=0}^{\infty} \left(\prod_{j=1}^n F_n^{*k_j}(x_j) \right) \left(\prod_{m=1}^n \mathbb{P}(N(t_m) - N(t_{m-1}) = k_m) \right) \\ &= \prod_{m=1}^n \sum_{k_m=0}^{\infty} F_n^{*k_m}(x_m) \mathbb{P}(N(t_m) - N(t_{m-1}) = k_m) \\ &= \prod_{m=1}^n \mathbb{P} \left(\sum_{k_m=N(t_{m-1}+1)+1}^{N(t_m)} \leq x_m \right) \end{aligned}$$

2.

Aufgabe 2.2.3

□

2.2.3 Cox-Prozess

Ein Cox-Prozess ist ein (im Allgemeinen inhomogener) Poisson-Prozess mit Intensitätsmaß Λ , das an sich ein zufälliges Maß darstellt. Diese intuitive Vorstellung kann in folgender Definition formalisiert werden.

Definition 2.2.4

Sei $\Lambda = \{\Lambda(B), B \in \mathcal{B}(\mathbb{R}_+)\}$ ein zufälliges f.s. lokal-endliches Maß. Das zufällige Zählmaß $N = \{N(B), B \in \mathcal{B}(\mathbb{R}_+)\}$ wird *Cox-Zählmaß (oder doppelt-stochastisches Poisson-Maß) mit zufälligem Intensitätsmaß* Λ genannt, falls für beliebige $n \in \mathbb{N}$, $k_1, \dots, k_n \in \mathbb{N}_0$ und $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_n < b_n$ gilt $\mathbb{P}(\cap_{i=1}^n \{N((a_i, b_i]) = k_i\}) = \mathbb{E} \left(\prod_{i=1}^n e^{-\Lambda((a_i, b_i])} \frac{\Lambda^{k_i}((a_i, b_i])}{k_i!} \right)$. Der Prozess $\{N(t), t \geq 0\}$ mit $N(t) = N((0, t])$ heißt *Cox-Prozess (oder doppelt-stochastischer Poisson-Prozess)* mit zufälligem Intensitätsmaß Λ .

Beispiel 2.2.1 1. Falls das zufällige Maß Λ f.s. absolut stetig bzgl. des Lebesgue-Maßes ist, d.h., $\Lambda(B) = \int_B \lambda(t) dt$, B - beschränkt, $B \in \mathcal{B}(\mathbb{R}_+)$, wobei $\{\lambda(t), t \geq 0\}$ ein stochastischer Prozess mit f.s. Borel-messbaren borel-integrierbaren Trajektorien ist, $\lambda(t) \geq 0$ f.s. für alle $t \geq 0$, der Intensitätsprozess von N genannt wird.

2. Insbesondere kann $\lambda(t) \equiv Y$ sein, wobei Y eine nicht-negative Zufallsvariable ist. Dann gilt $\Lambda(B) = Y \nu_1(B)$, also hat N eine zufällige Intensität Y . Solche Cox-Prozesse werden *gemischte Poisson-Prozesse* genannt.

Einen Cox-Prozess $N = \{N(t), t \geq 0\}$ mit Intensitätsprozess $\{\lambda(t), t \geq 0\}$ kann man wie folgt explizit konstruieren. Sei $\tilde{N} = \{\tilde{N}(t), t \geq 0\}$ ein homogener Poisson-Prozess mit Intensität 1, der unabhängig von $\{\lambda(t), t \geq 0\}$ ist. Dann ist $N \stackrel{d}{=} N_1$, wobei der Prozess $N_1 = \{N_1(t), t \geq 0\}$ gegeben ist durch $N_1(t) = \tilde{N}(\int_0^t \lambda(y) dy)$, $t \geq 0$. Die Aussage $N \stackrel{d}{=} N_1$ soll natürlich bewiesen werden. Wir werden sie jedoch ohne Beweis annehmen. Sie bildet auch die Grundlage für die Simulation des Cox-Prozesses N .

2.3 Ergänzende Aufgaben

Aufgabe 2.3.1

Sei $\{N_t\}_{t \geq 0}$ ein Erneuerungsprozess mit Zwischenankunftszeiten T_i , welche exponentialverteilt sind, d.h. $T_i \sim \text{Exp}(\lambda)$.

- Beweisen Sie: N_t ist Poisson-verteilt für jedes $t > 0$.
- Bestimmen Sie den Parameter dieser Poisson-Verteilung.
- Bestimmen Sie die Erneuerungsfunktion $H(t) = \mathbb{E} N_t$.

Aufgabe 2.3.2

Beweisen Sie: Ein (reellwertiger) stochastischer Prozess $X = \{X(t), t \in [0, \infty)\}$ mit unabhängigen Zuwächsen hat bereits dann stationäre Zuwächse, wenn die Verteilung der Zufallsvariablen $X(t+h) - X(h)$ unabhängig von h ist.

Aufgabe 2.3.3

Sei $N = \{N(t), t \in [0, \infty)\}$ ein Poisson-Prozess mit Intensität λ . Berechnen Sie die Wahrscheinlichkeiten dafür, dass im Intervall $[0, s]$ genau i Ereignisse auftreten unter der Bedingung,

dass im Intervall $[0, t]$ genau n Ereignisse eintreten, d.h. $P(N(s) = i \mid N(t) = n)$ für $s < t$, $i = 0, 1, \dots, n$.

Aufgabe 2.3.4

Seien $N^{(1)} = \{N^{(1)}(t), t \in [0, \infty)\}$ und $N^{(2)} = \{N^{(2)}(t), t \in [0, \infty)\}$ unabhängige Poisson-Prozesse mit den Intensitäten λ_1 bzw. λ_2 . Die Unabhängigkeit soll in diesem Fall bedeuten, dass die Folgen $T_1^{(1)}, T_2^{(1)}, \dots$ und $T_1^{(2)}, T_2^{(2)}, \dots$ unabhängig sind. Zeigen Sie, dass $N = \{N(t) := N^{(1)}(t) + N^{(2)}(t), t \in [0, \infty)\}$ ein Poisson-Prozess mit Intensität $\lambda_1 + \lambda_2$ ist.

Aufgabe 2.3.5 (Wartezeitenparadoxon):

Für einen Erneuerungsprozess $N = \{N(t), t \in [0, \infty)\}$ heißt $T(t) = S_{N(t)+1} - t$ die *Exzesszeit*, $C(t) = t - S_{N(t)}$ die *aktuelle Lebenszeit* und $D(t) = T(t) + C(t) = T_{N(t)+1}$ die *Lebensdauer* zum Zeitpunkt $t > 0$. Sei nun $N = \{N(t), t \in [0, \infty)\}$ ein Poisson-Prozess mit Intensität λ .

- Berechnen Sie die Verteilung der Exzesszeit $T(t)$.
- Zeigen Sie, dass die Verteilung der aktuellen Lebenszeit durch $P(C(t) = t) = e^{-\lambda t}$ und die Dichte $f_{C(t) \mid N(t) > 0}(s) = \lambda e^{-\lambda s} \mathbf{1}\{s \leq t\}$ gegeben ist.
- Zeigen Sie, dass $P(D(t) \leq x) = (1 - (1 + \lambda \min\{t, x\})e^{-\lambda x}) \mathbf{1}\{x \geq 0\}$.
- Um $ET(t)$ zu bestimmen, könnte man folgendermaßen argumentieren: Im Mittel liegt t in der Mitte des umgebenden Zwischenankunftsintervalls $(S_{N(t)}, S_{N(t)+1})$, d.h. $ET(t) = \frac{1}{2}E(S_{N(t)+1} - S_{N(t)}) = \frac{1}{2}ET_{N(t)+1} = \frac{1}{2\lambda}$. In Anbetracht des Ergebnisses aus Teil (a) kann dieses Argument nicht stimmen. Wo liegt der Fehler in der Argumentation?

Aufgabe 2.3.6

Gegeben sei ein zusammengesetzter Poisson-Prozess $X = \{X(t) := \sum_{i=1}^{N(t)} U_i, t \geq 0\}$. Sei $M_{N(t)}(s) = Es^{N(t)}$, $s \in (0, 1)$, die erzeugende Funktion des Poisson-Prozesses $N(t)$, $\mathcal{L}\{U\}(s) = E \exp\{-sU\}$ die Laplace-Transformierte von U_i , $i \in \mathbb{N}$, und $\mathcal{L}\{X(t)\}(s)$ die Laplace-Transformierte von $X(t)$. Beweisen Sie, dass

$$\mathcal{L}\{X(t)\}(s) = M_{N(t)}(\mathcal{L}\{U\}(s)), \quad s \geq 0$$

gilt.

Aufgabe 2.3.7

Gegeben sei ein zusammengesetzter Poisson-Prozess $X = \{X(t), t \in [0, \infty)\}$ mit U_i u.i.v., $U_1 \sim \text{Exp}(\gamma)$, wobei die Intensität von $N(t)$ durch λ gegeben sei. Zeigen Sie, dass für die Laplace-Transformierte $\mathcal{L}\{X(t)\}(s)$ von $X(t)$ gilt:

$$\mathcal{L}\{X(t)\}(s) = \exp\left\{-\frac{\lambda t s}{\gamma + s}\right\}.$$

Aufgabe 2.3.8

Schreiben Sie eine Funktion in **R** (alternativ: Java), der als Parameter ein Zeitpunkt t , eine Intensität λ und ein Wert γ übergeben werden und die als Ergebnis den zufälligen Wert eines zusammengesetzten Poisson-Prozesses mit Charakteristiken $(\lambda, \text{Exp}(\gamma))$ zum Zeitpunkt t ausgibt. *Hinweis: die Lösungen sollen als kommentierter, strukturierter und lesbarer Ausdruck abgegeben werden.*

Aufgabe 2.3.9

Der stochastische Prozess $N = \{N(t), t \in [0, \infty)\}$ sei ein Cox-Prozess mit Intensitätsfunktion $\lambda(t) = Z$, wobei Z eine diskrete Zufallsvariable ist, welche die Werte λ_1 und λ_2 jeweils mit Wahrscheinlichkeit $1/2$ annimmt. Bestimmen Sie die momenterzeugende Funktion sowie den Erwartungswert und die Varianz von $N(t)$.

Aufgabe 2.3.10

Gegeben seien zwei unabhängige homogene Poisson-Prozesse $N^{(1)} = \{N^{(1)}(t), t \in [0, \infty)\}$ und $N^{(2)} = \{N^{(2)}(t), t \geq 0\}$ mit den Intensitäten λ_1 und λ_2 . Weiter sei $X \geq 0$ eine beliebige nichtnegative Zufallsvariable, die von $N^{(1)}$ und $N^{(2)}$ unabhängig ist. Zeigen Sie, dass der Prozess $N = \{N(t), t \geq 0\}$ mit

$$N(t) = \begin{cases} N^{(1)}(t), & t \leq X, \\ N^{(1)}(X) + N^{(2)}(t - X), & t > X \end{cases}$$

ein Cox-Prozess ist, dessen Intensitätsprozess $\lambda = \{\lambda(t), t \geq 0\}$ gegeben ist durch

$$\lambda(t) = \begin{cases} \lambda_1, & t \leq X, \\ \lambda_2, & t > X. \end{cases}$$

3 Wiener-Prozess

3.1 Elementare Eigenschaften

Im Beispiel 2) des Abschnittes 1.2 haben wir die *Brownsche Bewegung* (oder *Wiener-Prozess*) $W = \{W(t), t \geq 0\}$ definiert als einen Gaußschen Prozess mit $\mathbb{E}W(t) = 0$ und $\text{cov}(W(s), W(t)) = \min\{s, t\}$, $s, t \geq 0$. Geben wir jetzt eine neue (äquivalente) Definition an.

Definition 3.1.1

Ein stochastischer Prozess $W = \{W(t), t \geq 0\}$ heißt Wiener-Prozess (oder Brownsche Bewegung), falls

1. $W(0) = 0$ f.s.
2. W hat unabhängige Zuwächse
3. $W(t) - W(s) \sim \mathcal{N}(0, t - s)$, $0 \leq s < t$

Die Existenz von W nach Definition 3.1.1 folgt aus dem Satz 1.7.1, weil nämlich $\varphi_{s,t}(z) = \mathbb{E}e^{iz(W(t)-W(s))} = e^{-\frac{(t-s)z^2}{2}}$, $z \in \mathbb{R}$, und $e^{-\frac{(t-u)z^2}{2}}e^{-\frac{(u-s)z^2}{2}} = e^{-\frac{(t-s)z^2}{2}}$ für $0 \leq s < u < t$, also $\varphi_{s,u}(z)\varphi_{u,t}(z) = \varphi_{s,t}(z)$, $z \in \mathbb{R}$. Aus dem Satz 1.3.1 folgt außerdem die Existenz einer Version mit stetigen Trajektorien.

Aufgabe 3.1.1

Zeigen Sie, dass das Theorem für $\alpha = 3$, $\sigma = \frac{1}{2}$.

Der Wiener-Prozess ist nach dem Mathematiker Norbert Wiener (1894 - 1964) benannt worden. Warum existiert dann die Brownsche Bewegung? Aus dem Satz von Kolmogorov (Satz 1.1.2) existiert für jede Funktion $\mu : \mathbb{R}_+ \rightarrow \mathbb{R}$ und jede positiv semi-definite Funktion $C : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ ein reellwertiger Gaußscher Prozess $X = \{X(t), t \geq 0\}$ mit Mittelwert $\mathbb{E}X(t) = \mu(t)$, $t \geq 0$, und Kovarianzfunktion $\text{cov}(X(s), X(t)) = C(s, t)$, $s, t \geq 0$. Es bleibt lediglich zu zeigen, dass $C(s, t) = \min\{s, t\}$, $s, t \geq 0$ positiv-semidefinit ist.

Aufgabe 3.1.2

Zeigen Sie es.

Deswegen wird oft angenommen, dass der Wiener-Prozess stetige Pfade besitzt (man nimmt einfach die entsprechende Version von ihm).

Theorem 3.1.1

Beide Definitionen des Wiener-Prozesses sind äquivalent.

Beweis 1. Aus der Definition im Abschnitt 1.2 folgt die Definition 3.1.1.

$W(0) = 0$ f.s. folgt aus $\text{var}(W(0)) = \min\{0, 0\} = 0$. Beweisen wir nun, dass die Zuwächse von W unabhängig sind. Falls $Y \sim \mathcal{N}(\mu, K)$ ein n -dimensionaler Gaußscher Zufallsvektor ist und A eine $(n \times n)$ -Matrix, dann gilt $AY \sim \mathcal{N}(A\mu, AK A^\top)$, dies folgt aus der expliziten Form der charakteristischen Funktion von Y . Sei nun $n \in \mathbb{N}$, $0 = t_0 \leq t_1 < \dots < t_n$,

$Y = (W(t_0), W(t_1), \dots, W(t_n))^T$. Es gilt für $Z = (W(t_0), W(t_1) - W(t_0), \dots, W(t_n) - W(t_{n-1}))^T$, dass $Z = AY$, wobei

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & \dots & 0 \\ -1 & 1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

Daher ist Z auch Gaußsch mit einer Kovarianzmatrix, die diagonal ist. In der Tat gilt $\text{cov}(W(t_{i+1}) - W(t_i), W(t_{j+1}) - W(t_j)) = \min\{t_{i+1}, t_{j+1}\} - \min\{t_{i+1}, t_j\} - \min\{t_i, t_{j+1}\} + \min\{t_i, t_j\} = 0$ für $i \neq j$. Daher sind die Koordinaten von Z unkorreliert, was für die multivariate Gaußsche Verteilung Unabhängigkeit bedeutet. Deshalb sind die Zuwächse von W unabhängig. Weiterhin, gilt für beliebiges $0 \leq s < t$, dass $W(t) - W(s) \sim \mathcal{N}(0, t - s)$. Die Normalverteilttheit folgt aus der Tatsache, dass $Z = AY$ Gaußsch ist, offensichtlich gilt $\mathbb{E}W(t) - \mathbb{E}W(s) = 0$ und $\text{var}(W(t) - W(s)) = \text{var}(W(t)) - 2\text{cov}(W(s), W(t)) + \text{var}(W(s)) = t - 2\min\{s, t\} + s = t - s$.

2. Aus der Definition 3.1.1 folgt die Definition im Abschnitt 1.2. Da $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ für $0 \leq s < t$, gilt $\text{cov}(W(s), W(t)) = \mathbb{E}[W(s)(W(t) - W(s) + W(s))] = \mathbb{E}W(s)\mathbb{E}(W(t) - W(s)) + \text{var}W(s) = s$, daher gilt $\text{cov}(W(s), W(t)) = \min\{s, t\}$. Aus $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ und $W(0) = 0$ folgt auch, dass $\mathbb{E}W(t) = 0, t \geq 0$. Die Tatsache, dass W ein Gaußscher Prozess ist, folgt aus der Relation $Y = A^{-1}Z$, Punkt 1) des Beweises. □

Definition 3.1.2

Der Prozess $\{W(t), t \geq 0\}$, $W(t) = (W_1(t), \dots, W_d(t))^T, t \geq 0$, heißt *d-dimensionale Brownsche Bewegung*, falls $W_i = \{W_i(t), t \geq 0\}$ unabhängige Wiener-Prozesse sind, $i = 1, \dots, d$.

Die obigen Definitionen und Übungsaufgabe 3.1.1 garantieren uns die Existenz eines Wiener-Prozesses mit stetigen Pfaden. Wie kann man aber eine explizite Konstruktion dieser Pfade angeben? Damit befassen wir uns im nächsten Abschnitt.

3.2 Explizite Konstruktion des Wiener-Prozesses

Konstruieren wir den Wiener-Prozess zunächst auf dem Intervall $[0, 1]$. Die Hauptidee der Konstruktion ist es, einen stochastischen Prozess $X = \{X(t), t \in [0, 1]\}$ einzuführen, der auf einem Teilwahrscheinlichkeitsraum von $(\Omega, \mathcal{A}, \mathbb{P})$ definiert ist mit $X \stackrel{d}{=} W$, wobei $X(t) = \sum_{n=1}^{\infty} c_n(t)Y_n, t \in [0, 1], \{Y_n\}_{n \in \mathbb{N}}$ eine Folge von u.i.v. $\mathcal{N}(0, 1)$ -Zufallsvariablen und $c_n(t) = \int_0^t H_n(s)ds, t \in [0, 1], n \in \mathbb{N}$, ist. Hier soll $\{H_n\}_{n \in \mathbb{N}}$ die orthonormierte Haar-Basis im $L_2([0, 1])$ sein, die jetzt kurz eingeführt wird.

3.2.1 Haar- und Schauder-Funktionen

Definition 3.2.1

Die Funktionen $H_n : [0, 1] \rightarrow \mathbb{R}, n \in \mathbb{N}$, heißen *Haar-Funktionen*, falls $H_1(t) = 1, t \in [0, 1], H_2(t) = 1_{[0, \frac{1}{2}]}(t) - 1_{(\frac{1}{2}, 1]}(t), H_k(t) = 2^{\frac{n}{2}}(1_{I_{n,k}}(t) - 1_{J_{n,k}}(t)), t \in [0, 1], 2^n < k \leq 2^{n+1}$, wobei $I_{n,k} = [a_{n,k}, a_{n,k} + 2^{-n-1}], J_{n,k} = (a_{n,k} + 2^{-n-1}, a_{n,k} + 2^{-n}], a_{n,k} = 2^{-n}(k - 2^n - 1), n \in \mathbb{N}$.

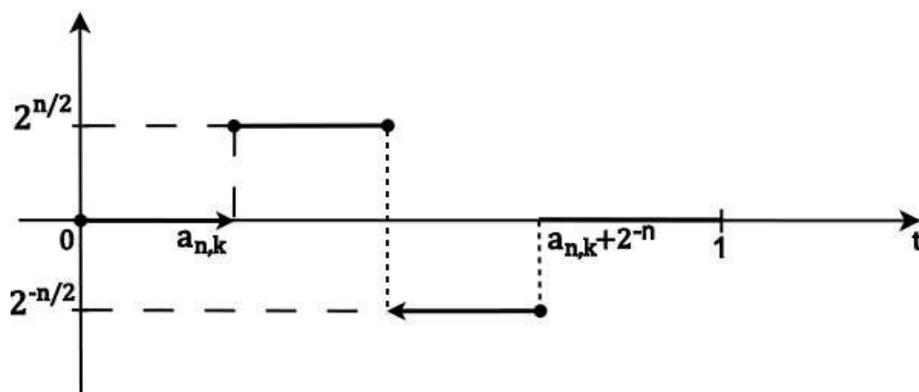


Abbildung 3.1: Haar-Funktionen

Lemma 3.2.1

Das Funktionssystem $\{H_n\}_{n \in \mathbb{N}}$ ist eine orthonormierte Basis in $L^2([0, 1])$ mit dem Skalarprodukt $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$, $f, g \in L^2([0, 1])$.

Beweis Die Orthonormiertheit des Systems $\langle H_k, H_n \rangle = \delta_{kn}$, $k, n \in \mathbb{N}$ folgt unmittelbar aus der Definition 3.2.1. Zeigen wir die Vollständigkeit von $\{H_n\}_{n \in \mathbb{N}}$. Es genügt zu zeigen, dass für beliebige Funktion $g \in L^2([0, 1])$ mit $\langle g, H_n \rangle = 0$, $n \in \mathbb{N}$, $g = 0$ fast überall auf $[0, 1]$ gilt. In der Tat kann die Indikator-Funktion eines Intervalls $1_{[a_{n,k}, a_{n,k}+2^{-n-1}]}$ stets als Linearkombination von H_n , $n \in \mathbb{N}$ aufgeschrieben werden.

$$\begin{aligned} 1_{[0, \frac{1}{2}]} &= \frac{(H_1 + H_2)}{2}, \\ 1_{(\frac{1}{2}, 1]} &= \frac{(H_1 - H_2)}{2}, \\ 1_{[0, \frac{1}{4}]} &= \frac{(1_{[0, \frac{1}{2}]} + \frac{1}{\sqrt{2}}H_2)}{2}, \\ 1_{(\frac{1}{4}, \frac{1}{2}]} &= \frac{(1_{[0, \frac{1}{2}]} - \frac{1}{\sqrt{2}}H_2)}{2}, \\ &\vdots \\ 1_{[a_{n,k}, a_{n,k}+2^{-n-1}]} &= \frac{(1_{[a_{n,k}, a_{n,k}+2^{-n}]} + 2^{-\frac{n}{2}}H_k)}{2}, \quad 2^n < k \leq 2^{n+1}. \end{aligned}$$

Daher gilt $\int_{\frac{k}{2^n}}^{\frac{k+1}{2^n}} g(t)dt = 0$, $n \in \mathbb{N}_0$, $k = 1, \dots, 2^n - 1$, und deshalb $G(t) = \int_0^t g(s)ds = 0$ für $t = \frac{k}{2^n}$, $n \in \mathbb{N}_0$, $k = 1, \dots, 2^n - 1$. Da G stetig auf $[0, 1]$ ist, folgt heraus $G(t) = 0$, $t \in [0, 1]$, und somit $g(s) = 0$ für fast jedes $s \in [0, 1]$. \square

Aus Lemma 3.2.1 folgt, dass zwei beliebige Funktionen $f, g \in L^2([0, 1])$ die Darstellungen $f = \sum_{n=1}^{\infty} \langle f, H_n \rangle H_n$ und $g = \sum_{n=1}^{\infty} \langle g, H_n \rangle H_n$ haben (diese Reihen konvergieren im $L^2([0, 1])$) und $\langle f, g \rangle = \sum_{n=1}^{\infty} \langle f, H_n \rangle \langle g, H_n \rangle$ (Parseval-Identität).

Definition 3.2.2

Die Funktionen $S_n(t) = \int_0^t H_n(s)ds = \langle 1_{[0,t]}, H_n \rangle$, $t \in [0, 1]$, $n \in \mathbb{N}$ heißen *Schauder-Funktionen*.

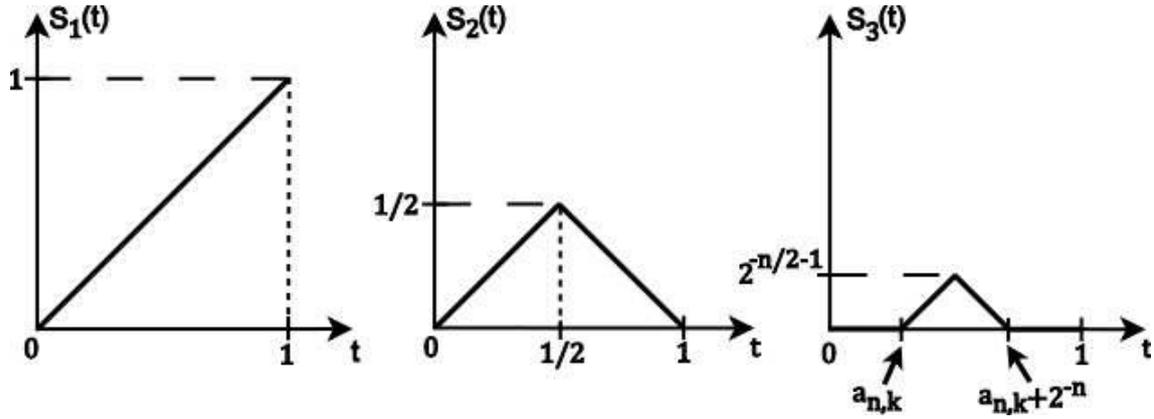


Abbildung 3.2: Schauder-Funktionen

Lemma 3.2.2

Es gilt:

1. $S_n(t) \geq 0, t \in [0, 1], n \in \mathbb{N}$,
2. $\sum_{k=1}^{2^n} S_{2^{n+k}}(t) \leq \frac{1}{2} 2^{-\frac{n}{2}}, t \in [0, 1], n \in \mathbb{N}$,
3. Sei $\{a_n\}_{n \in \mathbb{N}}$ eine Folge von reellen Zahlen mit $a_n = O(n^\varepsilon)$, $\varepsilon < \frac{1}{2}$, $n \rightarrow \infty$. Dann konvergiert die Reihe $\sum_{n=1}^{\infty} a_n S_n(t)$ absolut und gleichmäßig in $t \in [0, 1]$ und ist folglich eine stetige Funktion auf $[0, 1]$.

Beweis 1. folgt unmittelbar aus Definition 3.2.2.

2. folgt aus der Tatsache, dass die Funktionen $S_{2^{n+k}}$ für $k = 1, \dots, 2^n$ disjunkte Träger haben und $S_{2^{n+k}}(t) \leq S_{2^{n+k}}(\frac{2k-1}{2^{n+1}}) = 2^{-\frac{n}{2}-1}, t \in [0, 1]$.
3. Es genügt zu zeigen, dass $R_m = \sup_{t \in [0,1]} \sum_{k > 2^m} |a_k| S_k(t) \xrightarrow{m \rightarrow \infty} 0$. Für jedes $k \in \mathbb{N}$ und ein $c > 0$ gilt $|a_k| \leq ck^\varepsilon$. Deshalb gilt für alle $t \in [0, 1], n \in \mathbb{N}$

$$\sum_{2^n < k \leq 2^{n+1}} |a_k| S_k(t) \leq c \cdot 2^{(n+1)\varepsilon} \cdot \sum_{2^n < k \leq 2^{n+1}} S_k(t) \leq c \cdot 2^{(n+1)\varepsilon} \cdot 2^{-\frac{n}{2}-1} \leq c \cdot 2^{\varepsilon - n(\frac{1}{2} - \varepsilon)}.$$

$$\text{Da } \varepsilon < \frac{1}{2}, \text{ gilt } R_m \leq c \cdot 2^\varepsilon \sum_{n \geq m} 2^{-n(\frac{1}{2} - \varepsilon)} \xrightarrow{m \rightarrow \infty} 0.$$

□

Lemma 3.2.3Sei $\{Y_n\}_{n \in \mathbb{N}}$ eine Folge von (nicht unbedingt unabhängigen) Zufallsvariablen definiert auf $(\Omega, \mathcal{A}, \mathbb{P})$, $Y_n \sim \mathcal{N}(0, 1)$, $n \in \mathbb{N}$. Dann gilt $|Y_n| = O((\log n)^{\frac{1}{2}})$, $n \rightarrow \infty$.**Beweis** Zu zeigen ist, dass für $c > \sqrt{2}$ und fast allen $\omega \in \Omega$ ein $n_0 = n_0(\omega, c) \in \mathbb{N}$ existiert, so dass $|Y_n| \leq c(\log n)^{\frac{1}{2}}$ für $n \geq n_0$. Falls $Y \sim \mathcal{N}(0, 1)$, $x > 0$, dann gilt

$$\begin{aligned} \mathbb{P}(Y > x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_x^\infty \left(-\frac{1}{y}\right) d\left(e^{-\frac{y^2}{2}}\right) \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} e^{-\frac{x^2}{2}} - \int_x^\infty e^{-\frac{y^2}{2}} \frac{1}{y^2} dy\right) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}. \end{aligned}$$

(Man kann auch zeigen, dass $\bar{\Phi}(x) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{x^2}{2}}$, $x \rightarrow \infty$.) Daher gilt für $c > \sqrt{2}$

$$\sum_{n \geq 2} \mathbb{P}(|Y_n| > c(\log n)^{\frac{1}{2}}) \leq c^{-1} \frac{2}{\sqrt{2\pi}} \sum_{n \geq 2} (\log n)^{-\frac{1}{2}} e^{-\frac{c^2}{2} \log n} = \frac{c^{-1} \sqrt{2}}{\sqrt{\pi}} \sum_{n \geq 2} (\log n)^{-\frac{1}{2}} n^{-\frac{c^2}{2}} < \infty.$$

Nach dem Lemma von Borel-Cantelli (vgl. WR-Skript, Lemma 2.2.1) gilt $\mathbb{P}(\cap_n \cup_{k \geq n} A_k) = 0$, falls $\sum_k \mathbb{P}(A_k) < \infty$ mit $A_k = \{|Y_k| > e \cdot (\log k)^{\frac{1}{2}}\}$, $k \in \mathbb{N}$. Daher tritt A_k in unendlicher Anzahl nur mit Wahrscheinlichkeit 0, mit $|Y_n| \leq c(\log n)^{\frac{1}{2}}$ für $n \geq n_0$. \square

3.2.2 Wiener-Prozess mit f.s. stetigen Pfaden

Lemma 3.2.4

Sei $\{Y_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen $\mathcal{N}(0, 1)$ -verteilten Zufallsvariablen. Seien $\{a_n\}_{n \in \mathbb{N}}$ und $\{b_n\}_{n \in \mathbb{N}}$ Folgen von Zahlen mit $\sum_{k=1}^{2^m} |a_{2^m+k}| \leq 2^{-\frac{m}{2}}$, $\sum_{k=1}^{2^m} |b_{2^m+k}| \leq 2^{-\frac{m}{2}}$, $m \in \mathbb{N}$. Dann existieren f.s. die Grenzwerte $U = \sum_{n=1}^{\infty} a_n Y_n$ und $V = \sum_{n=1}^{\infty} b_n Y_n$, $U \sim \mathcal{N}(0, \sum_{n=1}^{\infty} a_n^2)$, $V \sim \mathcal{N}(0, \sum_{n=1}^{\infty} b_n^2)$, wobei $\text{cov}(U, V) = \sum_{n=1}^{\infty} a_n b_n$. U und V sind genau dann unabhängig, wenn $\text{cov}(U, V) = 0$.

Beweis Aus Lemma 3.2.2 und 3.2.3 ergibt sich f.s. die Existenz der Grenzwerte U und V (ersetze dafür a_n durch Y_n und S_n durch z.B. b_n in Lemma 3.2.2). Aus der Faltungsstabilität der Normalverteilung folgt für $U^{(m)} = \sum_{n=1}^m a_n Y_n$, $V^{(m)} = \sum_{n=1}^m b_n Y_n$, dass $U^{(m)} \sim \mathcal{N}(0, \sum_{n=1}^m a_n^2)$, $V^{(m)} \sim \mathcal{N}(0, \sum_{n=1}^m b_n^2)$. Da $U^{(m)} \xrightarrow{d} U$, $V^{(m)} \xrightarrow{d} V$ folgt somit $U \sim \mathcal{N}(0, \sum_{n=1}^{\infty} a_n^2)$, $V \sim \mathcal{N}(0, \sum_{n=1}^{\infty} b_n^2)$. Weiterhin, gilt

$$\begin{aligned} \text{cov}(U, V) &= \lim_{m \rightarrow \infty} \text{cov}(U^{(m)}, V^{(m)}) \\ &= \lim_{m \rightarrow \infty} \sum_{i,j=1}^m a_i b_j \text{cov}(Y_i, Y_j) \\ &= \lim_{m \rightarrow \infty} \sum_{i=1}^m a_i b_i = \sum_{i=1}^{\infty} a_i b_i, \end{aligned}$$

nach dem Satz von Lebesgue über die majorisierte Konvergenz, denn nach Lemma 3.2.3 gilt $|Y_n| \leq c \underbrace{(\log n)^{\frac{1}{2}}}_{\leq cn^\varepsilon, \varepsilon < \frac{1}{2}}$, für $n \geq N_0$, und die majorisierende Reihe konvergiert nach Lemma 3.2.2:

$$\sum_{n,k=2^m}^{2^{m+1}} a_n b_k Y_n Y_k \stackrel{f.s.}{\leq} \sum_{n,k=2^m}^{2^{m+1}} a_n b_k c^2 n^\varepsilon k^\varepsilon \leq 2^{2\varepsilon(m+1)} \cdot 2^{-\frac{m}{2}} \cdot 2^{-\frac{m}{2}} = 2^{-(1-2\varepsilon)m}, \quad 1 - 2\varepsilon > 0.$$

Für ausreichend großes m gilt $\sum_{n,k=m}^{\infty} a_n b_k Y_n Y_k \leq \sum_{j=m}^{\infty} 2^{-(1-2\varepsilon)j} < \infty$, und diese Reihe konvergiert f.s.

Zeigen wir nun

$$\text{cov}(U, V) = 0 \iff U \text{ und } V \text{ unabhängig}$$

Aus der Unabhängigkeit folgt immer die Unkorreliertheit von Zufallsvariablen. Zeigen wir hier

das Gegenteil. Aus $(U^{(m)}, V^{(m)}) \xrightarrow{d} (U, V)$ folgt $\varphi_{(U^{(m)}, V^{(m)})} \xrightarrow{m \rightarrow \infty} \varphi_{(U, V)}$, also

$$\begin{aligned}
\varphi_{(U^{(m)}, V^{(m)})}(s, t) &= \lim_{m \rightarrow \infty} \mathbb{E} \exp\left\{i\left(t \sum_{k=1}^m a_k Y_k + s \sum_{n=1}^m b_n Y_n\right)\right\} \\
&= \lim_{m \rightarrow \infty} \mathbb{E} \exp\left\{i \sum_{k=1}^m (ta_k + sb_k) Y_k\right\} = \lim_{m \rightarrow \infty} \prod_{k=1}^m \mathbb{E} \exp\{i(ta_k + sb_k) Y_k\} \\
&= \lim_{m \rightarrow \infty} \prod_{k=1}^m \exp\left\{-\frac{(ta_k + sb_k)^2}{2}\right\} = \exp\left\{-\sum_{k=1}^{\infty} \frac{(ta_k + sb_k)^2}{2}\right\} \\
&= \exp\left\{-\frac{t^2}{2} \sum_{k=1}^{\infty} a_k^2\right\} \exp\left\{ts \underbrace{\sum_{k=1}^{\infty} a_k b_k}_{\text{cov}(U, V)=0}\right\} \exp\left\{-\frac{s^2}{2} \sum_{k=1}^{\infty} b_k^2\right\} = \varphi_U(t) \varphi_V(s),
\end{aligned}$$

$s, t \in \mathbb{R}$. Daher sind U und V unabhängig, falls $\text{cov}(U, V) = 0$. \square

Theorem 3.2.1

Sei $\{Y_n, n \in \mathbb{N}\}$ eine Folge von u.i.v. Zufallsvariablen, die $\mathcal{N}(0, 1)$ -verteilt sind, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann gibt es einen Teil-Wahrscheinlichkeitsraum $(\Omega_0, \mathcal{A}_0, \mathbb{P})$ von $(\Omega, \mathcal{A}, \mathbb{P})$ und einen stochastischen Prozess $X = \{X(t), t \in [0, 1]\}$ darauf, so dass $X(t) = \sum_{n=1}^{\infty} Y_n S_n(t)$, $t \in [0, 1]$, und $X \stackrel{d}{=} W$. Hierbei ist $\{S_n\}_{n \in \mathbb{N}}$ die Familie der Schauder-Funktionen.

Beweis Nach Lemma 3.2.2, 2) erfüllen die Koeffizienten $S_n(t)$ für jedes $t \in [0, 1]$ die Bedingungen des Lemmas 3.2.4. Darüberhinaus existiert nach Lemma 3.2.3 eine Teilmenge $\Omega_0 \subset \Omega$, $\Omega_0 \in \mathcal{A}$ mit $\mathbb{P}(\Omega_0) = 1$, so dass für jedes $\omega \in \Omega_0$ die Relation $|Y_n(\omega)| = O(\sqrt{\log n})$, $n \rightarrow \infty$, gilt. Sei $\mathcal{A}_0 = \mathcal{A} \cap \Omega_0$. Schränken wir den Wahrscheinlichkeitsraum auf $(\Omega_0, \mathcal{A}_0, \mathbb{P})$ ein. Dann ist die Bedingung $a_n = Y_n(\omega) = O(n^\varepsilon)$, $\varepsilon < \frac{1}{2}$, erfüllt, weil $\sqrt{\log n} < n^\varepsilon$ für ausreichend große n , und nach Lemma 3.2.2, 3) konvergiert die Reihe $\sum_{n=1}^{\infty} Y_n(\omega) S_n(t)$ absolut und gleichmäßig in $t \in [0, 1]$ gegen die Funktion $X(\omega, t)$, $\omega \in \Omega_0$, die eine stetige Funktion in t für jedes $\omega \in \Omega_0$ ist. $X(\cdot, t)$ ist eine Zufallsvariable, weil nach Lemma 3.2.4 die Konvergenz dieser Reihe fast sicher gilt. Weiterhin, gilt $X(t) \sim \mathcal{N}(0, \sum_{n=1}^{\infty} S_n^2(t))$, $t \in [0, 1]$.

Zeigen wir, dass der so definierte stochastische Prozess auf $(\Omega_0, \mathcal{A}_0, \mathbb{P})$ ein Wiener-Prozess ist. Dazu prüfen wir die Bedingungen der Definition 3.1.1. Betrachten wir beliebige Zeitpunkte

$0 \leq t_1 < t_2, t_3 < t_4 \leq 1$ und berechnen wir

$$\begin{aligned}
\text{cov}(X(t_2) - X(t_1), X(t_4) - X(t_3)) &= \text{cov}\left(\sum_{n=1}^{\infty} Y_n(S_n(t_2) - S_n(t_1)), \sum_{n=1}^{\infty} Y_n(S_n(t_4) - S_n(t_3))\right) \\
&= \sum_{n=1}^{\infty} (S_n(t_2) - S_n(t_1))(S_n(t_4) - S_n(t_3)) \\
&= \sum_{n=1}^{\infty} (\langle H_n, \mathbf{1}_{[0,t_2]} \rangle - \langle H_n, \mathbf{1}_{[0,t_1]} \rangle) \times \\
&\quad (\langle H_n, \mathbf{1}_{[0,t_4]} \rangle - \langle H_n, \mathbf{1}_{[0,t_3]} \rangle) \\
&= \sum_{n=1}^{\infty} \langle H_n, \mathbf{1}_{[0,t_2]} - \mathbf{1}_{[0,t_1]} \rangle \langle H_n, \mathbf{1}_{[0,t_4]} - \mathbf{1}_{[0,t_3]} \rangle \\
&= \langle \mathbf{1}_{[0,t_2]} - \mathbf{1}_{[0,t_1]}, \mathbf{1}_{[0,t_4]} - \mathbf{1}_{[0,t_3]} \rangle \\
&= \langle \mathbf{1}_{[0,t_2]}, \mathbf{1}_{[0,t_4]} \rangle - \langle \mathbf{1}_{[0,t_1]}, \mathbf{1}_{[0,t_4]} \rangle \\
&\quad - \langle \mathbf{1}_{[0,t_2]}, \mathbf{1}_{[0,t_3]} \rangle + \langle \mathbf{1}_{[0,t_1]}, \mathbf{1}_{[0,t_3]} \rangle \\
&= \min\{t_2, t_4\} - \min\{t_1, t_4\} - \min\{t_2, t_3\} + \min\{t_1, t_3\},
\end{aligned}$$

weil $\langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \rangle = \int_0^{\min\{s,t\}} du = \min\{s,t\}$, $s, t \in [0, 1]$. Falls $0 \leq t_1 < t_2 \leq t_3 < t_4 < 1$, dann gilt $\text{cov}(X(t_2) - X(t_1), X(t_4) - X(t_3)) = t_2 - t_1 - t_2 + t_1 = 0$, also sind die Zuwächse von X (nach Lemma 3.2.4) unkorreliert. Weiterhin gilt $X(0) \sim \mathcal{N}(0, \sum_{n=1}^{\infty} S_n^2(0)) = \mathcal{N}(0, 0)$, daher $X(0) \stackrel{f.s.}{=} 0$. Daraus folgt für $t_1 = 0, t_2 = t, t_3 = 0, t_4 = t$, dass $\text{var}(X(t)) = t, t \in [0, 1]$, und für $t_1 = t_3 = s, t_2 = t_4 = t$, dass $\text{var}(X(t) - X(s)) = t - s - s + s = t - s, 0 \leq s < t \leq 1$. Somit gilt $X(t) - X(s) \sim \mathcal{N}(0, t - s)$, und nach Definition 3.1.1 gilt $X \stackrel{d}{=} W$. \square

Bemerkung 3.2.1 1. Theorem 3.2.1 bildet eine Grundlage für die approximative Simulation der Pfade einer Brownschen Bewegung durch die Teilsummen $X^{(n)}(t) = \sum_{k=1}^n Y_k S_k(t)$, $t \in [0, 1]$, für ausreichend großes $n \in \mathbb{N}$.

2. Die Konstruktion in Theorem 3.2.1 kann verwendet werden, um den Wiener-Prozess mit stetigen Pfaden auf einem Intervall $[0, t_0]$, für beliebiges $t_0 > 0$ zu erzeugen. Falls $W = \{W(t), t \in [0, 1]\}$ ein Wiener-Prozess auf $[0, 1]$ ist, dann ist $Y = \{Y(t), t \in [0, t_0]\}$ mit $Y(t) = \sqrt{t_0} W(\frac{t}{t_0})$, $t \in [0, t_0]$, ein Wiener-Prozess auf $[0, t_0]$.

Aufgabe 3.2.1

Beweisen Sie es.

3. Der Wiener-Prozess W mit stetigen Pfaden auf \mathbb{R}_+ kann wie folgt explizit konstruiert werden. Seien $W^{(n)} = \{W^{(n)}(t), t \in [0, 1]\}$ unabhängige Kopien des Wiener-Prozesses wie in Theorem 3.2.1. Definiere $W(t) = \sum_{n=1}^{\infty} \mathbf{1}(t \in [n-1, n]) [\sum_{k=1}^{n-1} W^{(k)}(1) - W^{(n)}(t - (n-1))]$, $t \geq 0$, also,

$$W(t) = \begin{cases} W^{(1)}(t), & t \in [0, 1], \\ W^{(1)}(1) + W^{(2)}(t-1), & t \in [1, 2], \\ W^{(1)}(1) + W^{(2)}(1) + W^{(3)}(t-2), & t \in [2, 3], \\ \text{usw.} \end{cases}$$

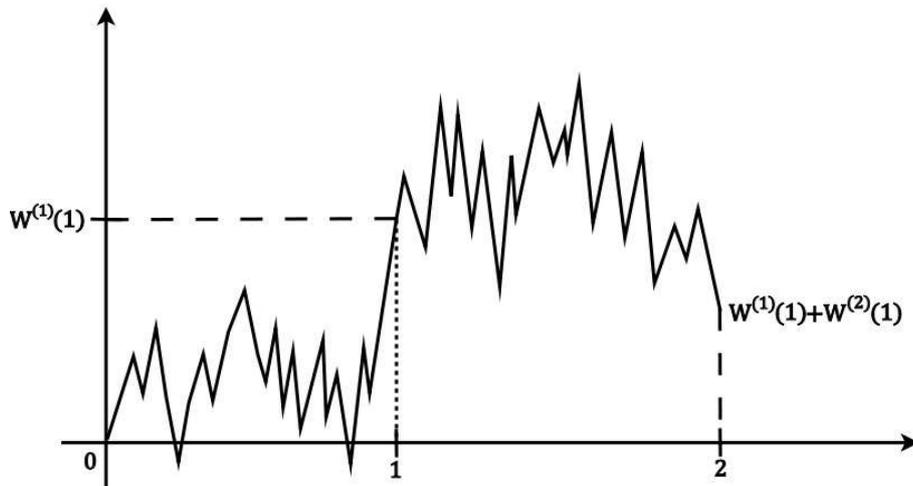


Abbildung 3.3:

Aufgabe 3.2.2

Zeigen Sie, dass der so eingeführte stochastische Prozess $W = \{W(t), t \geq 0\}$ ein Wiener-Prozess auf \mathbb{R}_+ ist.

3.3 Verteilungs- und Pfadigenschaften vom Wiener-Prozess**3.3.1 Verteilung des Maximums****Theorem 3.3.1**

Sei $W = \{W(t), t \in [0, 1]\}$ ein Wiener-Prozess über einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Dann gilt:

$$P \left(\max_{t \in [0, 1]} W(t) > x \right) \leq \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy, \quad (3.3.1)$$

für alle $x \geq 0$.

Die in 3.3.1 gegebene Abbildung $\max_{t \in [0, 1]} W(t) : \Omega \rightarrow [0, \infty)$ ist eine wohldefinierte Zufallsvariable, denn es gilt: $\max_{t \in [0, 1]} W(t, \omega) = \lim_{n \rightarrow \infty} \max_{i=1, \dots, k} W(\frac{i}{k}, \omega)$ für alle $\omega \in \Omega$, weil die Trajektorien von $\{W(t), t \in [0, 1]\}$ stetig sind. Aus 3.3.1 folgt, dass $\max_{t \in [0, 1]} W(t)$ einen exponentiell beschränkten Tail hat: somit hat $\max_{t \in [0, 1]} W(t)$ endliche k -te Momente.

Hilfsmittel für Beweis von Theorem 3.3.1

Sei $\{W(t), t \in [0, 1]\}$ ein Wiener-Prozess und Z_1, Z_2, \dots eine Folge von unabhängigen Zufallsvariablen mit $P(Z_i = 1) = P(Z_i = -1) = \frac{1}{2}$ für alle $i \geq 1$. Für jedes $n \in \mathbb{N}$ definieren wir $\{\tilde{W}^n(t), t \in [0, 1]\}$ durch $\tilde{W}^n(t) = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} + (nt - \lfloor nt \rfloor) \frac{Z_{\lfloor nt \rfloor + 1}}{\sqrt{n}}$, wobei $S_i = Z_1 + \dots + Z_i$, $i \geq 1$, $S_0 = 0$.

Lemma 3.3.1

Für jedes $k \geq 1$ und beliebige $t_1, \dots, t_k \in [0, 1]$ gilt:

$$\left(\tilde{W}^n(t_1), \dots, \tilde{W}^n(t_k) \right)^\top \xrightarrow{d} \left(W(t_1), \dots, W(t_k) \right)^\top.$$

Beweis Spezialfall $k = 2$ (für $k > 2$ verläuft der Beweis analog). Sei $t_1 < t_2$. Für alle $s_1, s_2 \in \mathbb{R}$ gilt:

$$\begin{aligned} s_1 \tilde{W}^{(n)}(t_1) + s_2 \tilde{W}^{(n)}(t_2) &= (s_1 + s_2) \frac{S_{\lfloor nt_1 \rfloor}}{\sqrt{n}} + s_2 \frac{(S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor + 1})}{\sqrt{n}} \\ &\quad + Z_{\lfloor nt_1 \rfloor + 1} \left((nt_1 - \lfloor nt_1 \rfloor) \frac{s_1}{\sqrt{n}} + \frac{s_2}{\sqrt{n}} \right) \\ &\quad + Z_{\lfloor nt_2 \rfloor + 1} (nt_2 - \lfloor nt_2 \rfloor) \frac{s_2}{\sqrt{n}}. \end{aligned}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E} e^{i(s_1 \tilde{W}^{(n)}(t_1) + s_2 \tilde{W}^{(n)}(t_2))} &= \lim_{n \rightarrow \infty} \mathbb{E} e^{i \frac{s_1 + s_2}{\sqrt{n}} S_{\lfloor nt_1 \rfloor}} \mathbb{E} e^{i \frac{s_2}{\sqrt{n}} (S_{\lfloor nt_2 \rfloor} - S_{\lfloor nt_1 \rfloor + 1})} \\ &= \left| \mathbb{E} e^{i \frac{s_1 + s_2}{\sqrt{n}} S_{\lfloor nt_1 \rfloor}} = \varphi_{S_{\lfloor nt_1 \rfloor}} \left(\frac{s_1 + s_2}{\sqrt{n}} \right) = \left(\varphi_{Z_1} \left(\frac{s_1 + s_2}{\sqrt{n}} \right) \right)^{\lfloor nt_1 \rfloor} \right| \\ &= \lim_{n \rightarrow \infty} \left(\varphi_{Z_1} \left(\frac{s_1 + s_2}{\sqrt{n}} \right) \right)^{\lfloor nt_1 \rfloor} \left(\varphi_{Z_1} \left(\frac{s_2}{\sqrt{n}} \right) \right)^{\lfloor nt_2 \rfloor - \lfloor nt_1 \rfloor - 1} \\ &= \left| \lim_{n \rightarrow \infty} \varphi^n \left(\frac{s}{\sqrt{n}} \right) = e^{-\frac{s^2}{2}} \right| \\ &= e^{-\frac{(s_1^2 t_1 + 2s_1 s_2 \min\{t_1, t_2\} + s_2^2 t_2)}{2}} \\ &= \varphi_{(W(t_1), W(t_2))}(s_1, s_2), \end{aligned}$$

wobei $\varphi_{(W(t_1), W(t_2))}$ die charakteristische Funktion von $(W(t_1), W(t_2))$ ist. \square

Lemma 3.3.2

Sei $\tilde{W}^{(n)} = \max_{t \in [0, 1]} \tilde{W}^{(n)}(t)$. Dann gilt:

$$\tilde{W}^{(n)} = \frac{1}{\sqrt{n}} \max_{k=1, \dots, n} S_k, \quad \text{für alle } n = 1, 2, \dots$$

und

$$\lim_{n \rightarrow \infty} \mathbb{P}(\tilde{W}^{(n)} \leq x) = \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{y^2}{2}} dy, \quad \text{für alle } x \geq 0.$$

Ohne Beweis

Beweis von Theorem 3.3.1

Aus Lemma 3.3.1 folgt für $x \geq 0$, $k \geq 1$ und $t_1, \dots, t_n \in [0, 1]$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left(\max_{t \in \{t_1, \dots, t_k\}} \tilde{W}^{(n)}(t) > x \right) &= \mathbb{P} \left(\max_{t \in \{t_1, \dots, t_k\}} W(t) > x \right) \\ \Rightarrow \mathbb{P} \left(\max_{t \in [0, 1]} \tilde{W}^{(n)}(t) > x \right) &\geq \mathbb{P} \left(\max_{t \in \{t_1, \dots, t_k\}} W(t) > x \right). \end{aligned}$$

Mit $(t_1, \dots, t_k)^\top = \left(\frac{1}{k}, \dots, \frac{k}{k} \right)^\top$ und $\max_{t \in [0, 1]} W(t, \omega) = \lim_{k \rightarrow \infty} \max_{i=1, \dots, k} W\left(\frac{1}{k}, \omega\right)$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\max_{t \in [0, 1]} \tilde{W}^{(n)}(t) > x \right) \geq \mathbb{P} \left(\max_{t \in [0, 1]} W(t) > x \right).$$

Aus Lemma 3.3.2 folgt die Behauptung. \square

Korollar 3.3.1

Sei $\{W(t), t \in [0, 1]\}$ ein Wiener-Prozess. Dann gilt:

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{W(t)}{t} = 0\right) = 1.$$

Beweis

$$\begin{aligned} \left| \frac{W(t)}{t} - \frac{W(n)}{n} \right| &\leq \left| \frac{W(t)}{t} - \frac{W(n)}{t} \right| + \left| \frac{W(n)}{t} - \frac{W(n)}{n} \right| \\ &\leq |W(n)| \left| \frac{1}{t} - \frac{1}{n} \right| + \frac{1}{n} \sup_{t \in [n, n+1]} |W(t) - W(n)| \\ &\leq \frac{2}{n} |W(n)| - \frac{Z(n)}{n}, \end{aligned}$$

wobei $Z(n) = \sup_{t \in [0, 1]} |W(n+t) - W(n)|$, $t \in [n, n+1]$. Es gilt

$$\frac{2}{n} |W(n)| = \frac{2}{n} \left| \sum_{i=1}^{\infty} (W(i) - W(i-1)) \right| \xrightarrow{f.s.} 2 |EW(1)| = 0.$$

Wir zeigen, dass $EZ(1) < \infty$.

$$\mathbb{P}(Z(1) > x) \leq \mathbb{P}\left(\max_{t \in [0, 1]} W(t) > x\right) + \mathbb{P}\left(\max_{t \in [0, 1]} (-W(t)) > x\right) = 2\mathbb{P}\left(\max_{t \in [0, 1]} W(t) > x\right),$$

weil $\{-W(t), t \in [0, 1]\}$ auch ein Wiener-Prozess ist. Es gilt

$$\mathbb{P}(Z(1) > x) \leq 2\sqrt{\frac{2}{\pi}} \int_x^{\infty} e^{-\frac{y^2}{2}} dy \quad \text{und somit} \quad \frac{Z(n)}{n} \xrightarrow{f.s.} 0 \quad \text{für } n \rightarrow \infty.$$

Daraus folgt $\frac{W(t)}{t} \xrightarrow{f.s.} 0$ für $t \rightarrow \infty$. □

3.3.2 Invarianzeigenschaften

Bestimmte Transformationen des Wiener-Prozesses führen wieder zu einem Wiener-Prozess.

Theorem 3.3.2

Sei $\{W(t), t \geq 0\}$ ein Wiener-Prozess. Dann sind die stochastischen Prozesse $\{Y^{(i)}(t), t \geq 0\}$, $i = 1, \dots, 4$, mit

$$\begin{aligned} Y^{(1)}(t) &= -W(t), && \text{(Symmetrie)} \\ Y^{(2)}(t) &= W(t + t_0) - W(t_0) \quad \text{für ein } t_0 > 0, && \text{(Verschiebung des Nullpunktes)} \\ Y^{(3)}(t) &= \sqrt{c}W\left(\frac{t}{c}\right) \quad \text{für ein } c > 0, && \text{(Skalierung)} \\ Y^{(4)}(t) &= \begin{cases} tW\left(\frac{1}{t}\right), & t > 0, \\ 0, & t = 0. \end{cases} && \text{(Spiegelung bei } t = 0) \end{aligned}$$

ebenfalls Wiener-Prozesse.

Beweis 1. $Y^{(i)}$, $i = 1, \dots, 4$, haben unabhängige Zuwächse mit $Y^{(i)}(t_2) - Y^{(i)}(t_1) \sim \mathcal{N}(0, t_2 - t_1)$.

2. $Y^{(i)}(0) = 0, i = 1, \dots, 4.$
3. $Y^{(i)}, i = 1, \dots, 3,$ haben stetige Trajektorien. $\{Y^{(i)}(t), t \geq 0\}$ hat stetige Trajektorien für $t > 0.$
4. Es ist zu zeigen, dass $\lim_{t \rightarrow 0} tW(\frac{1}{t}) = 0.$
 $\lim_{t \rightarrow 0} tW(\frac{1}{t}) = \lim_{t \rightarrow \infty} \frac{W(t)}{t} \stackrel{f.s.}{=} 0$ wegen Korollar 3.3.1.

□

Korollar 3.3.2

Sei $\{W(t), t \geq 0\}$ ein Wiener-Prozess. Dann gilt:

$$\mathbb{P} \left(\sup_{t \geq 0} W(t) = \infty \right) = \mathbb{P} \left(\inf_{t \geq 0} W(t) = -\infty \right) = 1.$$

Beweis Für $x, c > 0$ gilt:

$$\begin{aligned} \mathbb{P} \left(\sup_{t \geq 0} W(t) > x \right) &= \mathbb{P} \left(\sup_{t \geq 0} W \left(\frac{t}{c} \right) > \frac{x}{\sqrt{c}} \right) = \mathbb{P} \left(\sup_{t \geq 0} W(t) > \frac{x}{\sqrt{c}} \right) \\ \Rightarrow \mathbb{P} \left(\left\{ \sup_{t \geq 0} W(t) = 0 \right\} \cup \left\{ \sup_{t \geq 0} W(t) = \infty \right\} \right) &= \mathbb{P}(\sup_{t \geq 0} W(t) = 0) + \mathbb{P}(\sup_{t \geq 0} W(t) = \infty) = 1. \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} \mathbb{P} \left(\sup_{t \geq 0} W(t) = 0 \right) &= \mathbb{P} \left(\sup_{t \geq 0} W(t) \leq 0 \right) \leq \mathbb{P} \left(W(t) \leq 0, \sup_{t \geq 1} W(t) \leq 0 \right) \\ &= \mathbb{P} \left(W(1) \leq 0, \sup_{t \geq 1} (W(t) - W(1)) \leq -W(1) \right) \\ &= \int_{-\infty}^0 \mathbb{P} \left(\sup_{t \geq 1} W(t) - W(1) \leq -W(1) \mid W(1) = x \right) \mathbb{P}(W(1) \in dx) \\ &= \int_{-\infty}^0 \mathbb{P} \left(\sup_{t \geq 0} (W(t) - W(1)) \leq -x \mid W(1) = x \right) \mathbb{P}(W(1) \in dx) \\ &= \int_{-\infty}^0 \mathbb{P} \left(\sup_{t \geq 0} W(t) = 0 \right) \mathbb{P}(W(1) \in dx) \\ &= \mathbb{P} \left(\sup_{t \geq 0} W(t) = 0 \right) \frac{1}{2}, \end{aligned}$$

also $\mathbb{P} \left(\sup_{t \geq 0} W(t) = 0 \right) = 0$ und somit $\mathbb{P} \left(\sup_{t \geq 0} W(t) = \infty \right) = 1.$

Analog kann man zeigen, dass $\mathbb{P} \left(\inf_{t \geq 0} W(t) = -\infty \right) = 1.$

□

Bemerkung 3.3.1

$\mathbb{P} \left(\sup_{t \geq 0} X(t) = \infty, \inf_{t \geq 0} X(t) = -\infty \right) = 1$ bedeutet, dass die Trajektorien von W unendlich oft zwischen positiven und negativen Werten auf $[0, \infty)$ oszillieren.

Korollar 3.3.3

Sei $\{W(t), t \geq 0\}$ ein Wiener-Prozess. Dann gilt

$$\mathbb{P}(\omega \in \Omega : W(\omega) \text{ ist nirgendwo differenzierbar in } [0, \infty)) = 1.$$

Beweis

$$\begin{aligned} & \{\omega \in \Omega : W(\omega) \text{ ist nirgendwo differenzierbar in } [0, \infty)\} \\ &= \bigcap_{n=0}^{\infty} \{\omega \in \Omega : W(\omega) \text{ ist nirgendwo differenzierbar in } [n, n+1)\}. \end{aligned}$$

Es genügt zu zeigen, dass $\mathbb{P}(\omega \in \Omega : W(\omega) \text{ differenzierbar für ein } t_0 = t_0(\omega) \in [0, 1]) = 0$.
Definiere die Menge

$$A_{nm} = \left\{ \omega \in \Omega : \text{es gibt ein } t_0 = t_0(\omega) \in [0, 1] \text{ mit } |X(t_0\omega + h, \omega) - W(t_0(\omega), \omega)| \leq mh, \forall h \in \left[0, \frac{4}{k}\right] \right\}.$$

Dann gilt

$$\{\omega \in \Omega : W(\omega) \text{ differenzierbar für ein } t_0 = t_0(\omega)\} = \bigcup_{m \geq 1} \bigcup_{n \geq 1} A_{nm}.$$

Zu zeigen bleibt $\mathbb{P}(\bigcup_{m \geq 1} \bigcup_{n \geq 1} A_{nm}) = 0$.

Sei $k_0(\omega) = \min_{k=1,2,\dots} \left\{ \frac{k}{n} \geq t_0(\omega) \right\}$. Dann gilt für $\omega \in A_{nm}$ und $j = 0, 1, 2$

$$\begin{aligned} \left| W\left(\frac{k_0(\omega) + j + 1}{n}, \omega\right) - W\left(\frac{k_0(\omega) + j}{n}, \omega\right) \right| &\leq \left| W\left(\frac{k_0(\omega) + j + 1}{n}, \omega\right) - W(t_0(\omega), \omega) \right| \\ &\quad + \left| W\left(\frac{k_0(\omega) + j}{n}, \omega\right) - W(t_1(\omega), \omega) \right| \\ &\leq \frac{8m}{n}. \end{aligned}$$

Sei $\Delta_0(k) = W\left(\frac{k+1}{n}\right) - W\left(\frac{k}{n}\right)$. Dann gilt

$$\begin{aligned} \mathbb{P}(A_{nm}) &\leq \mathbb{P}\left(\bigcup_{k=0}^n \bigcup_{j=0}^2 |\Delta_n(k+j)| \leq \frac{8m}{n}\right) \\ &\leq \sum_{k=0}^n \mathbb{P}\left(\bigcap_{j=0}^2 |\Delta_n(k+j)| \leq \frac{8m}{n}\right) = \mathbb{P}\left(|\Delta_n(0)| \leq \frac{8m}{n}\right) \\ &\leq (n+1) \left(\frac{16m}{\sqrt{2\pi n}}\right)^3 \rightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

und weil $A_{nm} \subset A_{n+1,m}$ gilt, folgt $\mathbb{P}(A_{nm}) = 0$. □

Korollar 3.3.4

Mit Wahrscheinlichkeit 1 gilt:

$$\sup_{n \geq 1} \sup_{0 \leq t_0 < \dots < t_n \leq 1} \sum_{i=1}^n |W(t_i) - W(t_{i-1})| = \infty,$$

d.h. $\{W(t), t \in [0, 1]\}$ hat f.s. Trajektorien mit unbeschränkter Variation.

Beweis Weil jede stetige Funktion $g : [0, 1] \rightarrow \mathbb{R}$ mit beschränkter Variation fast überall differenzierbar ist, ergibt sich die Behauptung aus Korollar 3.3.3.

Alternativer Beweis

Es genügt zu zeigen, dass $\lim_{n \rightarrow \infty} \sum_{i=1}^{2^n} \left| W\left(\frac{it}{2^n}\right) - W\left(\frac{(i-1)t}{2^n}\right) \right| = \infty$.

Sei $Z_n = \sum_{i=1}^{2^n} \left(W\left(\frac{it}{2^n}\right) - W\left(\frac{(i-1)t}{2^n}\right) \right)^2 - t$. Daraus folgt $\mathbb{E}Z_n = 0$ und $\mathbb{E}Z_n^2 = t^2 2^{-n+1}$ und mit der Tschebyschev-Ungleichung

$$\mathbb{P}(|Z_n| < \varepsilon) \leq \frac{\mathbb{E}Z_n^2}{\varepsilon^2} = \left(\frac{t}{\varepsilon}\right)^2 2^{-n+1}, \quad \text{d.h.} \quad \sum_{i=1}^{\infty} \mathbb{P}(|Z_n| > \varepsilon) \stackrel{f.s.}{=} 0.$$

Aus dem Lemma von Borel-Cantelli ergibt sich, dass $\lim_{n \rightarrow \infty} Z_n = 0$ fast sicher und somit

$$\begin{aligned} 0 \leq t &\leq \sum_{i=1}^{2^n} \left(W\left(\frac{it}{2^n}\right) - W\left(\frac{(i-1)t}{2^n}\right) \right)^2 \\ &\leq \liminf_{n \rightarrow \infty} \max_{1 \leq k \leq 2^n} \left| W\left(\frac{kt}{2^n}\right) - W\left(\frac{(k-1)t}{2^n}\right) \right| \sum_{i=1}^{2^n} \left| W\left(\frac{it}{2^n}\right) - W\left(\frac{(i-1)t}{2^n}\right) \right|. \end{aligned}$$

Daraus folgt die Behauptung, weil W stetige Trajektorien hat und deshalb

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq 2^n} \left| W\left(\frac{kt}{2^n}\right) - W\left(\frac{(k-1)t}{2^n}\right) \right| = 0.$$

□

3.4 Ergänzende Aufgaben

Aufgabe 3.4.1

Geben Sie eine intuitive (exakte!) Methode an um Trajektorien eines Wiener-Prozesses $W = \{W(t), t \in [0, 1]\}$ zu realisieren. Nutzen Sie dabei die Unabhängigkeit und die Verteilung der Zuwächse von W . Schreiben Sie zudem ein Programm in \mathbf{R} zur Simulation von Pfaden von W . Zeichnen Sie drei Pfade $t \mapsto W(t, \omega)$ für $t \in [0, 1]$ in ein gemeinsames Schaubild.

Aufgabe 3.4.2

Es sei der Wiener-Prozess $W = \{W(t), t \in [0, 1]\}$ gegeben und $L := \operatorname{argmax}_{t \in [0, 1]} W(t)$. Zeigen Sie, dass gilt:

$$\mathbb{P}(L \leq x) = \frac{2}{\pi} \arcsin \sqrt{x}, \quad x \in [0, 1].$$

Hinweis: Verwenden Sie die Beziehung $\max_{r \in [0, t]} W(r) \stackrel{d}{=} |W(t)|$.

Aufgabe 3.4.3

Zur Simulation eines Wiener-Prozesses $W = \{W(t), t \in [0, 1]\}$ können wir auch die Approximation

$$W_n(t) = \sum_{k=1}^n S_k(t) z_k$$

verwenden, wobei die $S_k(t)$, $t \in [0, 1]$, $k \geq 1$ die *Schauder-Funktionen* sind, sowie $z_k \sim \mathcal{N}(0, 1)$ u.i.v. Zufallsvariablen, und die Reihe fast sicher für alle $t \in [0, 1]$ konvergiert ($n \rightarrow \infty$).

- Zeigen Sie, dass für alle $t \in [0, 1]$ die Approximation $W_n(t)$ auch im L^2 -Sinne gegen $W(t)$ konvergiert.
- Schreiben Sie ein Programm in \mathbf{R} (alternativ: C) zur Simulation des Wiener-Prozesses $W = \{W(t), t \in [0, 1]\}$.

- c) Simulieren Sie drei Pfade $t \mapsto W(t, \omega)$ für $t \in [0, 1]$ und zeichnen Sie diese in ein gemeinsames Schaubild. Betrachten Sie hierbei die Stützstellen $t_k = \frac{k}{n}$, $k = 0, \dots, n$ mit $n = 2^8 - 1$.

Aufgabe 3.4.4

Für den Wiener-Prozess $W = \{W(t), t \geq 0\}$ definieren wir den Prozess des Maximums, welcher gegeben ist durch $M = \{M(t) := \max_{s \in [0, t]} W(s), t \geq 0\}$. Zeigen Sie, dass dann gilt:

- a) Die Dichte $f_{M(t)}$ des Maximums $M(t)$ ist gegeben durch

$$f_{M(t)}(x) = \sqrt{\frac{2}{\pi t}} \exp\left\{-\frac{x^2}{2t}\right\} \mathbf{1}\{x \geq 0\}.$$

Hinweis: Verwenden Sie die Eigenschaft $\mathbf{P}(M(t) > x) = 2\mathbf{P}(W(t) > x)$.

- b) Erwartungswert und Varianz von $M(t)$ sind gegeben durch

$$\mathbf{E}M(t) = \sqrt{\frac{2t}{\pi}}, \quad \mathbf{var} M(t) = t(1 - 2/\pi).$$

Nun definieren wir $\tau(x) := \operatorname{argmin}_{s \in \mathbb{R}} \{W(s) = x\}$ als den ersten Zeitpunkt, zu dem der Wiener-Prozess den Wert x annimmt.

- c) Bestimmen Sie die Dichte von $\tau(x)$ und zeigen Sie: $\mathbf{E}\tau(x) = \infty$.

Aufgabe 3.4.5

Sei $W = \{W(t), t \geq 0\}$ ein Wiener-Prozess. Zeigen Sie, dass die folgenden Prozesse ebenfalls Wiener-Prozesse sind:

$$W_1(t) = \begin{cases} 0, & t = 0, \\ tW(1/t), & t > 0, \end{cases} \quad W_2(t) = \sqrt{c}W(t/c), \quad c > 0.$$

Aufgabe 3.4.6

Es sei der Wiener-Prozess $W = \{W(t), t \geq 0\}$ gegeben. Die Größe $Q(a, b)$ bezeichne die Wahrscheinlichkeit, dass der Prozess die Halbgerade $y = at + b$, $t \geq 0$, $a, b > 0$ überschreitet. Beweisen Sie:

- a) $Q(a, b) = Q(b, a)$ und $Q(a, b_1 + b_2) = Q(a, b_1)Q(a, b_2)$,
 b) $Q(a, b)$ ist gegeben durch $Q(a, b) = \exp\{-2ab\}$.

4 Lévy Prozesse

4.1 Lévy-Prozesse

Definition 4.1.1

Ein stochastischer Prozess $\{X(t), t \geq 0\}$ heißt Lévy-Prozess, wenn

1. $X(0) = 0$,
2. $\{X(t)\}$ hat stationäre und unabhängige Zuwächse,
3. $\{X(t)\}$ stochastisch stetig ist, d.h. für beliebiges $\varepsilon > 0$, $t_0 \geq 0$:

$$\lim_{t \rightarrow t_0} \mathbb{P}(|X(t) - X(t_0)| > \varepsilon) = 0.$$

Beachte

- Man kann sich leicht überlegen, dass zusammengesetzte Poisson-Prozesse die 3 Bedingungen erfüllen, denn für bel. $\varepsilon > 0$ gilt

$$\mathbb{P}(|X(t) - X(t_0)| < \varepsilon) \geq \mathbb{P}(|X(t) - X(t_0)| > 0) \leq 1 - e^{-\lambda|t-t_0|} \xrightarrow{t \rightarrow t_0} 0.$$

- Ferner gilt für den Wiener-Prozess für bel. $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}(|X(t) - X(t_0)| > \varepsilon) &= \int_{\frac{\varepsilon}{\sqrt{t-t_0}}}^{\infty} \frac{2}{\pi} \exp\left(-\frac{y^2}{2(t-t_0)}\right) dy \\ &\stackrel{x=\frac{y}{\sqrt{t-t_0}}}{=} \frac{2}{\pi} \int_{\frac{\varepsilon}{\sqrt{t-t_0}}}^{\infty} e^{-\frac{x^2}{2}} dx \xrightarrow{t \rightarrow t_0} 0. \end{aligned}$$

4.1.1 Unbegrenzte Teilbarkeit

Definition 4.1.2

Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable. Dann nennt man X *unbegrenzt teilbar*, wenn für bel. $n \in \mathbb{N}$ Zufallsvariablen Y_1, Y_2, \dots, Y_n existieren mit $X \stackrel{d}{=} Y_1^{(n)} + \dots + Y_n^{(n)}$.

Theorem 4.1.1

Sei $\{X(t), t \geq 0\}$ ein Lévy-Prozess. Dann ist die Zufallsvariable $X(t)$ für jedes $t \geq 0$ unbegrenzt teilbar.

Beweis Für bel. $t \geq 0$ und $n \in \mathbb{N}$ gilt offenbar, dass

$$X(t) = X\left(\frac{t}{n}\right) + \left(X\left(\frac{2t}{n}\right) - X\left(\frac{t}{n}\right)\right) + \dots + \left(X\left(\frac{nt}{n}\right) - X\left(\frac{(n-1)t}{n}\right)\right).$$

Da $\{X(t)\}$ unabhängige stationäre Zuwächse hat, sind Summanden offenbar unabhängige identisch verteilte Zufallsvariablen. \square

Lemma 4.1.1

Die Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ ist genau dann unbegrenzt teilbar, wenn sich die charakteristische Funktion φ_X von X für jedes $n \geq 1$ darstellen läßt in der Form

$$\varphi_X(s) = (\varphi_n(s))^n \quad \text{für alle } s \in \mathbb{R},$$

wobei φ_n charakteristische Funktionen von Zufallsvariablen sind.

Beweis „ \Rightarrow “

$Y_1^{(n)}, \dots, Y_n^{(n)}$ u.i.v., $X \stackrel{d}{=} Y_1^{(n)} + \dots + Y_n^{(n)}$. Daraus folgt, dass $\varphi_X(s) = \prod_{i=1}^n \varphi_{Y_i^{(n)}}(s) = (\varphi_n(s))^n$.

“ \Leftarrow “

$\varphi_X(s) = (\varphi_n(s))^n \Rightarrow$ existiert $Y_1^{(n)}, \dots, Y_n^{(n)}$ u.i.v. mit charakteristischer Funktion φ_n und $\varphi_{Y_1, \dots, Y_n}(s) = (\varphi_n(s))^n = \varphi_X(s)$. Mit dem Eindeutigkeitssatz für charakteristische Funktionen folgt, dass $X \stackrel{d}{=} Y_1^{(n)} + \dots + Y_n^{(n)}$. \square

Lemma 4.1.2

Sei $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ eine Folge von Zufallsvariablen. Falls es eine Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ gibt, so dass $\varphi(s)$ stetig in $s = 0$ ist und $\lim_{n \rightarrow \infty} \varphi_{X_n}(s) = \varphi(s)$ für alle $s \in \mathbb{R}$, dann ist φ die charakteristische Funktion einer Zufallsvariable X und es gilt $X_n \xrightarrow{d} X$.

Definition 4.1.3

Sei ν ein Maß über den Meßraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Dann nennt man ν ein Lèvy-Maß, wenn $\nu(\{0\}) = 0$ und

$$\int_{\mathbb{R}} \min\{y^2, 1\} \nu(dy) < \infty.$$

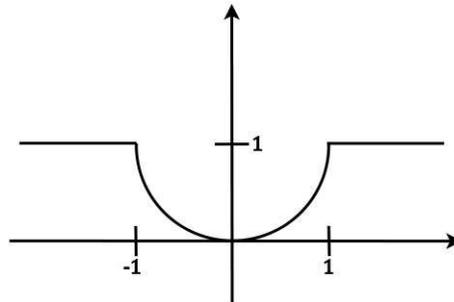


Abbildung 4.1:

Beachte

- Offenbar ist jedes Lèvy-Maß σ -endlich und

$$\nu((-\varepsilon, \varepsilon)^c) < \varepsilon, \quad \text{für alle } \varepsilon > 0, \quad (4.1.1)$$

wobei $(-\varepsilon, \varepsilon)^c = \mathbb{R} \setminus (-\varepsilon, \varepsilon)$.

- Insbesondere ist jedes endliche Maß ν ein Lèvy-Maß, falls $\nu(\{0\}) = 0$.

- Eine zu (4.1.1) äquivalente Bedingung ist

$$\int_{\mathbb{R}} \frac{y^2}{1+y^2} \nu(dy) < \infty, \quad \text{denn} \quad \frac{y^2}{1+y^2} \leq \min\{y^2, 1\} \leq 2 \frac{y^2}{1+y^2}. \quad (4.1.2)$$

Theorem 4.1.2

Seien $a \in \mathbb{R}$, $b \geq 0$ beliebig und sei ν ein beliebiges Lèvy-Maß. Dann ist durch die Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ mit

$$\varphi(s) = \exp \left\{ ias - \frac{bs^2}{2} + \int_{\mathbb{R}} \left(e^{isy} - 1 - isy \mathbf{1}(y \in (-1, 1)) \right) \nu(dy) \right\} \quad \text{für alle } s \in \mathbb{R}, \quad (4.1.3)$$

die charakteristische Funktion einer unbegrenzt teilbaren Zufallsvariable gegeben.

Bemerkung 4.1.1 • Die Formel (4.1.3) wird auch *Lèvy-Chintschin-Formel* genannt.

- Die Umkehrung vom Theorem 4.1.2 gilt auch, es hat also jede unbegrenzt teilbare Zufallsvariable eine solche Darstellung. Deshalb nennt man das charakteristische Tripel (a, b, ν) auch *Lèvy-Charakteristik* einer unbegrenzt teilbaren Zufallsvariable.

Beweis des Theorems 4.1.2 1. Schritt

Zeige, dass φ eine charakteristische Funktion ist.

-

$$\left| e^{isy} - 1 - isy \right| = \left| \sum_{k=0}^{\infty} \frac{(isy)^k}{k!} - 1 - isy \right| = \left| \sum_{k=2}^{\infty} \frac{(isy)^k}{k!} \right| \leq y^2 \underbrace{\left| \sum_{k=2}^{\infty} \frac{s^k}{k!} \right|}_{:=c} \leq y^2 c$$

Damit folgt mit (4.1.1) und (4.1.2), dass das Integral in (4.1.3) existiert und somit wohldefiniert ist.

- Sei nun $\{c_n\}$ eine beliebige Zahlenfolge mit $c_n > c_{n+1} > \dots > 0$ und $\lim_{n \rightarrow \infty} c_n = 0$. Dann ist die Funktion $\varphi_n : \mathbb{R} \rightarrow \mathbb{C}$ mit

$$\varphi_n(s) := \exp \left\{ is \left(a - \int_{[-c_n, c_n]^c \cap (-1, 1)} y \nu(dy) \right) - \frac{bs^2}{2} \right\} \exp \left\{ \int_{[-c_n, c_n]^c} (e^{isy} - 1) \nu(dy) \right\}$$

die charakteristische Funktion der Summe von $Z_1^{(n)}$ und $Z_2^{(n)}$, 2 unabhängigen Zufallsvariablen, denn

- der erste Faktor ist die charakteristische Funktion der Normalverteilung mit Erwartungswert $a - \int_{[-c_n, c_n]^c \cap (-1, 1)} y \nu(dy)$ und Varianz b .
- der zweite Faktor ist die charakteristische Funktion der zusammengesetzten Poisson-Verteilung mit den Charakteristiken

$$\lambda = \nu([-c_n, c_n]^c) \quad \text{und} \quad P_U(\cdot) = \nu(\cdot \cap [-c_n, c_n]^c) / \nu([-c_n, c_n]^c)$$

ist.

- Außerdem gilt $\lim_{n \rightarrow \infty} \varphi_n(s) = \varphi(s)$ für alle $s \in \mathbb{R}$, wobei offenbar φ stetig in 0 ist, den für die Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ im Exponent von (4.1.3), also

$$\psi(s) = \int_{\mathbb{R}} \left(e^{isy} - 1 - isy \mathbf{1}(y \in (-1, 1)) \right) \nu(dy) \quad \text{für alle } s \in \mathbb{R}$$

gilt $|\psi(s)| = cs^2 \int_{(-1,1)} y^2 \nu(dy) + \int_{(-1,1)^c} |e^{isy} - 1| \nu(dy)$. Hieraus und aus (4.1.2) folgt mit Satz von Lebesgue, dass $\lim_{s \rightarrow \infty} \psi(s) = 0$.

- Aus Lemma 4.1.2 ergibt sich, dass die in (4.1.3) gegebene Funktion φ die charakteristische Funktion einer Zufallsvariable ist.

2. Schritt

Die unbegrenzte Teilbarkeit dieser Zufallsvariable folgt aus dem Lemma 4.1.1 und daraus, dass für beliebige $n \in \mathbb{N}$ auch $\frac{\nu}{n}$ ein Lévy-Maß ist und dass

$$\varphi(s) = \exp \left\{ i \frac{a}{n} s - \frac{b}{n} \frac{s^2}{2} + \int_{\mathbb{R}} \left(e^{isy} - 1 - isy \mathbf{1}(y \in (-1, 1)) \right) \left(\frac{\nu}{n} \right) (dy) \right\} \quad \text{für alle } s \in \mathbb{R}.$$

□

Bemerkung 4.1.2

Die Abbildung $\eta : \mathbb{R} \rightarrow \mathbb{C}$ mit

$$\eta(s) = ias - \frac{bs^2}{2} + \int_{\mathbb{R}} \left(e^{isy} - 1 - isy \mathbf{1}(y \in (-1, 1)) \right) \nu(dy)$$

aus (4.1.3) heißt *Lévy-Exponent* dieser unbegrenzt teilbaren Verteilung.

4.1.2 Lévy-Chintschin-Darstellung

$\{X(t), t \geq 0\}$ – Lévy-Prozess. Wir wollen die charakteristische Funktion von $X(t), t \geq 0$, durch die Lévy-Chintschin-Formel darstellen.

Lemma 4.1.3

Sei $\{X(t), t \geq 0\}$ ein stochastisch stetiger Prozess, d.h. für alle $\varepsilon > 0$ und $t_0 \geq 0$ gelte $\lim_{t \rightarrow t_0} \mathbb{P}(|X(t) - X(t_0)| > \varepsilon) = 0$. Dann ist für jedes $s \in \mathbb{R}$ durch $t \mapsto \varphi_{X(t)}(s)$ eine stetige Abbildung von $[0, \infty)$ nach \mathbb{C} .

Beweis • $y \mapsto e^{isy}$ stetig in 0, d.h. für alle $\varepsilon > 0$ existiert $\delta_1 > 0$, so dass

$$\sup_{y \in (-\delta_1, \delta_1)} |e^{isy} - 1| < \frac{\varepsilon}{2}.$$

- $\{X(t), t \geq 0\}$ ist stochastisch stetig, d.h. für alle $t_0 \geq 0$ existiert $\delta_2 > 0$, so dass

$$\sup_{t \geq 0, |t - t_0| < \delta_2} \mathbb{P}(|X(t) - X(t_0)| > \delta_1) < \frac{\varepsilon}{4}.$$

Daraus folgt, dass für alle $s \in \mathbb{R}$, $t \geq 0$ und $|t - t_0| < \delta_2$ gilt

$$\begin{aligned}
|\varphi_{X(t)}(s) - \varphi_{X(t_0)}(s)| &= \left| \mathbb{E} \left(e^{isX(t)} - e^{isX(t_0)} \right) \right| \leq \mathbb{E} \left| e^{isX(t_0)} \left(e^{is(X(t)-X(t_0))} - 1 \right) \right| \\
&\leq \mathbb{E} \left| e^{is(X(t)-X(t_0))} - 1 \right| = \int_{\mathbb{R}} |e^{isy} - 1| \mathbf{P}_{X(t)-X(t_0)}(dy) \\
&\leq \int_{(-\delta_1, \delta_1)} |e^{isy} - 1| \mathbf{P}_{X(t)-X(t_0)}(dy) \\
&\quad + \underbrace{\int_{(-\delta_1, \delta_1)^c} |e^{isy} - 1| \mathbf{P}_{X(t)-X(t_0)}(dy)}_{=2} \\
&\leq \sup_{y \in (-\delta_1, \delta_1)} |e^{isy} - 1| + 2\mathbf{P}(|X(t) - X(t_0)| > \delta_1) \leq \varepsilon.
\end{aligned}$$

□

Theorem 4.1.3

Sei $\{X(t), t \geq 0\}$ ein Lèvy-Prozess. Für alle $t \geq 0$ gilt

$$\varphi_{X(t)}(s) = e^{t\eta(s)}, \quad s \in \mathbb{R},$$

wobei $\eta : \mathbb{R} \rightarrow \mathbb{C}$ eine stetige Funktion ist. Insbesondere gilt

$$\varphi_{X(t)}(s) = e^{t\eta(s)} = \left(e^{\eta(s)} \right)^t = \left(\varphi_{X(1)}(s) \right)^t, \quad \text{für alle } s \in \mathbb{R}, t \geq 0.$$

Beweis

$$\varphi_{X(t+t')}(s) = \mathbb{E} e^{isX(t+t')} = \mathbb{E} \left(e^{isX(t)} e^{is(X(t+t')-X(t))} \right) = \varphi_{X(t)}(s) \varphi_{X(t')}(s)$$

Sei $g_s : [0, \infty) \rightarrow \mathbb{C}$ definiert durch $g_s(t) = \varphi_{X(t)}(s)$, $s \in \mathbb{R}$, $g_s(t+t') = g_s(t)g_s(t')$, $t, t' \geq 0$. $X(0) = 0$.

$$\begin{cases} g_s(t+t') = g_s(t)g_s(t'), & t, t' \geq 0, \\ g_s(0) = 1, \\ g_s : [0, \infty) \rightarrow \mathbb{C} \text{ stetig.} \end{cases}$$

Daraus folgt: existiert $\eta : \mathbb{R} \rightarrow \mathbb{C}$, so dass $g_s(t) = e^{\eta(s)t}$ für alle $s \in \mathbb{R}$, $t \geq 0$. $\varphi_{X(1)}(s) = e^{\eta(s)}$ und es folgt, dass η stetig ist. □

Lemma 4.1.4

Sei μ_1, μ_2, \dots eine Folge von endlichen Maßen (auf $\mathcal{B}(\mathbb{R})$) mit

1. $\sup_{n \geq 1} \mu_n(\mathbb{R}) < c$, $c = \text{const} < \infty$ (gleichmäßig beschränkt)
2. für alle $\varepsilon > 0$ existiert $B_\varepsilon \in \mathcal{B}(\mathbb{R})$ kompakt, so dass $\sup_{n \geq 1} \mu_n(B_\varepsilon^c) \leq \varepsilon$. Daraus folgt, dass es eine Teilfolge $\mu_{n_1}, \mu_{n_2}, \dots$ und ein endliches Maß über $\mathcal{B}(\mathbb{R})$ existiert, so dass für alle $f : \mathbb{R} \rightarrow \mathbb{C}$, beschränkt, stetig, gilt

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}} f(y) \mu_{n_k}(dy) = \int_{\mathbb{R}} \lim_{k \rightarrow \infty} f(y) \mu_{n_k}(dy) = \int_{\mathbb{R}} f(y) \mu(dy)$$

Beweis Siehe [14], S. 122 - 123. □

Theorem 4.1.4

Sei $\{X(t), t \geq 0\}$ ein Lévy-Prozess. Dann gibt es $a \in \mathbb{R}$, $b \geq 0$ und ein Lévy-Maß ν , so dass

$$\varphi_{X(1)}(s) = e^{ias - \frac{bs^2}{2}} + \int_{\mathbb{R}} \left(e^{isy} - 1 - iy1(y \in (-1, 1)) \right) \nu(dy), \quad \text{für alle } s \in \mathbb{R}.$$

Beweis Für alle Nullfolgen t_1, t_2, \dots gilt

$$\eta(s) = \left(e^{t\eta(s)} \right)' \Big|_{t=0} = \lim_{n \rightarrow \infty} \frac{e^{t_n \eta(s)} - 1}{t_n} = \lim_{n \rightarrow \infty} \frac{\varphi_{X(t_n)}(s) - 1}{t_n}. \quad (4.1.4)$$

$\eta : \mathbb{R} \rightarrow \mathbb{C}$ stetig \Rightarrow Die Konvergenz in (4.1.4) ist gleichmäßig in $s \in [-t_0, t_0]$ für ein $s_0 > 0$ (Taylor-Entwicklung von $e^{t\eta(s)}$). Sei $t_n = \frac{1}{n}$ und P_n die Verteilung von $X(\frac{1}{n})$. Daraus folgt, dass

$$\begin{aligned} \lim_{n \rightarrow \infty} n \int_{\mathbb{R}} (e^{isy} - 1) P_n(ds) &= \lim_{n \rightarrow \infty} n \frac{\varphi_{X(\frac{1}{n})}(s) - 1}{\frac{1}{n}} = \eta(s) \\ \lim_{n \rightarrow \infty} \int_{\mathbb{R}} n \int_{-s_0}^{s_0} (e^{isy} - 1) P_n(dy) ds &= \int_{-s_0}^{s_0} \eta(s) ds \\ \Rightarrow \lim_{n \rightarrow \infty} n \int_{\mathbb{R}} \left(1 - \frac{\sin(s_0 y)}{s_0 y} \right) P_n(dy) &= -\frac{1}{2s_0} \int_{-s_0}^{s_0} \eta(s) ds \end{aligned}$$

$\eta : \mathbb{R} \rightarrow \mathbb{C}$ ist stetig mit $\eta(0) = 0$ und daraus folgt, dass es für alle $\varepsilon > 0$ $\delta_0 > 0$ existiert, so dass $\left| -\frac{1}{2s_0} \int_{-s_0}^{s_0} \eta(s) ds \right| < \varepsilon$. Weil $1 - \frac{\sin(s_0 y)}{s_0 y} \geq \frac{1}{2}$, $|s_0 y| \geq 2$, gilt: für alle $\varepsilon > 0$ existiert $s_0 > 0$, $n_0 > 0$, so dass

$$\limsup_{n \rightarrow \infty} \frac{n}{2} \int_{\left\{ y: |y| \geq \frac{2}{s_0} \right\}} P_n(dy) \leq \limsup_{n \rightarrow \infty} n \int_{\mathbb{R}} \left(1 - \frac{\sin(s_0 y)}{s_0 y} \right) P_n(dy) < \varepsilon.$$

Für alle $\varepsilon > 0$ existiert $s_0 > 0$, $n_0 > 0$, so dass

$$n \int_{\left\{ y: |y| \geq \frac{2}{s_0} \right\}} P_n(dy) \leq 4\varepsilon, \quad \text{für alle } n \geq n_0.$$

Durch Verkleinerung von s_0 erhält man

$$n \int_{\left\{ y: |y| \geq \frac{2}{s_0} \right\}} P_n(dy) \leq 4\varepsilon, \quad \text{für alle } n \geq 1.$$

$$\frac{y^2}{1+y^2} \leq c \left(1 - \frac{\sin y}{y} \right), \quad \text{für alle } y \neq 0 \quad \text{und ein } c > 0.$$

Daraus folgt, dass

$$\sup_{n \geq 1} n \int_{\mathbb{R}} \frac{y^2}{1+y^2} P_n(dy) \leq c' \quad \text{für ein } c' < \infty.$$

Sei nun $\mu_n : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty)$ definiert durch

$$\mu_n(B) = n \int_B \frac{y^2}{1+y^2} P_n(dy) \quad \text{für alle } B \in \mathcal{B}(\mathbb{R}).$$

Es folgt, dass $\{\mu_n\}_{n \in \mathbb{N}}$ gleichmäßig beschränkt ist, $\sup_{n \geq 1} \mu_n(\mathbb{R}) < c'$. Außerdem gilt $\frac{y^2}{1+y^2} \leq 1$, $\sup_{n \geq 1} \mu_n \left(\left\{ y : |y| > \frac{2}{s_0} \right\} \right) \leq 4\varepsilon$ und $\{\mu_n\}_{n \in \mathbb{N}}$ ist relativ kompakt. Nach dem Lemma 4.1.3 gilt: existiert $\{\mu_{n_k}\}_{k \in \mathbb{N}}$, so dass

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}} f(y) \mu_{n_k}(dy) = \int_{\mathbb{R}} f(y) \mu(dy)$$

für ein Maß μ und f stetig und beschränkt. Sei für $s \in \mathbb{R}$ die Funktion $f_s : \mathbb{R} \rightarrow \mathbb{C}$ definiert durch

$$f_s(y) = \begin{cases} (e^{isy} - 1 - isy) \frac{1+y^2}{y^2}, & y \neq 0, \\ -\frac{s^2}{2}, & \text{sonst.} \end{cases}$$

Daraus folgt, dass f_s beschränkt und stetig ist und

$$\begin{aligned} \eta(s) &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} (e^{isy} - 1) P_n(dy) \\ &= \lim_{n \rightarrow \infty} \left(\int_{\mathbb{R}} f_s(y) \mu_n(dy) + isn \int_{\mathbb{R}} \sin y P_n(dy) \right) \\ &= \lim_{n \rightarrow \infty} \left(\int_{\mathbb{R}} f_s(y) \mu_{n_k}(dy) + isn_k \int_{\mathbb{R}} \sin y P_{n_k}(dy) \right) \\ &= \int_{\mathbb{R}} f_s(y) \mu(dy) + \lim_{k \rightarrow \infty} isn_k \int_{\mathbb{R}} \sin y P_{n_k}(dy) \\ \eta(s) &= ia's - \frac{bs^2}{2} + \int_{\mathbb{R}} (e^{isy} - 1 - is \sin y) \nu(dy), \end{aligned}$$

für alle $s \in \mathbb{R}$ mit $a' = \lim_{k \rightarrow \infty} isn_k \int_{\mathbb{R}} \sin y P_{n_k}(dy)$, $b = \mu(\{0\})$, $\nu : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty)$,

$$\nu(dy) = \begin{cases} \frac{1+y^2}{y^2} \mu(dy), & y \neq 0, \\ 0, & y = 0. \end{cases}$$

$$\int_{\mathbb{R}} |y \mathbf{1}(y \in (-1, 1)) - \sin y| \nu(dy) < \infty.$$

$$|y \mathbf{1}(y \in (-1, 1)) - \sin y| \frac{1+y^2}{y^2} < c'', \quad \text{für alle } y \neq 0 \quad \text{und ein } c'' > 0.$$

Daraus folgt, dass

$$\eta(s) = ias - \frac{bs^2}{2} + \int_{\mathbb{R}} (e^{isy} - 1 - isy \mathbf{1}(y \in (-1, 1))) \nu(dy), \quad \text{für alle } s \in \mathbb{R}.$$

$$a = a' + \int_{\mathbb{R}} (y \mathbf{1}(y \in (-1, 1)) - \sin y) \nu(dy).$$

□

4.1.3 Beispiele

1. Wiener-Prozess (es genügt $X(1)$ zu betrachten)

$X(1) \sim \mathcal{N}(0, 1)$, $\varphi_{X(1)}(s) = e^{-\frac{s^2}{2}}$ und daraus folgt

$$(a, b, \nu) = (0, 1, 0).$$

Sei $X = \{X(t), t \geq 0\}$ ein Wiener-Prozess mit Drift μ , d.h. $X(t) = \mu t + \sigma W(t)$, $W = \{W(t), t \geq 0\}$ – Brownsche Bewegung. Es folgt

$$(a, b, \nu) = (\mu, \sigma^2, 0).$$

$$\varphi_{X(1)}(s) = \mathbb{E}e^{isX(1)} = \mathbb{E}e^{(\mu + \sigma W(1))is} = e^{\mu is} \varphi_{W(1)}(\sigma s) = e^{is\mu - \sigma^2 \frac{s^2}{2}}, \quad s \in \mathbb{R}.$$

2. Zusammengesetzter Poisson-Prozess mit Parametern (λ, \mathbb{P}_U)

$X(t) = \sum_{i=1}^{N(t)} U_i$, $N(t) \sim \text{Pois}(\lambda t)$, U_i u.i.v. $\sim \mathbb{P}_U$.

$$\begin{aligned} \varphi_{X(1)}(s) &= \exp \left\{ \lambda \int_{\mathbb{R}} (e^{isx} - 1) \mathbb{P}_U(dx) \right\} \\ &= \exp \left\{ \lambda is \int_{\mathbb{R}} x \mathbf{1}(x \in [-1, 1]) \mathbb{P}_U(dx) + \lambda \int_{\mathbb{R}} (e^{isx} - 1 - isx \mathbf{1}(x \in [-1, 1])) \mathbb{P}_U(dx) \right\} \\ &= \exp \left\{ \lambda is \int_{-1}^1 x \mathbb{P}_U(dx) + \lambda \int_{\mathbb{R}} (e^{isx} - 1 - isx \mathbf{1}(x \in [-1, 1])) \mathbb{P}_U(dx) \right\}, \quad s \in \mathbb{R}. \end{aligned}$$

Daraus folgt

$$(a, b, \nu) = \left(\lambda \int_{-1}^1 x \mathbb{P}_U(dx), 0, \lambda \mathbb{P}_U \right), \quad \mathbb{P}_U \text{ – endlich auf } \mathbb{R}.$$

3. Prozesse von Gauß-Poisson-Typ

$X = \{X(t), t \geq 0\}$, $X(t) = X_1(t) + X_2(t)$, $t \geq 0$.

$X_1 = \{X_1(t), t \geq 0\}$ und $X_2 = \{X_2(t), t \geq 0\}$ unabhängig.

X_1 – Wiener-Prozess mit Drift μ und Varianz σ^2 ,

X_2 – Zusammengesetzter-Poisson-Prozess mit Parametern λ, \mathbb{P}_U .

$$\begin{aligned} \varphi_{X(t)}(s) &= \varphi_{X_1(t)}(s) \varphi_{X_2(t)}(s) \\ &= \exp \left\{ is \left(\mu + \lambda \int_{-1}^1 x \mathbb{P}_U(dx) \right) - \frac{\sigma^2 s^2}{2} \right. \\ &\quad \left. + \int_{\mathbb{R}} \lambda (e^{isx} - 1 - isx \mathbf{1}(x \in [-1, 1])) \mathbb{P}_U(dx) \right\}, \quad s \in \mathbb{R}. \end{aligned}$$

Daraus folgt

$$(a, b, \nu) = \left(\mu + \lambda \int_{-1}^1 x \mathbb{P}_U(dx), \sigma^2, \lambda \mathbb{P}_U \right).$$

4. Stabile Lèvy-Prozesse

$X = \{X(t), t \geq 0\}$ – Lèvy-Prozess mit $X(t) \sim \alpha$ stabile Verteilung, $\alpha \in (0, 2]$. Falls $X = W$ (Wiener-Prozess), dann $X(1) \sim \mathcal{N}(0, 1)$. Seien Y, Y_1, \dots, Y_n u.i.v. $\mathcal{N}(\mu, \sigma^2)$ -Variablen. Nach Faltungstabilität der Normalverteilung gilt

$$\begin{aligned} Y_1 + \dots + Y_n \sim \mathcal{N}(n\mu, n\sigma^2) &\stackrel{d}{=} \sqrt{n}Y + n\mu - \sqrt{n}\mu \\ &= \sqrt{n}Y + \mu(n - \sqrt{n}) \\ &= n^{\frac{1}{2}}Y + \mu \left(n^{\frac{2}{2}} - n^{\frac{1}{2}} \right), \quad \alpha = 2. \end{aligned}$$

Definition 4.1.4

Die Verteilung einer Zufallsvariable Y heißt α -stabil, falls für alle $n \in \mathbb{N}$ nur Kopien Y_1, \dots, Y_n gibt (von Y)

$$Y_1 + \dots + Y_n \stackrel{d}{=} n^{\frac{1}{\alpha}} Y + d_n,$$

wobei d_n deterministisch ist (also eine Konstante bzgl. W , d.h. nicht zufällig). Dabei heißt $\alpha \in (0, 2]$ *Stabilitätsindex*.

$$d_n = \begin{cases} \mu \left(n - n^{\frac{1}{\alpha}} \right), & \alpha \neq 1, \\ \mu n \log n, & \alpha = 1. \end{cases}$$

Ohne Beweis**Beispiel 4.1.1** • $\alpha = 2$: Normalverteilung

- $\alpha = 1$: Cauchy-Verteilung mit Parametern (μ, σ^2) . Die Dichte:

$$f_Y(x) = \frac{\sigma}{\pi \left((x - \mu)^2 + \sigma^2 \right)}, \quad x \in \mathbb{R}.$$

Dabei gilt $EY^2 = \infty$, EY existiert nicht.

- $\alpha = \frac{1}{2}$: Lévy-Verteilung mit Parametern (μ, σ^2) . Die Dichte:

$$f_Y(x) = \begin{cases} \left(\frac{\sigma}{2\pi} \right)^{\frac{1}{2}} \frac{1}{(x-\mu)^{\frac{3}{2}}} \exp \left\{ -\frac{\sigma}{2(x-\mu)} \right\}, & x > \mu, \\ 0, & \text{sonst.} \end{cases}$$

Diese Beispiele sind die wenigen Beispiele von α -stabilen Verteilungen, die eine explizite Form der Dichte besitzen. Für andere $\alpha \in (0, 2)$, $\alpha \neq \frac{1}{2}, 1$, wird die α -stabile Verteilung durch ihre charakteristische Funktionen eingeführt. Generell gilt: Falls Y α -stabil, $\alpha \in (0, 2]$, dann $E|Y|^p < \infty$, $0 < p < \alpha$.

Definition 4.1.5

Die Verteilung einer Zufallsvariable heißt symmetrisch, falls $Y \stackrel{d}{=} -Y$, falls Y eine symmetrische α -stabile Verteilung besitzt, $\alpha \in (0, 2]$,

$$\varphi_Y(s) = \exp \{ -c |s|^\alpha \}.$$

In der Tat, aus der Stabilität von Y folgt

$$(\varphi_Y(s))^n = e^{id_n s} \varphi_Y \left(n^{\frac{1}{\alpha}} s \right), \quad s \in \mathbb{R}.$$

Daraus folgt, dass $d_n = 0$, denn $\varphi_{-Y}(s) = \varphi_Y(s)$. Es gilt: $e^{id_n s} = e^{-id_n s}$, $s \in \mathbb{R}$ und $d_n = 0$. Der Rest ist eine Übungsaufgabe.

Lemma 4.1.5

Lévy-Chintschin-Darstellung der charakteristischen Funktion ist eine stabile Verteilung. Eine Lévy-Charakteristik (a, b, ν) , $a \in \mathbb{R}$ beliebig.

$$b = \begin{cases} \sigma^2, & \alpha = 2, \\ 0, & \alpha < 2. \end{cases}$$

$$\nu(dx) = \begin{cases} 0, & \alpha = 2, \\ \frac{c_1}{x^{1+\alpha}} \mathbf{1}(x \geq 0) dx + \frac{c_2}{|x|^{1+\alpha}} \mathbf{1}(x < 0) dx, & \alpha < 2, \\ c_1, c_2 \geq 0 : c_1 + c_2 > 0, & \end{cases}$$

Ohne Beweis

Man kann zeigen, dass

$$P(|Y| \geq x) \underset{x \rightarrow \infty}{\sim} \begin{cases} e^{-\frac{x^2}{2\sigma^2}}, & \alpha = 2, \\ \frac{c}{x^\alpha}, & \alpha < 2. \end{cases}$$

Definition 4.1.6

Der Lévy-Prozess $X = \{X(t), t \geq 0\}$ heißt stabil, wenn $X(1)$ eine α -stabile Verteilung besitzt, $\alpha \in (0, 2]$ ($\alpha = 2$: Brownsche Bewegung (mit Drift)).

4.1.4 Subordinatoren**Definition 4.1.7**

Ein Lévy-Prozess $X = \{X(t), t \geq 0\}$ heißt *Subordinator*, falls für alle $0 < t_1 < t_2$, $X(t_1) \leq X(t_2)$ f.s. gilt

$$X(0) = 0 \quad \text{f.s.} \quad \Rightarrow \quad X(t) \geq 0, \quad t \geq 0.$$

Diese Klasse von Lévy-Prozessen ist deshalb wichtig, weil man leicht $\int_a^b g(t) dX(t)$ einführen kann als Lebesgue-Stieltjes-Integrale.

Theorem 4.1.5

Der Lévy-Prozess $X = X(t)$, $t \geq 0$ ist genau dann ein Subordinator, wenn die Lévy-Chintschin-Darstellung in der Form sich darstellen läßt

$$\varphi_{X(t)}(s) = \exp \left\{ ias + \int_{\mathbb{R}} (e^{isx} - 1) \nu(dx) \right\}, \quad s \in \mathbb{R},$$

wobei ν das Lévy-Maß ist, mit

$$\nu((-\infty, 0)) = 0, \quad \int_0^\infty \min\{1, y^2\} \nu(dy) < \infty.$$

Beweis Hinlänglichkeit

Es ist zu zeigen, dass $X(t_2) \geq X(t_1)$ f.s., falls $t_2 \geq t_1 \geq 0$.

Zunächst zeigen wir, dass $X(1) \geq 0$ f.s.. Falls $\nu \equiv 0$, dann $X(1) = a$ f.s., daher

$$\varphi_{X(t)}(s) = \left(\frac{\varphi(s)}{X(t)} \right)^t = e^{iats}, \quad s \in \mathbb{R}.$$

$X(t) = at$ f.s. und daraus folgt, dass $X(t) \uparrow$ und X ist ein Subordinator. Falls $\nu([0, \infty)) > 0$, dann existiert $N > 0$, so dass $n \geq N$, $0 < \nu\left(\left[\frac{1}{n}, \infty\right)\right) < \infty$. Es folgt

$$\varphi_{X(t)}(s) = \exp \left\{ ias + \lim_{n \rightarrow \infty} \int_{\frac{1}{n}}^\infty (e^{isx} - 1) \nu(dx) \right\} = e^{ias} \lim_{n \rightarrow \infty} \varphi_n(s), \quad s \in \mathbb{R},$$

wobei $\varphi_n(s) = \int_{\frac{1}{n}}^\infty (e^{isx} - 1) \nu(dx)$ die charakteristische Funktion einer zusammengesetzter Poisson-Verteilung mit Parametern $\left(\nu\left(\left[\frac{1}{n}, \infty\right)\right), \frac{\nu\left(\left[\frac{1}{n}, \infty\right)\right)}{\nu\left(\left[\frac{1}{n}, \infty\right)\right)} \right)$ für alle $n \in \mathbb{N}$ ist. Sei Z_n die Zufallsvariable mit charakteristischer Funktion φ_n . Es gilt: $Z_n = \sum_{i=1}^{N_n} U_i$, $N_n \sim \text{Pois}\left(\nu\left(\left[\frac{1}{n}, \infty\right)\right)\right)$,

$U_i \sim \frac{\nu(\cap(\frac{1}{n}, \infty))}{\nu(\cap(\frac{1}{n}, \infty))}$ und daraus folgt $Z_n \geq 0$ f.s. und $X(1) = \underbrace{a}_{=0} + \underbrace{\lim Z_n}_{\geq 0} \geq 0$ f.s. Da X ein Lèvy-Prozess ist, gilt

$$X(1) = X\left(\frac{1}{n}\right) + \left(X\left(\frac{2}{n}\right) - X\left(\frac{1}{n}\right)\right) + \dots + \left(X\left(\frac{n}{n}\right) - X\left(\frac{n-1}{n}\right)\right),$$

wobei wegen Stationarität und Unabhängigkeit der Zuwächse $X\left(\frac{k}{n}\right) - X\left(\frac{k-1}{n}\right) \stackrel{f.s.}{\geq} 0$ für $1 \leq k \leq n$ für alle n . $X(q_2) - X(q_1) \geq 0$ f.s. für alle $q_1, q_2 \in \mathbb{Q}$, $q_2 \geq q_1 \geq 0$. Nun seien $t_1, t_2 \in \mathbb{Q}$, so dass $0 \leq t_1 \leq t_2$. Seien $\{q_1^{(n)}, q_2^{(n)}\}$ Folgen von Zahlen aus \mathbb{Q} mit $q_1^{(n)} \leq q_2^{(n)}$. $q_1^{(n)} \downarrow t_1$, $q_2^{(n)} \uparrow t_2$, $n \rightarrow \infty$. Für $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}(X(t_2) - X(t_1) < -\varepsilon) &= \mathbb{P}\left(X(t_2) - X(q_2^{(n)}) + X(q_2^{(n)}) - X(q_1^{(n)}) + X(q_1^{(n)}) - X(t_1) < -\varepsilon\right) \\ &\leq \mathbb{P}\left(X(t_2) - X(q_2^{(n)}) + X(q_1^{(n)}) - X(t_1) < -\varepsilon\right) \\ &\leq \mathbb{P}\left(X(t_2) - X(q_2^{(n)}) < -\varepsilon\right) + \mathbb{P}\left(X(q_1^{(n)}) - X(t_1) \leq -\frac{\varepsilon}{2}\right) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

$$\Rightarrow \mathbb{P}(X(t_2) - X(t_1) < \varepsilon) = 0 \quad \text{für alle } \varepsilon > 0$$

$$\Rightarrow \mathbb{P}(X(t_2) - X(t_1) < 0) = \lim_{\varepsilon \rightarrow +0} \mathbb{P}(X(t_2) - X(t_1) < \varepsilon) = 0$$

$$\Rightarrow X(t_2) \geq X(t_1) \quad \text{f.s.}$$

Notwendigkeit

Sei X ein Lèvy-Prozess, der ein Subordinator ist. Es ist zu zeigen, dass $\varphi_{X_1(t)}(\cdot)$ die obige Form hat.

Nach der Lèvy-Chintschin-Darstellung für $X_1(t)$ gilt

$$\varphi_{X(1)}(s) = \exp \left\{ ias - \frac{b^2 s^2}{2} + \int_0^\infty \left(e^{isx} - 1 - isx \mathbf{1}(x \in [-1, 1]) \right) \nu(dx) \right\}, \quad s \in \mathbb{R}.$$

Das Maß ν ist auf $[0, \infty)$ konzentriert, weil $X(1) \stackrel{f.s.}{\geq} 0$ und aus dem Beweis des Theorems 4.1.4 $\nu((-\infty, 0)) = 0$ gewählt werden kann.

$$\varphi_{X(1)}(s) \leq \underbrace{\exp \left\{ ias - \frac{b^2 s^2}{2} \right\}}_{:=\varphi_{Y_1}(s)} \underbrace{\exp \left\{ \int_0^\infty \left(e^{isx} - 1 - isx \mathbf{1}(x \in [-1, 1]) \right) \nu(dx) \right\}}_{:=\varphi_{Y_2}(s)}$$

Daraus folgt, dass $X(1) = Y_1 + Y_2$, Y_1 und Y_2 unabhängig, $Y_1 \sim \mathcal{N}(a, b^2)$ und deswegen $b = 0$. Für alle $\varepsilon \in (0, 1)$

$$\varphi_{X_1}(s) = \exp \left\{ is \left(a - \int_\varepsilon^1 x \nu(dx) \right) + \int_0^\varepsilon \left(e^{isx} - 1 - isx \right) \nu(dx) + \int_0^\infty \left(e^{isx} - 1 \right) \nu(dx) \right\}$$

Es ist zu zeigen, dass für $\varepsilon \rightarrow 0$ gilt, dass $\int_\varepsilon^\infty (e^{isx} - 1) \nu(dx) \rightarrow \int_0^\infty (e^{isx} - 1) \nu(dx) < \infty$ mit $\int_0^1 \min\{x, 1\} \nu(dx) < \infty$. $\varphi_{X(1)}(s) = \exp \left\{ is \left(a - \int_\varepsilon^1 x \nu(dx) \right) \right\} \varphi_{Z_1}(s) \varphi_{Z_2}(s)$, wobei Z_1 und Z_2 unabhängig, $\varphi_{Z_1}(s) = \exp \left\{ (e^{isx} - 1 - isx) \nu(dx) \right\}$, $\varphi_{Z_2}(s) = \exp \left\{ \int_\varepsilon^\infty (e^{isx} - 1) \nu(dx) \right\}$, $s \in \mathbb{R}$.

$X(1) \stackrel{d}{=} a - \int_{\varepsilon}^1 x \nu(dx) + Z_1 + Z_2$. Es existiert $\varphi_{Z_1}^{(2)}(0) = \frac{-\mathbb{E}Z_1^2}{2} < \infty$, $\varphi_{Z_1}^{(1)}(0) = 0 = i\mathbb{E}Z_1$ und daraus ergibt sich, dass $\mathbb{E}Z_1 = 0$ und $\mathbb{P}(Z_1 \leq 0) > 0$. Andererseits, hat Z_2 eine zusammengesetzte Poisson-Verteilung mit Parametern $(\nu([\varepsilon, \infty)), \frac{\nu(\cap[\varepsilon, +\infty])}{\nu([\varepsilon, +\infty])})$, $\varepsilon \in (0, 1)$.

$$\begin{aligned} &\Rightarrow \mathbb{P}(Z_2 \leq 0) > 0 \\ &\Rightarrow \mathbb{P}(Z_1 + Z_2 \leq 0) \geq \mathbb{P}(Z_1 \leq 0, Z_2 \leq 0) = \mathbb{P}(Z_1 \leq 0) \mathbb{P}(Z_2 \leq 0) > 0 \\ &\Rightarrow a - \int_{\varepsilon}^1 x \nu(dx) \geq 0 \quad \text{für alle } \varepsilon \in (0, 1) \\ &\Rightarrow \int_0^a \min\{x, 1\} dx < \infty \\ &\Rightarrow \text{für } \varepsilon \rightarrow \infty \quad Z_1 \xrightarrow{d} 0 \\ &\varphi_{X(1)}(s) = \exp \left\{ is \left(a - \int_0^1 x \nu(dx) \right) + \int_0^{\infty} (e^{isx} - 1) \nu(dx) \right\}, \quad s \in \mathbb{R}. \end{aligned}$$

□

Beispiel 4.1.2 (α -stabiler Subordinator):

$X = \{X(t), t \geq 0\}$ ein Lévy-Prozess, Subordinator, mit $a = 0$ – Lévy-Maß.

$$\nu(dx) = \begin{cases} \frac{\alpha}{\Gamma(1-\alpha)} \frac{1}{x^{1+\alpha}} dx, & x > 0, \\ 0 \cdot \frac{1}{x^{1+\alpha}} dx = 0, & x \leq 0. \end{cases}$$

Daraus folgt, dass X ein α -stabiler Lévy-Prozess ist.

Zeigen wir, dass $\hat{l}_{X(\cdot)}(s) = \mathbb{E}e^{-sX(t)} = e^{-ts^\alpha}$ für alle $s, t \geq 0$.

$$\varphi_{X(t)}(s) = (\varphi_{X(1)}(s))^t = \exp \left\{ t \int_0^{\infty} (e^{isx} - 1) \frac{\alpha}{\Gamma(1-\alpha)} \frac{1}{x^{1+\alpha}} dx \right\}, \quad s \in \mathbb{R}.$$

Es ist zu zeigen, dass

$$U^d = \frac{\alpha}{\Gamma(1-\alpha)} \int_0^{\infty} (1 - e^{-ux}) \frac{dx}{x^{1+\alpha}}, \quad u \geq 0.$$

Das genügt, weil $\varphi_{X(t)}(\cdot)$ analytisch auf $\{Z \in \mathbb{C} : \Im Z \geq 0\}$ fortgesetzt werden kann, d.h. $\varphi_{X(t)}(iu) = \hat{l}_{X(t)}$, $u \geq 0$. In der Tat gilt

$$\begin{aligned} \int_0^{\infty} (1 - e^{-ux}) \frac{dx}{x^{1+\alpha}} &= \int_0^{\infty} u \int_0^x e^{-uy} dy x^{-1-\alpha} dx \\ &= \int_0^{\infty} \int_y^{\infty} u e^{-uy} x^{-1-\alpha} dx dy \\ &= \int_0^{\infty} \int_y^{\infty} x^{-1-\alpha} dx u e^{-uy} dy \\ &= \frac{u}{\alpha} \int_0^{\infty} e^{-uy} y^{-\alpha} dy \\ &= \frac{u}{\alpha} \int_0^{\infty} e^{-z} z^{-\alpha} \frac{1}{u^{-\alpha}} d\left(\frac{z}{u}\right) \\ &= \frac{u^\alpha}{\alpha} \int_0^{\infty} e^{-z} z^{(1-\alpha)-1} dz \\ &= \frac{u^\alpha}{\alpha} \Gamma(1-\alpha) \end{aligned}$$

und daraus folgt $\hat{l}_{X(t)}(s) = e^{-ts^\alpha}$, $t, s \geq 0$.

4.2 Ergänzende Aufgaben

Aufgabe 4.2.1

Gegeben sei eine reellwertige Zufallsvariable X mit Verteilungsfunktion F und charakteristischer Funktion φ . Zeigen Sie, dass dann die folgenden Aussagen gelten:

- Falls X unbegrenzt teilbar ist, dann gilt $\varphi(t) \neq 0$ für alle $t \in \mathbb{R}$. *Hinweis: Zeigen Sie, dass $\lim_{n \rightarrow \infty} |\varphi_n(s)|^2 = 1$ für alle $s \in \mathbb{R}$, falls $\varphi(s) = (\varphi_n(s))^n$. Beachten Sie außerdem, dass $|\varphi_n(s)|^2$ wiederum eine charakteristische Funktion ist und $\lim_{n \rightarrow \infty} x^{\frac{1}{n}} = 1$ für $x > 0$ gilt.*
- Geben Sie ein Beispiel (mit Begründung) für eine Verteilung an, die nicht unbegrenzt teilbar ist.

Aufgabe 4.2.2

Sei $X = \{X(t), t \geq 0\}$ ein Lévy-Prozess. Zeigen Sie, dass dann die Zufallsvariable $X(t)$ für jedes $t \geq 0$ unbegrenzt teilbar ist.

Aufgabe 4.2.3

Zeigen Sie, dass die Summe von zwei unabhängigen Lévy-Prozessen wieder ein Lévy-Prozess ist, und geben Sie die zugehörige Lévy-Charakteristik an.

Aufgabe 4.2.4

Betrachten Sie die folgende Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ mit

$$\varphi(t) = e^{\psi(t)}, \quad \text{wobei } \psi(t) = 2 \sum_{k=-\infty}^{\infty} 2^{-k} (\cos(2^k t) - 1).$$

Zeigen Sie, dass $\varphi(t)$ die charakteristische Funktion einer unbegrenzt teilbaren Verteilung ist. *Hinweis: Betrachten Sie die Lévy-Chintschin-Darstellung mit Maß $\nu(\{\pm 2^k\}) = 2^{-k}$, $k \in \mathbb{Z}$.*

Aufgabe 4.2.5

Der Lévy-Prozess $\{X(t), t \geq 0\}$ sei ein Gamma-Prozess mit Parametern $b, p > 0$, das heißt, für jedes $t \geq 0$ gelte $X(t) \sim \Gamma(b, pt)$. Zeige, dass $\{X(t), t \geq 0\}$ ein Subordinator ist mit dem Laplace-Exponenten $\xi(u) = \int_0^\infty (1 - e^{-uy}) \nu(dy)$ für $\nu(dy) = py^{-1} e^{-by} dy$, $y > 0$. (Der Laplace-Exponent von $\{X(t), t \geq 0\}$ ist die Funktion $\xi : [0, \infty) \rightarrow [0, \infty)$, für die $\mathbb{E}e^{-uX(t)} = e^{-t\xi(u)}$ für beliebige $t, u \geq 0$ gilt)

Aufgabe 4.2.6

Sei $\{X(t), t \geq 0\}$ ein Lévy-Prozess mit charakteristischem Lévy-Exponenten η und $\{\tau(s), s \geq 0\}$ ein unabhängiger Subordinator mit charakteristischem Lévy-Exponenten γ . Der stochastische Prozess Y sei definiert durch $Y = \{X(\tau(s)), s \geq 0\}$.

- Zeige, dass

$$\mathbb{E} \left(e^{i\theta Y(\tau(s))} \right) = e^{\gamma(-i\eta(\theta))s}, \quad \theta \in \mathbb{R},$$

wobei $\Im z$ den Imaginärteil von z bezeichnet.

Hinweis: Weil τ ein Prozess mit nicht-negativen Werten ist, gilt $\mathbb{E}e^{i\theta\tau(s)} = e^{\gamma(\theta)s}$ für alle $\theta \in \{z \in \mathbb{C} : \Im z \geq 0\}$ durch analytische Fortsetzung in Theorem 4.1.3.

- Zeige, dass Y ein Lévy-Prozess mit charakteristischem Lévy-Exponenten $\gamma(-i\eta(\cdot))$ ist.

Aufgabe 4.2.7

Sei $\{X(t), t \geq 0\}$ ein zusammengesetzter Poisson-Prozess mit Lévy-Maß

$$\nu(dx) = \frac{\lambda\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}} dx, \quad x \in \mathbb{R},$$

wobei $\lambda, \sigma > 0$. Zeigen Sie, dass $\{\sigma W(N(t)), t \geq 0\}$ die gleichen endlich-dimensionalen Verteilungen wie X hat, wobei $\{N(s), s \geq 0\}$ ein Poisson-Prozess mit Intensität 2λ und W ein von N unabhängiger Standard-Wiener-Prozess ist.

Hinweise zu Aufgabe 4.2.6 a) und Aufgabe 4.2.7

- Zur Berechnung des Erwartungswertes für die charakteristische Funktion kann die Identität $E(X) = E(E(X|Y)) = \int_{\mathbb{R}} E(X|Y = y)F_Y(dy)$ für zwei Zufallsvariablen X und Y benutzt werden. Dabei sollte auf $\tau(s)$ bedingt werden.
- $\int_{-\infty}^{\infty} \cos(sy)e^{-\frac{y^2}{2a}} dy = \sqrt{2\pi a} \cdot e^{-\frac{as^2}{2}}$ für $a > 0$ und $s \in \mathbb{R}$.

Aufgabe 4.2.8

Sei W ein Standard-Wiener-Prozess und τ ein unabhängiger $\frac{\alpha}{2}$ -stabiler Subordinator, wobei $\alpha \in (0, 2)$. Zeige, dass $\{W(\tau(s)), s \geq 0\}$ ein α -stabiler Lévy-Prozess ist.

Aufgabe 4.2.9

Zeige, dass der Subordinator T mit Randdichte

$$f_{T(t)}(s) = \frac{t}{2\sqrt{\pi}} s^{-\frac{3}{2}} e^{-\frac{t^2}{4s}} \mathbf{1}\{s > 0\}$$

ein $\frac{1}{2}$ -stabiler Subordinator ist. (Hinweis: Differenziere die Laplace-Transformierte von $T(t)$ und löse die Differentialgleichung)

5 Martingale

5.1 Grundbegriffe

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein vollständiger Wahrscheinlichkeitsraum.

Definition 5.1.1

Sei $\{\mathcal{F}_t, t \geq 0\}$ eine Familie von σ -Algebra $\mathcal{F}_t \subset \mathcal{F}$. Sie heißt

1. *eine Filtration*, falls $\mathcal{F}_s \subseteq \mathcal{F}_t, 0 \leq s < t$.
2. *eine vollständige Filtration*, falls sie eine Filtration ist, so dass \mathcal{F}_0 (und somit alle $\mathcal{F}_s, s > 0$) sämtliche Mengen des Wahrscheinlichkeitsmaßes Null enthält.
Später werden wir immer voraussetzen, dass wir mit einer vollständigen Filtration zu tun haben.
3. *eine rechtsseitig stetige Filtration*, falls für alle $t \geq 0$ $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.
4. *eine natürliche Filtration* für einen stochastischen Prozess $\{X(t), t \geq 0\}$, falls sie durch die Vergangenheit des Prozesses bis zum Zeitpunkt $t \geq 0$ erzeugt wird, d.h. für alle $t \geq 0$ \mathcal{F}_t ist die kleinste σ -Algebra ($\subset \mathcal{F}_t$), die Mengen $\{\omega \in \Omega : (X(t_1), \dots, X(t_n))^T \subset B\}$ enthält, für alle $n \in \mathbb{N}, 0 \leq t_1, \dots, t_n \leq t, B \in \mathcal{B}(\mathbb{R}^n)$.

Eine Zufallsvariable $\tau : \Omega \rightarrow \mathbb{R}_+$ heißt *Stoppzeit* (bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$), falls für alle $t \geq 0$ $\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$, d.h. aus den Beobachtungen des Prozesses X (bis natürlicher Filtration $\{\mathcal{F}_t, t \geq 0\}$) kann man beurteilen, ob der Moment τ eingetreten ist.

Lemma 5.1.1

Sei $\{\mathcal{F}_t, t \geq 0\}$ eine rechtsseitig stetige Filtration. τ ist eine Stoppzeit bzgl. $\{\mathcal{F}_t, t \geq 0\}$ genau dann, wenn $\underbrace{\{\tau < t\}}_{\{\omega \in \Omega : \tau(\omega) \leq t\}} \in \mathcal{F}_t$, für alle $t \geq 0$.

Beweis „ \Leftarrow “

Sei $\{\tau < t\} \in \mathcal{F}_t, t \geq 0$. Zu zeigen: $\{\tau \leq t\} \in \mathcal{F}_t$.

$$\{\tau \leq t\} = \bigcap_{s \in (t, t+\varepsilon)} \{\tau < s\} \text{ für alle } \varepsilon > 0 \Rightarrow \{\tau \leq t\} \in \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$$

„ \Rightarrow “

Zu zeigen: $\{\tau \leq t\} \in \mathcal{F}_t, t \geq 0 \Rightarrow \{\tau < t\} \in \mathcal{F}_t, t \geq 0$.

$$\{\tau < t\} = \bigcup_{s \in (0, t)} \{\tau \leq t - s\} \in \bigcup_{s \in (0, t)} \mathcal{F}_{t-s} \subset \mathcal{F}_t \quad \square$$

Definition 5.1.2

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\{\mathcal{F}_t, t \geq 0\}$ eine Filtration ($\mathcal{F}_t \subset \mathcal{F}, t \geq 0$) und $X = \{X(t), t \geq 0\}$ ein stochastischer Prozess auf $(\Omega, \mathcal{F}, \mathbb{P})$. X ist adaptiert bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$, falls $X(t)$ \mathcal{F}_t -meßbar ist, für alle $t \geq 0$, d.h., für alle $B \subset \mathcal{B}(\mathbb{R})$ $\{X(t) \in B\} \in \mathcal{F}_t$.

Definition 5.1.3

Der Zeitpunkt $\tau_B(\omega) = \inf\{t \geq 0 : X(t) \in B\}$, $\omega \in \Omega$, heißt *Ersterreichungszeit* der Menge $B \in \mathcal{B}(\mathbb{R})$ durch den stochastischen Prozess $X = \{X(t), t \geq 0\}$ (engl. first passage time, first entrance/hitting time).

Theorem 5.1.1

Sei $\{\mathcal{F}_t, t \geq 0\}$ eine rechtsseitig stetige Filtration und $X = \{X(t), t \geq 0\}$ ein adaptierter (bzgl. $\{\mathcal{F}_t, t \geq 0\}$) càdlàg-Prozess. Für offenes $B \subset \mathbb{R}$ ist τ_B eine Stoppzeit. Falls B abgeschlossen ist, dann ist $\tilde{\tau}_B(\omega) = \inf\{t \geq 0 : X(t) \in B \text{ oder } X(t-) \in B\}$ eine Stoppzeit, wobei $X(t-) = \lim_{s \uparrow t} X(s)$.

Beweis 1. Sei $B \in \mathcal{B}(\mathbb{R})$ offen.

Wegen Lemma 5.1.1 genügt zu zeigen, dass $\{\tau_B < t\} \in \mathcal{F}_t, t \geq 0$. Wegen rechtsseitiger Stetigkeit der Trajektorien von X gilt:

$$\{\tau_B < t\} = \cup_{s \in \mathbb{Q} \cap (0, t)} \{X(s) \in B\} \in \cup_{s \in \mathbb{Q} \cap (0, t)} \mathcal{F}_s \subseteq \mathcal{F}_t, \quad \text{weil } \mathcal{F}_s \subseteq \mathcal{F}_t, s < t.$$

2. Sei $B \in \mathcal{B}(\mathbb{R})$ abgeschlossen.

Für alle $\varepsilon > 0$. Sei $B_\varepsilon = \{x \in \mathbb{R} : d(x, B) < \varepsilon\}$ – Parallelmenge von B , wobei $d(x, B) = \inf_{y \in B} |x - y|$. B_ε ist offen, für alle $t \geq 0$.

$\{\tilde{\tau}_B \leq t\} = \{X(t) \in B\} \cup \cap_{n \geq 1, s \in \mathbb{Q} \cap (0, t)} \cup \{X(s) \in B_{\frac{1}{n}}\} \in \mathcal{F}_t$, weil X adaptiert bzgl. $\{\mathcal{F}_t, t \geq 0\}$ ist. □

Lemma 5.1.2

Seien τ_1, τ_2 Stoppzeiten bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$. Dann sind $\min\{\tau_1, \tau_2\}$, $\tau_1 + \tau_2$ und $\alpha\tau_1$, $\alpha \geq 1$, Stoppzeiten (bzgl. $\{\mathcal{F}_t, t \geq 0\}$).

Beweis Für alle $t \geq 0$ gilt:

$$\{\min\{\tau_1, \tau_2\} \leq t\} = \underbrace{\{\tau_1 \leq t\}}_{\in \mathcal{F}_t} \cup \underbrace{\{\tau_2 \leq t\}}_{\in \mathcal{F}_t} \in \mathcal{F}_t,$$

$$\{\max\{\tau_1, \tau_2\} \leq t\} = \{\tau_1 \leq t\} \cap \{\tau_2 \leq t\} \in \mathcal{F}_t,$$

$$\{\alpha\tau_1 \leq t\} = \{\tau_1 \leq \frac{t}{\alpha}\} \in \mathcal{F}_{\frac{t}{\alpha}} \subset \mathcal{F}_t, \text{ weil } \frac{t}{\alpha} \leq t,$$

$$\{\tau_1 + \tau_2 \leq t\} = \underbrace{\{\tau_1 > t\}}_{\in \mathcal{F}_t} \cup \underbrace{\{\tau_2 > t\}}_{\in \mathcal{F}_t} \cup \underbrace{\{\tau_1 \geq t, \tau_2 > 0\}}_{\mathcal{F}_t} \cup \{0 < \tau_2 < t, \tau_1 - \tau_2 > t\},$$

Zu zeigen: $\{0 < \tau_2 < t, \tau_1 - \tau_2 > t\} \in \mathcal{F}_t$.

$$\{0 < \tau_2 < t, \tau_1 - \tau_2 > t\} = \cup_{s \in \mathbb{Q} \cap (0, t)} \{s < \tau_1 < t, \tau_2 > t - s\} \in \mathcal{F}_t \quad \square$$

Theorem 5.1.2

Sei τ eine f.s. endliche Stoppzeit bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$ auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, d.h. $\mathbb{P}(\tau = \infty) = 1$. Dann existiert eine Folge von diskreten Stoppzeiten $\{\tau_n\}_{n \in \mathbb{N}}$, $\tau_1 \geq \tau_2 \geq \tau_3 \geq \dots$, so dass $\tau_n \downarrow \tau, n \rightarrow \infty$ f.s.

Beweis Für alle $n \in \mathbb{N}$ sei

$$\tau_n = \begin{cases} 0, & \text{falls } \tau(\omega) = 0 \\ \frac{k+1}{2^n}, & \text{falls } \frac{k}{2^n} < \tau(\omega) \leq \frac{k+1}{2^n}, \text{ für ein } k \in \mathbb{N}_0 \end{cases}$$

Für alle $t \geq 0$ und für alle $n \in \mathbb{N} \exists k \in \mathbb{N}_0 : \frac{k}{2^n} \leq t \leq \frac{k+1}{2^n}$ gilt $\{\tau_n \leq t\} = \{\tau_n \leq \frac{k}{2^n}\} = \{\tau \leq \frac{k}{2^n}\} \in \mathcal{F}_{\frac{k}{2^n}} \subset \mathcal{F}_t \Rightarrow \tau_n$ ist eine Stoppzeit. Es ist also klar, dass $\tau_n \downarrow \tau, n \rightarrow \infty$ f.s. □

Folgerung 5.1.1

Sei τ eine f.s. endliche Stoppzeit bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$ und $X = \{X(t), t \geq 0\}$ ein càdlàg-Prozess über $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{F}_t \subset \mathcal{F}$ für alle $t \geq 0$. Dann ist $X(\omega, \tau(\omega))$, $\omega \in \Omega$, eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$.

Beweis Zu zeigen: $X(\tau) : \Omega \rightarrow \mathbb{R}$ meßbar, d.h. für alle $B \in \mathcal{B}(\mathbb{R})$ $\{X(\tau) \in B\} \in \mathcal{F}$. Sei $\tau_n \downarrow \tau$, $n \rightarrow \infty$ wie im Satz 5.1.2. Da X càdlàg ist, gilt $X(\tau_n) \xrightarrow{n \rightarrow \infty} X(\tau)$ f.s. $X(\tau)$ ist dann \mathcal{F} -meßbar als Grenzwert von $X(\tau_n)$, die ihrerseits \mathcal{F} -meßbar sind. Für alle $B \in \mathcal{B}(\mathbb{R})$ gilt

$$\{X(\tau_n) \in B\} = \underbrace{\cup_{k=0}^{\infty} \{\tau_n = \frac{k}{2^n}\}}_{\in \mathcal{F}} \cap \underbrace{\{X(\frac{k}{2^n}) \in B\}}_{\in \mathcal{F}} \in \mathcal{F}$$

□

5.2 (Sub-, Super-)Martingale**Definition 5.2.1**

Sei $X = \{X(t), t \geq 0\}$ ein stochastischer Prozess adaptiert bzgl. einer Filtration $\{\mathcal{F}_t, t \geq 0\}$, $\mathcal{F}_t \subset \mathcal{F}$, $t \geq 0$, auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}|X(t)| < \infty$, $t \geq 0$. X heißt Martingal (bzw. Sub- oder Supermartingal), falls $\mathbb{E}(X(t) | \mathcal{F}_s) \stackrel{\cong}{=} X(s)$ für alle $s, t \geq 0$ mit $t \geq s$: $\Rightarrow \mathbb{E}(X(t)) = \mathbb{E}(X(s)) = \text{const}$ für alle s, t .

Beispiele

Sehr oft werden Martingale auf Basis eines stochastischen Prozesses $Y = \{Y(t), t \geq 0\}$ wie folgt konstruiert: $X(t) = Y(t) - \mathbb{E}Y(t)$.

1. *Poisson-Prozess*

Sei $Y = \{Y(t), t \geq 0\}$ der homogene Poisson-Prozess mit Intensität $\lambda > 0$. $\mathbb{E}Y(t) = \text{var}Y(t) = \lambda t$, weil $Y(t) \sim \text{Pois}(\lambda t)$, $t \geq 0$.

a) $X(t) = Y(t) - \lambda t$, $t \geq 0 \Rightarrow X(t)$ ist ein Martingal bzgl. natürlicher Filtration $\{\mathcal{F}_s, s \geq 0\}$.

$$\begin{aligned} \mathbb{E}(X(t) | \mathcal{F}_s)_{s \leq t} &= \mathbb{E}(Y(t) - \lambda t - (Y(s) - \lambda s + (Y(s) - \lambda s)) | \mathcal{F}_s) \\ &= Y(s) - \lambda s + \mathbb{E}(Y(t) - Y(s) - \lambda(t-s) | \mathcal{F}_s) \\ &= Y(s) - \lambda s + \mathbb{E}(Y(t) - Y(s)) + Y(s) - \lambda s \\ &= Y(s) - \lambda s + \underbrace{\mathbb{E}(Y(t-s))}_{=\lambda(t-s)} - \lambda(t-s) \\ &= Y(s) - \lambda \stackrel{f.s.}{=} X(s) \end{aligned}$$

b) $X'(t) = X^2(t) - \lambda t$, $t \geq 0 \Rightarrow X'(t)$ ist ein Martingal bzgl. $\{\mathcal{F}_s, s \geq 0\}$.

$$\begin{aligned} \mathbb{E}(X'(t) | \mathcal{F}_s) &= \mathbb{E}(X^2(t) - \lambda t | \mathcal{F}_s) = \mathbb{E}((X(t) - X(s) + X(s))^2 - \lambda t | \mathcal{F}_s) \\ &= \mathbb{E}((X(t) - X(s))^2 + 2((X(t) - X(s))X(s)) + X^2(s) - \lambda s - \lambda(t-s) | \mathcal{F}_s) \\ &= X'(s) + \underbrace{\mathbb{E}((X(t) - X(s))^2)}_{=\text{var}(Y(t)-Y(s))=\lambda(t-s)} + 2X(s) \underbrace{\mathbb{E}(X(t) - X(s))}_{=0} - \lambda(t-s) \\ &\stackrel{f.s.}{=} X'(s), \quad s \leq t. \end{aligned}$$

2. *Zusammengesetzter Poisson-Prozess*

$Y(t) = \sum_{i=1}^{N(t)} U_i$, $t \geq 0$, N – homogener Poisson-Prozess mit Intensität $\lambda > 0$, U_i – unabhängige identisch verteilte Zufallsvariablen, $E|U_i| < \infty$, $\{U_i\}$ unabhängig von N .
 $X(t) = Y(t) - EY(t) = Y(t) - \lambda t E U_1$, $t \geq 0$.

Aufgabe 5.2.1

Zeigen Sie, dass $X = \{X(t), t \geq 0\}$ ein Martingal bzgl. der natürlichen Filtration ist.

3. *Wiener-Prozess*

Sei $W = \{W(t), t \geq 0\}$ ein Wiener-Prozess, $\{\mathcal{F}_s, s \geq 0\}$ sei die natürliche Filtration.

a) $Y = \{Y(t), t \geq 0\}$

$Y(t) = W^2(t) - EW^2(t) = W^2(t) - t$, $t \geq 0$, ist ein Martingal bzgl. $\{\mathcal{F}_s, s \geq 0\}$.

$$\begin{aligned} E(Y(t) | \mathcal{F}_s) &= E((W(t) - W(s) + W(s))^2 - s - (t - s) | \mathcal{F}_s) \\ &= \text{siehe Beispiel 1b, benutze die Unabhängigkeit und Stationarität der Zuwächse} \\ &= W^2(s) - s \stackrel{f.s.}{=} Y(s), \quad s \leq t. \end{aligned}$$

b) $Y'(t) = e^{uW(t) - u^2 \frac{t}{2}}$, $t \geq 0$ und ein fixiertes $u \in \mathbb{R}$.

$E|Y'(t)| = e^{-u^2 \frac{t}{2}} E e^{uW(t)} = e^{-u^2 \frac{t}{2}} e^{u^2 \frac{t}{2}} = 1 < \infty$. Zeigen wir, dass $Y' = \{Y'(t), t \geq 0\}$ ein Martingal bzgl. $\{\mathcal{F}_s, s \geq 0\}$ ist.

$$\begin{aligned} E(Y'(t) | \mathcal{F}_s) &= E(e^{u(W(t) - W(s) + W(s)) - u^2 \frac{s}{2} - u^2 \frac{(t-s)}{2}} | \mathcal{F}_s) \\ &= \underbrace{e^{-u^2 \frac{s}{2}} e^{uW(s)}}_{=Y'(s)} e^{-u^2 \frac{(t-s)}{2}} \underbrace{E(e^{u(W(t) - W(s))} | \mathcal{F}_s)}_{=E(e^{uW(t-s)}) = e^{u^2 \frac{(t-s)}{2}}} \\ &= Y'(s) e^{-u^2 \frac{(t-s)}{2}} e^{u^2 \frac{(t-s)}{2}} = Y'(s), \quad s \leq t. \end{aligned}$$

4. *Abgeschlossenes Martingal*

Sei X eine Zufallsvariable (auf (Ω, \mathcal{F}, P)) mit $E|X| < \infty$. Sei $\{\mathcal{F}_s, s \geq 0\}$ eine Filtration auf (Ω, \mathcal{F}, P) .

$Y(t) = E(X | \mathcal{F}_t)$, $t \geq 0$. $Y = \{Y(t), t \geq 0\}$ ist ein Martingal.

$$E|Y(t)| = E|E(X | \mathcal{F}_t)| \leq E(E|X| | \mathcal{F}_t) = E|X| < \infty, \quad t \geq 0.$$

$$E(Y(t) | \mathcal{F}_s) = E(E(X | \mathcal{F}_t) | \mathcal{F}_s) = E(X | \mathcal{F}_s) \stackrel{f.s.}{=} Y(s), \quad s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t.$$

5. *Lèvy-Prozesse*

Sei $X = \{X(t), t \geq 0\}$ ein Lèvy-Prozess mit Lèvy-Exponenten η und natürlicher Filtration $\{\mathcal{F}_s, s \geq 0\}$.

a) Falls $E|X(1)| < \infty$, definiere $Y(t) = X(t) - \underbrace{tEX(1)}_{=EX(t)}$, $t \geq 0$. Es kann wie in obigen

Fällen gezeigt werden, dass $Y = \{Y(t), t \geq 0\}$ ein Martingal bzgl. der Filtration $\{\mathcal{F}_s, s \geq 0\}$ ist.

b) Benutze die Kombination aus Beispiel 3b – normiere die charakteristische Funktion von $X(t)$ ohne Erwartung durch deren Wert. $Y(t) = \frac{e^{iuX(t)}}{\varphi_{X(t)}(u)} = \frac{e^{iuX(t)}}{t\eta(u)} = e^{iuX(t) - t\eta(u)}$,
 $t \geq 0$, $u \in \mathbb{R}$.

Zu zeigen: $Y = \{Y(t), t \geq 0\}$ ist ein komplexwertiges Martingal.
 $E|Y(t)| = |e^{-t\eta(u)}| < \infty$, weil $\eta : \mathbb{R}_+ \rightarrow \mathbb{C}$. $EY(t) = 1, t \geq 0$.

$$\begin{aligned} E(Y(t) | \mathcal{F}_s) &= E(e^{iu(X(t)-X(s))(t-s)\eta(u)} e^{iuX(s)-s\eta(u)} | \mathcal{F}_s) \\ &= e^{iuX(s)-s\eta(u)} e^{-(t-s)\eta(u)} E(e^{iu(X(t)-X(s))}) \\ &= Y(s) e^{-(t-s)\eta(u)} e^{(t-s)\eta(u)} \stackrel{f.s.}{=} Y(s) \end{aligned}$$

6. Submartingale/Supermartingale

Jeder integrierbare stochastische Prozess $X = \{X(t), t \geq 0\}$, der bzgl. einer Filtration $\{\mathcal{F}_s, s \geq 0\}$ adaptiert ist und f.s. monoton nichtfallende (bzw. nichtsteigende) Trajektorien besitzt, ist ein Sub- (bzw. ein Super-)Martingal.

Tatsächlich, es gilt $X(t) \stackrel{f.s.}{\geq} X(s), t \geq s \Rightarrow E(X(t) | \mathcal{F}_s) \stackrel{f.s.}{\geq} E(X(s) | \mathcal{F}_s) \stackrel{f.s.}{=} X(s)$.
 Insbesondere ist jeder Subordinator ein Submartingal.

Lemma 5.2.1

Sei $X = \{X(t), t \geq 0\}$ ein stochastischer Prozess, der bzgl. einer Filtration $\{\mathcal{F}_t, t \geq 0\}$ adaptiert ist und sei $f : \mathbb{R} \rightarrow \mathbb{R}$ konvex, so dass $E|f(X(t))| < \infty, t \geq 0$. Dann ist $Y = \{f(X(t), t \geq 0\}$ ein Submartingal, falls

- X ein Submartingal ist, oder
- X ein Submartingal und f monoton nichtfallend ist.

Beweis Benutze die Ungleichung von Jensen für die bedingten Erwartungen. $E(f(X(t)) | \mathcal{F}_s) \geq f(\underbrace{E(X(t) | \mathcal{F}_s)}_{\geq X(s)}) \geq f(X(s))$, weil f monoton nichtfallend (Fall b)) oder es gilt die Gleichung

(Fall a)). □

5.3 Gleichgradige Integrierbarkeit

Frage: Man weiss, dass im Allg. aus $X_n \xrightarrow[n \rightarrow \infty]{f.s.} X$ nicht $X_n \xrightarrow[n \rightarrow \infty]{L_1} X$ folgt. Hier sind X, X_1, X_2, \dots

Zufallsvariablen, definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Wann gilt „ $X_n \xrightarrow[n \rightarrow \infty]{L_1} X$ “
 \Rightarrow „ $X_n \xrightarrow[n \rightarrow \infty]{L_1} X$ “? Die Antwort darauf liefert der Begriff der gleichgradigen Integrierbarkeit von $\{X_n, n \in \mathbb{N}\}$.

Definition 5.3.1

Die Folge $\{X_n, n \in \mathbb{N}\}$ von Zufallsvariablen heißt gleichgradig integrierbar, falls $E|X_n| < \infty, n \in \mathbb{N}$, und $\sup_n E(|X_n|1(|X_n| > \varepsilon)) \xrightarrow[\varepsilon \rightarrow +\infty]{} 0$.

Lemma 5.3.1

Die Folge $\{X_n, n \in \mathbb{N}\}$ von Zufallsvariablen ist gleichgradig integrierbar genau dann, wenn

- $\sup_n E|X_n| < \infty$ (gleichmäßige Beschränktheit),
- wenn es für jedes $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass $E(|X_n|1(A)) < \varepsilon$ für alle $n \in \mathbb{N}$ und alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) < \delta$.

Beweis Sei $\{X_n\}$ eine Folge von Zufallsvariablen.

Es ist zu zeigen, dass

$$\sup_n \mathbf{E}(|X_n| \mathbf{1}(|X_n| > x)) \xrightarrow{n \rightarrow \infty} 0 \iff \begin{array}{l} 1) \sup_n \mathbf{E}|X_n| < \infty \\ 2) \forall \varepsilon > 0 \exists \delta > 0 : \mathbf{E}(|X_n| \mathbf{1}(A)) < \varepsilon \forall A \in \mathcal{F} : \mathbf{P}(A) < \delta \end{array}$$

„ \Leftarrow “

$A_n = \{|X_n| > x\}$. Aus der Markov-Ungleichung: $\mathbf{P}(A_n) \leq \frac{1}{x} \mathbf{E}|X_n|$ für alle $n \Rightarrow \sup_n \mathbf{P}(A_n) \leq \frac{1}{x} \sup_n \mathbf{E}|X_n| \leq \frac{\varepsilon}{x} \xrightarrow{x \rightarrow \infty} 0 \Rightarrow \exists N > 0 : \forall x > N \mathbf{P}(A_n) < \delta \stackrel{2)}{\Rightarrow} \sup_n \mathbf{E}(|X_n| \mathbf{1}(A_n)) \leq \varepsilon \Rightarrow$ weil $\varepsilon > 0$ beliebig klein gewählt werden kann $\Rightarrow \sup_n \mathbf{E}(|X_n| \mathbf{1}(|X_n| > x)) \xrightarrow{x \rightarrow \infty} 0$.

„ \Rightarrow “

1.

$$\begin{aligned} \sup_n \mathbf{E}|X_n| &\leq \sup_n (\mathbf{E}(|X_n| \mathbf{1}(|X_n| > x)) + \mathbf{E}(|X_n| \mathbf{1}(|X_n| \leq x))) \\ &\leq \sup_n (\mathbf{E}(|X_n| \mathbf{1}(|X_n| > x)) + x \underbrace{\mathbf{P}(|X_n| \leq x)}_{\leq 1}) \\ &\leq \varepsilon + x < \infty \end{aligned}$$

2.

$$\begin{aligned} \mathbf{E}(|X_n| \mathbf{1}(A)) &= \underbrace{\mathbf{E}(|X_n| \mathbf{1}(|X_n| \leq x) \mathbf{1}(A))}_{\leq x} + \underbrace{\mathbf{E}(|X_n| \mathbf{1}(|X_n| > x) \mathbf{1}(A))}_{\leq 1} \\ &\leq \underbrace{x \mathbf{P}(A)}_{\leq \frac{\varepsilon}{2}} + \underbrace{\mathbf{E}(|X_n| \mathbf{1}(|X_n| > x))}_{\leq \frac{\varepsilon}{2}}, \end{aligned}$$

für alle $\varepsilon > 0 \exists x > 0$, so dass $\mathbf{E}(|X_n| \mathbf{1}(|X_n| > x)) < \frac{\varepsilon}{2}$ wegen gleichgradiger Integrierbarkeit. Wähle $\delta > 0$, $x\delta < \frac{\varepsilon}{2}$.

□

Lemma 5.3.2

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit $\mathbf{E}|X_n| < \infty$, $n \in \mathbb{N}$, $X_n \xrightarrow{n \rightarrow \infty} X$ f.s. $X_n \xrightarrow{n \rightarrow \infty} X$ L^1 genau dann, wenn $\{X_n\}_{n \in \mathbb{N}}$ gleichgradig integrierbar ist. Insbesondere folgt aus $X_n \xrightarrow{n \rightarrow \infty} X$ L^1 die Konvergenz $\mathbf{E}X_n \xrightarrow{n \rightarrow \infty} \mathbf{E}X$.

Beweis Sei $\{X_n\}_{n \in \mathbb{N}}$ gleichgradig integrierbar. Es ist zu zeigen, dass $\mathbf{E}|X_n - X| \xrightarrow{n \rightarrow \infty} 0$.

$$X_n \xrightarrow{n \rightarrow \infty} X \text{ f.s.} \Rightarrow X_n \xrightarrow{n \rightarrow \infty} X \text{ P} \Rightarrow \mathbf{P}(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \text{ für alle } \varepsilon.$$

$$\begin{aligned} \mathbf{E}|X_n - X| &\leq \mathbf{E}(|X_n - X| \mathbf{1}(|X_n - X| \leq \varepsilon)) + \mathbf{E}(|X_n - X| \mathbf{1}(|X_n - X| > \varepsilon)) \\ &\leq \varepsilon + \underbrace{\mathbf{E}(|X_n - X| \mathbf{1}(|X_n - X| > \varepsilon))}_{\xrightarrow{n \rightarrow \infty} 0, \text{ wegen Lemma 5.3.1, 2) für } A_n = \{|X_n - X| > \varepsilon\}} \\ &\quad + \underbrace{\mathbf{E}(|X| \mathbf{1}(|X_n - X| > \varepsilon))}_{\xrightarrow{n \rightarrow \infty} 0, \text{ weil } \mathbf{E}|X| < \infty, \text{ nach dem Satz von Lebesgue}} \end{aligned}$$

Warum $E|X| < \infty$? Es gilt $X_n \xrightarrow[n \rightarrow \infty]{f.s.} X$, aus Lemma 5.3.1, 1): $\sup_n E|X_n| < \infty$. Nach dem Lemma von Fatou gilt $E|X| < \infty$, denn für alle $\varepsilon_0 > 0 \exists N$: für alle $n > N$ $|X_n - X| < \varepsilon_0 \Rightarrow X_n \leq \eta_1, \eta_1 = |X| + \varepsilon_0, X_n \geq \eta_2, \eta_2 = |X| - \varepsilon_0$, für alle $n > N$. $E|X| = E|\lim_{n \rightarrow \infty} X_n| \leq \lim_{n \rightarrow \infty} E|X_n| < \infty$. Somit haben wir bewiesen, dass $X_n \xrightarrow[n \rightarrow \infty]{L^1} X$.

Jetzt sei $E|X_n - X| \xrightarrow[n \rightarrow \infty]{} 0$. Es sind die Eigenschaften 1) und 2) des Lemmas 5.3.1 zu zeigen.

1. $\sup_n E|X_n| \leq \sup_n E|X_n - X| + E|X| < \infty$, wegen $X_n \xrightarrow{L^1} X$.
2. Für alle $A \subset \mathcal{F}, P(A) \leq \delta$: $E(|X_n|1(A)) \leq E(|X_n - X|1(A)) + E(|X|1(A)) \leq \underbrace{E|X_n - X|}_{\leq 1} + \underbrace{E|X|}_{< \frac{\varepsilon}{2}} = \varepsilon$ bei entsprechender Wahl von δ , weil $E|X| < \infty$ und weil für alle $\varepsilon > 0 \exists N$, so dass für alle $n > N$ $E|X_n - X| < \frac{\varepsilon}{2}$.

□

5.4 Gestoppte Martingale

Bezeichnung: $x_+ = (x)_+ = \max(x, 0), x \in \mathbb{R}$.

Theorem 5.4.1 (Ungleichung von Doob):

Sei $X = \{X(t), t \geq 0\}$ ein càdlàg-Prozess, adaptiert bzgl. der Filtration $\{\mathcal{F}, t \geq 0\}$. Sei X ein Submartingal. Dann gilt für beliebige $t > 0$ und beliebige $x > 0$:

$$P\left(\sup_{0 \leq s \leq t} X(s) > x\right) \leq \frac{E(X(t))_+}{x}$$

Beweis O.B.d.A. setze $X(t) \geq 0, t \geq 0$ f.s. voraus.

$P(\sup_{0 \leq s \leq t} X(s) > x) = P(\sup_{0 \leq s \leq t} ((X(s))_+ > x))$, für alle $t \geq 0, x > 0$. $A = \{\sup_{t_1, \dots, t_n} X(s) > x\}, 0 \leq t_1 < t_2 < \dots < t_n \leq t$ - beliebige Zeitpunkte. $A = \cup_{k=1}^n A_k$,

$$\begin{aligned} A_1 &= \{X(t_1) > x\} \\ A_2 &= \{X(t_2) \leq x, X(t_2) > x\} \\ &\vdots \\ A_k &= \{X(t_1) \leq x, X(t_1) \leq x, \dots, X(t_{k-1}) \leq x, X(t_k) > x\}, \end{aligned}$$

$k = 2, \dots, n, A_i \cap A_j = \emptyset, i \neq j$.

Es ist zu zeigen, dass $P(A) \leq \frac{E(X(t_n))}{x}$.

$E(X(t_n)) \geq E(X(t_n)1(A)) = \sum_{k=1}^n E(X(t_n)1(A_k)) \geq x \sum_{k=1}^n P(A_k) = xP(A), k = 1, \dots, n - 1$, weil X ein Martingal ist und daraus folgt, dass $E(X(t_n)1(A_k)) \geq E(X(t_k)1(A_k)) \geq E(x1(A_k)) = xP(A_k), k = 1, \dots, n - 1, t_n > t_k$.

Sei $B \subset [0, t]$ eine endliche Teilmenge, $0 \in B, t \in B \Rightarrow$ es wird ähnlich bewiesen, dass $P(\max_{s \in B} X(s) > x) \leq \frac{E X(t)}{x}$.

\mathbb{Q} ist dicht in $\mathbb{R} \Rightarrow [0, t] \cap \mathbb{Q} \cup \{t\} = \cup_{k=1}^\infty B_k, B_k \subset [0, t] \cap \mathbb{Q} \cup \{t\}$ endlich, $B_k \subset B_n, k < n$. Wegen Monotonie des Wahrscheinlichkeitsmaßes gilt:

$$\lim_{n \rightarrow \infty} P\left(\max_{s \in B} X(s) \geq x\right) = P\left(\cup_n \{\max_{s \in B_n} X(s) > x\}\right) = P\left(\sup_{s \in \cup_n B_n} X(s) > x\right) \leq \frac{E X(t)}{x}$$

Wegen der rechtsseitigen Stetigkeit der Pfade von X gilt $P(\sup_{0 \leq s \leq t} X(s) > x) \leq \frac{E X(t)}{x}$. □

Folgerung 5.4.1

Für Wiener-Prozess $W = \{W(t), t \geq 0\}$ betrachten wir den Wiener-Prozess mit negativer Drift: $Y(t) = W(t) - \mu t$, $\mu > 0$, $t \geq 0$. Aus dem Beispiel Nr.3 des Abschnitts 5.3 ist $X(t) = \exp\{u(Y(t) + t\mu) - \frac{u^2 t}{2}\}$, $t \geq 0$, ein Martingal bzgl. der natürlichen Filtration für W . Für $u = 2\mu$ gilt

$$X(t) = \exp\{2\mu Y(t)\}, \quad t \geq 0.$$

$$\begin{aligned} \mathbb{P}\left\{\sup_{0 \leq s \leq t} Y(s) > x\right\} &= \mathbb{P}\left\{\sup_{0 \leq s \leq t} e^{2\mu Y(s)} > e^{2\mu x}\right\} \leq \frac{\mathbb{E}e^{2\mu Y(t)}}{e^{2\mu x}} = e^{-2\mu x}, \quad x > 0 \\ \Rightarrow \lim_{t \rightarrow \infty} \mathbb{P}\left\{\sup_{0 \leq s \leq t} Y(s) > x\right\} &= \text{Aus Beispiel Nr.3 gilt } \mathbb{E}e^{2\mu Y(t)} = \mathbb{P}(\sup_{t \geq 0} Y(t) > x) \leq e^{-2\mu x}. \end{aligned}$$

Theorem 5.4.2

Sei $X = \{X(t), t \geq 0\}$ ein Martingal bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$ mit càdlàg-Pfaden. Falls $T : \Omega \rightarrow [0, \infty)$ eine endliche Stoppzeit bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$ ist, dann ist auch der stochastische Prozess $\{X_{T \wedge t}(t) \geq 0\}$ ein Martingal, das auch ein gestopptes Martingal genannt wird. Dabei ist $a \wedge b = \min\{a, b\}$.

Lemma 5.4.1

Sei $X = \{X(t), t \geq 0\}$ ein Martingal mit càdlàg-Trajektorien bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$. Sei T eine endliche Stoppzeit und sei $\{T_n\}_{n \in \mathbb{N}}$ die Folge von diskreten Stoppzeiten aus dem Theorem 5.1.2, für die $T_n \downarrow T$, $n \rightarrow \infty$, gilt. Dann ist $\{X(T_n \wedge t)\}_{n \in \mathbb{N}}$ gleichgradig integrierbar für jedes $t \geq 0$.

Beweis

$$T_n = \begin{cases} 0 & , \quad \text{falls } T = 0 \\ \frac{k+1}{2^n} & , \quad \text{falls } \frac{k}{2^n} < T \leq \frac{k+1}{2^n}, \text{ für ein } k \in \mathbb{N}_0 \end{cases}$$

1. Es ist zu zeigen: $\mathbb{E}|X(T_n \wedge t)| < \infty$ für alle n .
 $\mathbb{E}|X(T_n \wedge t)| \leq \sum_{k: \frac{k}{2^n} < t} \mathbb{E}|X(\frac{k}{2^n})| + \mathbb{E}|X(t)| < \infty$, weil X ein Martingal ist, also integrierbar.
2. Es ist zu zeigen: $\sup_n \mathbb{E}(|X(T_n \wedge t)| \mathbf{1}(\underbrace{|X(T_n \wedge t)| > x}_{A_n})) \xrightarrow{x \rightarrow \infty} 0$.

$$\begin{aligned} & \sup_n \mathbb{E}(|X(T_n \wedge t)| \mathbf{1}(A_n)) \\ &= \sup_n \left(\sum_{k: \frac{k}{2^n} < t} \mathbb{E} \left(\left| X \left(\frac{k}{2^n} \right) \right| \mathbf{1} \left(\left\{ T_n = \frac{k}{2^n} \right\} \cap A_n \right) \right) + \mathbb{E}(|X(t)| \mathbf{1}(T_n > t) \mathbf{1}(A_n)) \right) \\ &\leq \sup_n \left(\sum_{k: \frac{k}{2^n} < t} \mathbb{E} \left(|X(t)| \mathbf{1} \left(\left\{ T_n = \frac{k}{2^n} \right\} \cap A_n \right) \right) + \mathbb{E}(|X(t)| \mathbf{1}(\{T_n > t\} \cap A_n)) \right) \\ &= \sup_n \mathbb{E}(|X(t)| \mathbf{1}(A_n)) \leq \sup_n \mathbb{E}(|X(t)| \mathbf{1}(Y > x)) \\ &= \mathbb{E}(|X(t)| \mathbf{1}(Y > x)), \end{aligned}$$

wobei $\mathbf{1}(A_n) \leq \mathbf{1}(\underbrace{\sup_n |X(T_n \wedge t)| > x}_Y)$. Es ist zu zeigen: $\mathbb{P}(Y > x) \xrightarrow{n \rightarrow \infty} 0$ mit Hilfe von

der Ungleichung von Doob.

$$\mathbb{P}(Y > x) \leq \mathbb{P}(\sup_{0 \leq s \leq t} |X(s)| > x) \leq \frac{\mathbb{E}|X(t)|}{x} \xrightarrow{x \rightarrow +\infty} 0. \text{ Da } \mathbb{E}|X(t)| < \infty \text{ für alle}$$

$t \geq 0$ und $\mathbb{P}(Y > x) \xrightarrow{x \rightarrow \infty} 0$, ergibt sich $\mathbb{E}(|X(t)|1(Y > x)) \xrightarrow{n \rightarrow \infty} 0 \Rightarrow \sup_n \mathbb{E}|X(T_n \wedge t)1(A)| \xrightarrow{x \rightarrow \infty} 0$ $\{X(T_n \wedge t)\}_{n \in \mathbb{N}}$ ist gleichgradig integrierbar.

□

Beweis vom Theorem 5.4.2

Es ist zu zeigen, dass $\{X(T \wedge t), t \geq 0\}$ ein Martingal ist.

1. $\mathbb{E}|X(T \wedge t)| < \infty$ für alle $t \geq 0$. Wie in der Folgerung 5.1.1 approximiert man $T_n \downarrow T$, $n \rightarrow \infty \Rightarrow X(T_n \wedge t) \xrightarrow[n \rightarrow \infty]{f.s.} X(T \wedge t)$, da aber $\mathbb{E}|X(T_n \wedge t)| < \infty$ für alle n folgt $\mathbb{E}|X(T \wedge t)| < \infty$ wegen Lemma 5.4.1, weil aus der gleichgradigen Integrierbarkeit die L^1 -Konvergenz folgt.
2. *Martingal Eignenschaft*

Es ist zu zeigen:

$$\begin{aligned} \mathbb{E}(X(T \wedge t) | \mathcal{F}_s) &\stackrel{f.s.}{=} X(T \wedge s), \quad s \leq t \\ &\Downarrow \\ \mathbb{E}(X(T \wedge t)1(A)) &\stackrel{f.s.}{=} \mathbb{E}(X(T \wedge s)1(A)), \quad A \in \mathcal{F}_s \end{aligned}$$

Zunächst zeigen wir, dass $\mathbb{E}(|X(T_n \wedge t)|1(A)) = \mathbb{E}(|X(T_n \wedge s)|1(A))$, $A \in \mathcal{F}_s$, $n \in \mathbb{N}$. Seien $t_1, \dots, t_k \in (s, t)$ diskrete Werte, die T_n mit positiver Wahrscheinlichkeit in (s, t) annimmt.

$$\begin{aligned} \mathbb{E}(X(T_n \wedge t) | \mathcal{F}_s) &= \mathbb{E}(\mathbb{E}(X(T_n \wedge t) | \mathcal{F}_{t_k}) | \mathcal{F}_s) \\ &= \mathbb{E}(\underbrace{\mathbb{E}(X(T_n \wedge t)1(T_n \leq t_k) | \mathcal{F}_{t_k})}_{X(t_k)} | \mathcal{F}_s) \\ &\quad + \mathbb{E}(\underbrace{\mathbb{E}(X(T_n \wedge t)1(T_n > t_k) | \mathcal{F}_{t_k})}_{X(t)} | \mathcal{F}_s) \\ &= \mathbb{E}(X(t_k)1(T_n \leq t_k) | \mathcal{F}_s) + \mathbb{E}(1(T_n > t_k)\mathbb{E}(X(t) | \mathcal{F}_{t_k}) | \mathcal{F}_s) \\ &= \mathbb{E}(X(t_k \wedge T_n) | \mathcal{F}_s) = \dots = \mathbb{E}(X(t_{k-1} \wedge T_n) | \mathcal{F}_s) = \dots \\ &= \mathbb{E}(X(t_1 \wedge T_n) | \mathcal{F}_s) = \dots = \mathbb{E}(X(T_n \wedge s) | \mathcal{F}_s) \\ &\stackrel{f.s.}{=} X(T_n \wedge s) \end{aligned}$$

Da X càdlàg ist und $T_n \downarrow T$, $n \rightarrow \infty$, gilt $X(T_n \wedge t) \xrightarrow[n \rightarrow \infty]{f.s.} X(T \wedge t)$. Dazu sind $\{X(T_n \wedge t)\}_{n \in \mathbb{N}}$ gleichgradig integrierbar wegen L^1 -Konvergenz. Daraus folgt

$$\begin{aligned} \mathbb{E}(X(T_n \wedge t)1(A)) &= \mathbb{E}(X(T_n \wedge s)1(A)) \quad \text{für alle } A \in \mathcal{F}_s \\ \downarrow &\quad \quad \quad \downarrow \\ \mathbb{E}(X(T \wedge t)1(A)) &= \mathbb{E}(X(T \wedge s)1(A)) \end{aligned}$$

$\Rightarrow \{X(T \wedge t), t \geq 0\}$ ist ein Martingal.

□

Definition 5.4.1

Sei $T : \Omega \rightarrow \mathbb{R}_+$ eine Stoppzeit bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$, $\mathcal{F}_t \subset \mathcal{F}$, $t \geq 0$. Die „gestoppte“ σ -Algebra \mathcal{F}_T wird definiert durch $A \in \mathcal{F}_T \Rightarrow A \cap \{T \leq t\} \in \mathcal{F}_t$ für alle $t \geq 0$.

Lemma 5.4.2 1. Seien S, T – Stoppzeiten bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$, $S \stackrel{f.s.}{\leq} T$. Dann gilt $\mathcal{F}_S \subset \mathcal{F}_T$.

2. Sei $X = \{X(t), t \geq 0\}$ ein Martingal mit càdlàg-Trajektorien bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$ und sei T eine Stoppzeit bzgl. $\{\mathcal{F}_t, t \geq 0\}$. Dann ist $X(T)$ \mathcal{F}_T -meßbar.

Beweis 1. $A \in \mathcal{F}_S \Rightarrow A \cap \{S \leq t\} \in \mathcal{F}_t, t \geq 0$. $A \cap \{T \leq t\} = \underbrace{A \cap \{S \leq t\}}_{\in \mathcal{F}_t} \cap \underbrace{\{T \leq t\}}_{\in \mathcal{F}_t} \in \mathcal{F}_t$

für alle $t \geq 0 \Rightarrow A \in \mathcal{F}_T$.

2. $X(T) = g \circ f, f: \Omega \rightarrow \Omega \times \mathbb{R}_+, f(\omega) = (\omega, T(\omega)), g: \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}, g(\omega, s) = X(s, \omega)$.

Es ist zu zeigen: f - $\mathcal{F} \mid \mathcal{F} \times B_{\mathbb{R}_+}$ -meßbar, g - $\mathcal{F} \times B_{\mathbb{R}_+} \mid \mathcal{F}_T$ -meßbar $\Rightarrow g \circ f$ - $\mathcal{F} \mid \mathcal{F}_T$ -meßbar.

f - $\mathcal{F} \mid \mathcal{F} \times B_{\mathbb{R}_+}$ -meßbar ist offensichtlich, weil T eine Zufallsvariable ist. Betrachten wir die Einschränkung von $X = \{X(s), s \geq 0\}$ auf $s \in [0, t], t \geq 0$.

Es ist zu zeigen: $\{X(T) \in B\} \cap \{T \leq t\} \in \mathcal{F}_t$ für alle $t \geq 0, B \in \mathcal{B}(\mathbb{R})$.

X – càdlàg $\Rightarrow X(s, \omega) = X(0, \omega)1(s=0) + \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} X(t \frac{k}{2^n}, \omega)1(\frac{k-1}{2^n}t < s \leq \frac{k}{2^n}t) \Rightarrow X(s, \omega)$ ist $B_{[0,t]} \times \mathcal{F}_t$ -meßbar $\Rightarrow X(T)$ ist $\mathcal{F} \mid \mathcal{F}_T$ -meßbar. □

Theorem 5.4.3 (Optionales Sampling-Theorem):

Sei $X = \{X(t), t \geq 0\}$ ein Martingal mit càdlàg-Trajektorien bzgl. einer Filtration $\{\mathcal{F}_t, t \geq 0\}$ und Sei T eine endliche Stoppzeit bzgl. $\{\mathcal{F}_t, t \geq 0\} \Rightarrow \mathbf{E}(X(t) \mid \mathcal{F}_T) \stackrel{f.s.}{=} X(T \wedge t), t > 0$.

Beweis Zeigen wir zunächst, dass $\mathbf{E}(X(t) \mid \mathcal{F}_{T_n}) \stackrel{f.s.}{=} X(T_n \wedge t), t \geq 0, n \in \mathbb{N}$, wobei die diskrete Approximation von T ist, $T_n \downarrow T, n \rightarrow \infty$. Sei $t_1 \leq t_2 \leq \dots \leq t_k = t$ die Werte, die $T_n \wedge t$ mit positiven Wahrscheinlichkeiten annimmt. Es ist zu zeigen, dass für alle $A \in \mathcal{F}_{T_n}$ gilt: $\mathbf{E}(X(t)1(A)) = \mathbf{E}(X(T_n \wedge t)1(A))$.

$$\begin{aligned} (X(t) - X(T_n \wedge t))1(A) &= \sum_{i=1}^{k-1} X(t_k) - X(t_i)1(\{T_n \wedge t = t_i\} \cap A) \\ &= \sum_{i=2}^k (X(t_i) - X(t_{i-1}))1(A)1(\{T_n \wedge t < t_i\}) \\ \mathbf{E}((X(t) - X(T_n \wedge t))1(A)) &= \sum_{i=2}^k \mathbf{E}((X(t_i) - X(t_{i-1}))1(T_n \wedge t < t_i)1(A)) \\ &= \sum_{i=2}^k \mathbf{E}(\mathbf{E}(X(t_i) - X(t_{i-1}))1(T_n \wedge t < t_i)1(A) \mid \mathcal{F}_{t_{i-1}}) \\ &= \sum_{i=2}^k \mathbf{E}(1(T_n \wedge t < t_i)1(A))\mathbf{E}((X(t_i) - X(t_{i-1})) \mid \mathcal{F}_{t_{i-1}}) = 0 \end{aligned}$$

$\mathbf{E}(X(t) \mid \mathcal{F}_{T_n}) \stackrel{f.s.}{=} \mathbf{E}(X(T_n \wedge t) \mid \mathcal{F}_{T_n}) \stackrel{f.s.}{=} X(T_n \wedge t)$, denn $X(T_n)$ ist \mathcal{F}_{T_n} -meßbar. $T \leq T_n \Rightarrow \mathcal{F}_T \subseteq \mathcal{F}_{T_n}$. Da $\{X(T_n \wedge t)\}_{n \in \mathbb{N}}$ für $t \in [0, \infty)$ gleichgradig integrierbar ist, gilt

$$\mathbf{E}(X(t) \mid \mathcal{F}_T) = \mathbf{E}(X(t) \mid \mathcal{F}_{T_n}) = \lim_{n \rightarrow \infty} \mathbf{E}(X(T_n \wedge t) \mid \mathcal{F}_{T_n}) = \lim_{n \rightarrow \infty} X(T_n \wedge t) = X(T \wedge t),$$

weil X càdlàg ist. □

Folgerung 5.4.2

Sei $X = \{X(t), t \geq 0\}$ ein càdlàg-Martingal und seien S, T endliche Stoppzeiten, so dass $P(S \leq T) = 1$. Dann gilt $E(X(t \wedge T) | \mathcal{F}_s) \stackrel{f.s.}{=} E(X(S \wedge t)), t \geq 0$. Insbesondere gilt $E(X(T \wedge t)) = E(X(0))$.

Beweis X – Martingal. Aus dem Theorem 5.4.2 ist $\{X(T \wedge t), t > 0\}$ auch ein Martingal. Wende den Satz 5.4.3 an dieses treue Martingal an:

$$E(X(T \wedge t) | \mathcal{F}_s) \stackrel{f.s.}{=} X(T \wedge S \wedge t) \stackrel{f.s.}{=} X(S \wedge t),$$

weil $S \stackrel{f.s.}{\leq} T$. Setze $S = 0$, dann $E(E(X(T \wedge t) | \mathcal{F}_0)) = EX(0 \wedge t) = EX(0)$. \square

5.5 Lèvy-Prozesse und Martingale**Theorem 5.5.1**

Sei $X = \{X(t), t \geq 0\}$ ein Lèvy-Prozess mit Charakteristiken (a, b, ν) .

1. Es existiert eine càdlàg-Modifikation von $\tilde{X} = \{\tilde{X}(t), t \geq 0\}$ von X mit denselben Charakteristiken (a, b, ν) .
2. Die natürliche Filtration eines càdlàg-Lèvy-Prozesses ist rechtsseitig stetig.

Ohne Beweis

Theorem 5.5.2 (Regenerationssatz für Lèvy-Prozesse):

Sei $X = \{X(t), t > 0\}$ ein càdlàg-Lèvy-Prozess mit natürlicher Filtration $\{\mathcal{F}_t^X, t \geq 0\}$ und sei T eine endliche Stoppzeit bzgl. $\{\mathcal{F}_t^X, t \geq 0\}$. Der Prozess $Y = \{Y(t), t \geq 0\}$, gegeben durch $Y(t) = X(T + t) - X(T), t \geq 0$, ist ebenfalls ein Lèvy-Prozess, adaptiert bzgl. der Filtration $\{\mathcal{F}_{T+t}^X, t \geq 0\}$, der unabhängig von \mathcal{F}_T^X ist und dieselben Charakteristiken, wie X besitzt. T wird Regenerationszeitpunkt genannt, da $Y \stackrel{d}{=} X$, Y unabhängig von \mathcal{F}_T^X .

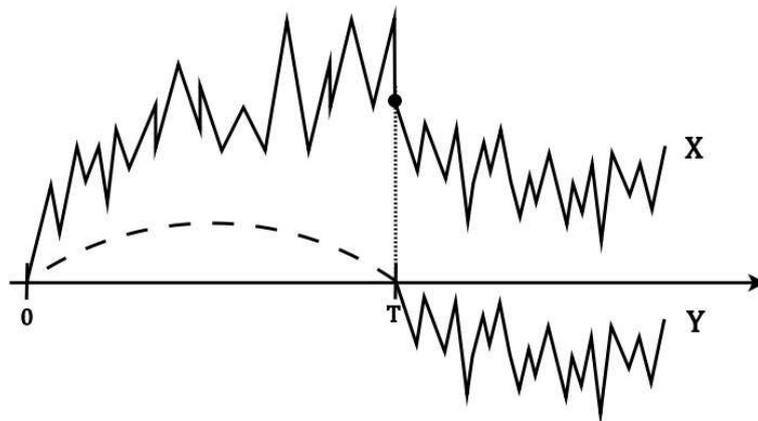


Abbildung 5.1:

Beweis 1. Annahme: Es $\exists c > 0$, so dass $\mathbb{P}(T \leq c) = 1$. Seien $u_1, \dots, u_n \in \mathbb{R}$. Nach dem Beispiel Nr.5 im Abschnitt 5.2 ist $\tilde{Y}_j = \{\tilde{Y}_j(t) = \exp\{iu_j X(t) - t\eta(u_j)\}, t \geq 0\}$, $j = 1, \dots, n$, ein komplexwertiges Martingal, wobei $\eta(\cdot)$ die Lèvy-Exponente von $X(t)$ ist. Seien $0 \leq t_0 < t_1 < \dots < t_n$ beliebige Zeitpunkte. Für alle $A \in \mathcal{F}_T^X$ gilt

$$\begin{aligned}
& \mathbb{E}(\mathbf{1}(A) \exp\{\sum_{j=1}^n iu_j(Y(t_j) - Y(t_{j-1}))\}) \stackrel{Z.z.}{=} \mathbb{P}(A) \mathbb{E}(\exp\{\sum_{j=1}^n iu_j(X(t_j) - X(t_{j-1}))\}) \\
& \mathbb{E}(\mathbf{1}(A) \exp\{\sum_{j=1}^n iu_j(Y(t_j) - Y(t_{j-1}))\}) \\
& = \mathbb{E}(\mathbf{1}(A) \exp\{\sum_{j=1}^n iu_j(X(T+t_j) - X(T) - X(T+t_{j-1}) - X(T))\}) \\
& = \mathbb{E}\left(\mathbf{1}(A) \prod_{j=1}^n \frac{\tilde{Y}_j(T+t_j)}{\tilde{Y}_j(T+t_{j-1})} \frac{\exp\{\eta(u_j)(T+t_j)\}}{\exp\{\eta(u_j)(T+t_{j-1})\}}\right) \\
& = \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}(A) \prod_{j=1}^n \frac{\tilde{Y}_j(T+t_j)}{\tilde{Y}_j(T+t_{j-1})} \exp\{(t_j - t_{j-1})\eta(u_j)\} \mid \mathcal{F}_{T+t_{j-1}}^X\right)\right) \\
& = \mathbb{E}\left(\mathbf{1}(A) \prod_{j=1}^{n-1} \frac{\tilde{Y}_j(T+t_j)}{\tilde{Y}_j(T+t_{j-1})} e^{(t_j - t_{j-1})\eta(u_j)} \frac{e^{(t_n - t_{n-1})\eta(u_n)}}{\tilde{Y}_n(T+t_{n-1})} \mathbb{E}(\tilde{Y}_n(T+t_n) \mid \mathcal{F}_{T+t_{n-1}}^X)\right) \\
& = \mathbb{E}\left(\mathbf{1}(A) \prod_{j=1}^{n-1} \frac{\tilde{Y}_j(T+t_j)}{\tilde{Y}_j(T+t_{j-1})} e^{(t_j - t_{j-1})\eta(u_j)} \dots e^{(t_n - t_{n-1})\eta(u_n)}\right) \\
& = \dots = \mathbb{E}(\mathbf{1}(A) \prod_{j=1}^n e^{(t_j - t_{j-1})\eta(u_j)}) = \mathbb{P}(A) \prod_{j=1}^n e^{(t_j - t_{j-1})\eta(u_j)} \\
& = \mathbb{P}(A) \mathbb{E}(\exp\{i \sum_{j=1}^n (u_j(X(t_j) - X(t_{j-1})))\})
\end{aligned}$$

□

Folgerung 5.5.1

$T_1 = T + t_n$, $S_1 = T + t_{n-1} \leq T_1$ f.s., $T_1, S_1 \leq t$, weil $t > c + t_n$, $T \geq c$.

Aufgabe 5.5.1

Zeigen Sie, dass aus $\mathbb{E}(\mathbf{1}(A) \exp\{\sum_{j=1}^n iu_j(Y(t_j) - Y(t_{j-1}))\}) = \mathbb{P}(A) \mathbb{E}(\exp\{\sum_{j=1}^n iu_j(X(t_j) - X(t_{j-1}))\})$ die Aussage des Satzes folgt.

5.6 Martingale und Wiener-Prozesse

Unser Ziel: Falls $W = \{W(t), t \geq 0\}$ ein Wiener-Prozess ist, dann gilt

$$\mathbb{P}(\max_{s \in [0, t]} W(s) > x) = \sqrt{\frac{2}{\pi t}} \int_x^{+\infty} e^{-\frac{y^2}{2t}} dy, \quad \text{für alle } x \geq 0.$$

Theorem 5.6.1 (Reflexionsprinzip):

Sei T eine beliebige Stopzeit bzgl. der natürlichen Filtration $\{\mathcal{F}_t^W, t \geq 0\}$. Sei $X = \{X(t), t \geq 0\}$ der reflektierte Wiener-Prozess zum Zeitpunkt T , d.h. $X(t) = W(T \wedge t) - (W(t) - W(T \wedge t))$, $t \geq 0$. Dann gilt $X \stackrel{d}{=} W$.

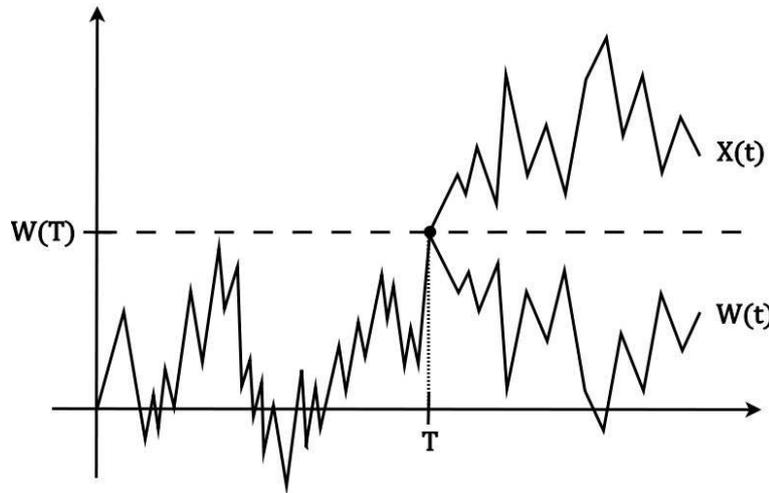


Abbildung 5.2:

Beweis Sei $X_1(t) = W(T \wedge t)$, $X_2(t) = W(T+t) - W(T)$, $t \geq 0$. Aus dem Theorem 5.5.2 folgt, dass X_2 unabhängig von (T_1, X_1) ist (W - Lévy-Prozess und T - Regenerationszeitpunkt). Es gilt $W(t) \stackrel{g}{=} X_1(t) + X_2((t-T)_+)$, $X(t) \stackrel{g}{=} X_1(t) - X_2((t-T)_+)$, $t \geq 0$. Aus dem Satz ?? folgt

$$\begin{array}{ccc} (T_1, X_1, X_2) & \stackrel{d}{=} & (T, X_1, -X_2) \\ \downarrow & & \downarrow \\ W & \stackrel{d}{=} & X \end{array}$$

□

Sei $W = \{W(t), t \geq 0\}$ ein Wiener-Prozess auf (Ω, \mathcal{F}, P) , sei $\{\mathcal{F}_t^W, t \geq 0\}$ die natürliche Filtration bzgl. W . Für $z \in \mathbb{R}$ sei $T_{\{z\}}^W = \inf\{t \geq 0 : W(t) = z\}$. $T_{\{z\}}^W := T_z^W$ ist eine f.s. endliche Stopzeit bzgl. $\{\mathcal{F}_t^W, t \geq 0\}$, $z > 0$. Offensichtlich, es gilt $\{\mathcal{F}_z^W \leq t\} \in \mathcal{F}_t^W$. Da W stetige Pfade (f.s.) besitzt, ist $\{\mathcal{F}_t^W, t \geq 0\}$ rechtsseitig stetig.

Folgerung 5.6.1

Sei $M_t = \max_{s \in [0,t]} W(s)$, $t \geq 0$. Dann gilt für alle $z > 0$, $y \geq 0$, dass $P(M_t \geq z, W(t) \leq z - y) = P(W(t) > y + z)$.

Beweis M_t sei eine Zufallsvariable, weil W stetige Pfade hat. $T := T_z^W$. Nach dem Theorem 5.6.1 gilt: für $Y(t) = W(T \wedge t) - (W(t) - W(T \wedge t))$, $t \geq 0$, $Y \stackrel{d}{=} W$ bzw. $\{T_z^W, W\} \stackrel{d}{=} \{T_z^Y, Y\}$, weil $W(t) = z$, $T_z^W = T_z^Y$. Deshalb

$$P(T \leq t, W(t) < z - y) = P(T_z^Y \leq t, Y(t) < z - y)$$

$\{T_z^Y \leq t\} \cap \{Y(t) < z - y\} = \{T_z^Y \leq t\} \cap \{2z - W(t) < z - y\}$. Falls $T = T_z^Y \leq t$, dann $Y(t) = W(T) - W(t) + W(T) = 2z - W(t)$ und daraus folgt

$$\mathbb{P}(T \leq t, W(t) < z - y) = \mathbb{P}(T \leq t, 2z - W(t) < z - y) = \mathbb{P}(T \leq t, W(t) > z + y) = \mathbb{P}(W(t) > z + y)$$

Per Definition im $T = T_z^W$ gilt:

$$\mathbb{P}(T \leq t, W(t) < z - y) = \mathbb{P}(M_t \geq z, W(t) < z - y) = \mathbb{P}(W(t) > y + z)$$

$$\Rightarrow T_z^W \leq t \iff \max_{s \in [0, t]} W(s) \geq z \quad \square$$

Theorem 5.6.2 (Verteilung des Maximums von W):

Für $t > 0$ und $x \geq 0$ gilt

$$\mathbb{P}(M_t > x) = \sqrt{\frac{2}{\pi t}} \int_x^\infty e^{-\frac{y^2}{2t}} dy$$

Beweis In Folgerung 5.6.1 setze $y = 0 \Rightarrow \mathbb{P}(M_t \geq z, W(t) < z) = \mathbb{P}(W(t) > z)$. Es gilt $\mathbb{P}(W(t) > z) = \mathbb{P}(W(t) \geq z)$ für alle t und alle z , weil $W(t) \sim \mathcal{N}(0, t)$, also atomfrei

$$\Rightarrow \mathbb{P}(M_t \geq z, W(t) < z) + \mathbb{P}(W(t) \geq z) = \mathbb{P}(W(t) > z) + \mathbb{P}(W(t) > z)$$

$$\Rightarrow \mathbb{P}(M_t \geq z, W(t) < z) + \mathbb{P}(M_t \geq z, W(t) \geq z) = \mathbb{P}(M_t \geq z) = 2\mathbb{P}(W(t) > z)$$

$$\Rightarrow \mathbb{P}(M_t > z) = 2\mathbb{P}(W(t) > z) = 2 \frac{1}{\sqrt{2\pi t}} \int_z^\infty e^{-\frac{y^2}{2t}} dy = \sqrt{\frac{2}{\pi t}} \int_z^\infty e^{-\frac{y^2}{2t}} dy \quad \square$$

Sei $X(t) = W(t) - t\mu$, $t \geq 0$, $\mu > 0$, der Wiener-Prozess mit negativer Drift. Betrachte $\mathbb{P}(\sup_{t \geq 0} X(t) > x) = e^{-2\mu x}$, $x \geq 0$.

Motivation Berechnung der Ruin-Wahrscheinlichkeiten in der Risikotheorie.

Annahmen Startkapital $x \geq 0$. Sei μ das Prämienvolumen per Zeiteinheit. $\Rightarrow \mu t$ - Prämienannahmen zum Zeitpunkt $t \geq 0$. Sei $W(t)$ der Prozess der Verluste (Preisentwicklung). $\Rightarrow Y(t) = x + t\mu - W(t)$ - Restkapital zum Zeitpunkt t . Die Wahrscheinlichkeit des Ruins ist $\mathbb{P}(\inf_{t \geq 0} Y(t) < 0) = \mathbb{P}(x - \sup_{t \geq 0} X(t) < 0) = \mathbb{P}(\sup_{t \geq 0} X(t) > x)$

Theorem 5.6.3

Es gilt

$$\mathbb{P}(\sup_{t \geq 0} X(t) > x) = e^{-2\mu x}, \quad x \geq 0, \mu > 0.$$

Beweis Sei $T = T_z^X = \inf\{t \geq 0 : X(t) = z\}$. Es ist bekannt, dass $Y(t) = \exp\{uX(t) - t(\frac{u^2}{2} - \mu u)\}$, $t \geq 0$, $u \geq 0$, ein Martingal ist. Sei $T' = T \wedge t$ - eine endliche Stoppzeit bzgl. $\{\mathcal{F}_t^X, t \geq 0\}$. Aus Folgerung 5.4.1: $\mathbb{E}Y(T') = \mathbb{E}Y(0) = \mathbb{E}e^0 = 1$

$$\Rightarrow \mathbb{E}(Y(T')1(T < t)) + \mathbb{E}(Y(T')1(T \geq t)) = \mathbb{E}(Y(T)1(T < t)) + \mathbb{E}(Y(T')1(t \geq t))$$

Es ist zu zeigen, dass $\mathbb{E}(Y(T')1(T \geq t)) \xrightarrow{t \rightarrow \infty} 0$.

Aus der Folgerung ?? ist bekannt, dass

$$\frac{W(t)}{t} \xrightarrow{f.s.} 0 \Rightarrow \lim_{t \rightarrow \infty} \frac{X(t)}{t} = \lim_{t \rightarrow \infty} \frac{W(t)}{t} - \mu = -\mu \Rightarrow X(t) \xrightarrow{f.s.} -\infty$$

$Y(T')\mathbf{1}(T \geq t) = \exp\{uX(t) - t(\frac{u^2}{2} - \mu u)\}\mathbf{1}(T \geq t) \xrightarrow[t \rightarrow +\infty]{f.s.} 0$, falls $\frac{u^2}{2} - \mu u > 0 \Rightarrow u \geq 2\mu$.
Andererseits, $Y(T')\mathbf{1}(T \leq t) \leq \exp\{uz\} \Rightarrow$ nach dem Satz von Lebesgue gilt:

$$\begin{aligned} & \mathbf{E}(Y(T')\mathbf{1}(T \geq t)) \xrightarrow[t \rightarrow +\infty]{} 0 \\ \Rightarrow & \lim_{t \rightarrow +\infty} \mathbf{E}(Y(T)\mathbf{1}(T < t)) = 1, \quad Y(T) = \exp\{uz - T(\frac{u^2}{2} - \mu u)\} \\ \Rightarrow & \lim_{t \rightarrow +\infty} \mathbf{E}(\exp\{-T(\frac{u^2}{2} - \mu u)\}\mathbf{1}(T < t)) = e^{-uz} \\ \xrightarrow{u=2\mu} & \lim_{t \rightarrow +\infty} \mathbf{P}(T < t) = \mathbf{P}(T < \infty) = e^{-2\mu z} \\ \Rightarrow & \mathbf{P}(\sup_{t \geq 0} X(t) > z) = \mathbf{P}(T_z^X < \infty) = e^{-2\mu z} \end{aligned}$$

□

Theorem 5.6.4

Sei $\mu \in \mathbb{R}$, $\delta > 0$, $T(t) = \inf\{s \geq 0 : W(s) + \mu s = \delta t\}$, $t \geq 0$. Dann ist $T = \{T(t), t \geq 0\}$ ein Lèvy-Prozess mit $\hat{m}_{T(t)}(z) = \mathbf{E}e^{-zT(t)} = \exp\{-t\delta(\sqrt{2z + \mu^2} - \mu)\}$, $t \geq 0$, $z \geq 0$.

Spezialfall: Für $\mu = 0$, $\delta = \frac{1}{\sqrt{2}}$ ist $T = \{T(t), t \geq 0\}$ ein $\frac{1}{2}$ -stabiler Subordinator, der auch manchmal Lèvy-Subordinator genannt wird. Hier gilt $\hat{m}_{X(t)}(z) = e^{-t\sqrt{z}}$. (Für α -stabile Subordinatoren gilt: $\hat{m}_{T(t)}(z) = e^{-tz^\alpha}$, $\alpha \in (0, 1)$)

Zur Erinnerung: Das Lèvy-Maß eines α -stabilen Subordinators ist

$$\nu(dx) = \frac{\alpha}{\Gamma(1-\alpha)} \frac{dx}{x^{1+\alpha}} \mathbf{1}(x > 0), \quad \alpha \in (0, 1).$$

Beweis des Satzes 5.6.4 im Spezialfall (allgemein geht analog)

Sei $T(t) = \inf\{s \geq 0 : W(s) = \frac{t}{\sqrt{2}}\}$, $t \geq 0$. Es ist zu zeigen, dass $T = \{T(t), t \geq 0\}$ ein Lèvy-Prozess ist.

$T(0) \stackrel{f.s.}{=} 0$. Aus dem Theorem 5.5.2 folgt, dass T unabhängige und stationäre Zuwächse hat. T ist stochastisch stetig, denn

$$\lim_{t \rightarrow 0} \mathbf{P}(T(t) > \varepsilon) = \lim_{t \rightarrow \infty} \mathbf{P}(\max_{s \in [0, \varepsilon]} W(s) < \frac{t}{\sqrt{2}}) = \lim_{t \rightarrow 0} (1 - \sqrt{\frac{2}{\pi \varepsilon}} \int_{\frac{t}{\sqrt{2}}}^{\infty} e^{-\frac{y^2}{2\varepsilon}} dy) = 1 - 1 = 0.$$

Somit haben wir bewiesen, dass T ein Lèvy-Prozess ist.

Es ist noch zu zeigen, dass $T(t)$ α -stabil für $\alpha = \frac{1}{2}$ ist, d.h. $\mathbf{E}e^{-zT(t)} = e^{-t\sqrt{z}}$, für alle z und $t \geq 0$. Ähnlich zum Beweis des Theorems 5.6.3 betrachten wir das Martingal $X = \{X(s), s \geq 0\}$, $X(s) = \exp\{zW(s) - s\frac{z^2}{2}\}$, $s \geq 0$.

Sei $Y_{n,t} = T(t) \wedge n$, für alle $n \in \mathbb{N}$, $t \geq 0$, eine Folge von Stoppzeiten bzgl. $\{\mathcal{F}_t, t \geq 0\}$. Aus Folgerung 5.4.1 ist $\{X(Y_{n,t}), n \in \mathbb{N}\}$ für alle $t, z > 0$, ebenfalls ein Martingal.

$$\begin{aligned} \mathbf{E}X(Y_{n,t}) &= \mathbf{E}X(Y_{0,t}) = \mathbf{E}X(0) = e^{z0} = 1 \\ &= \mathbf{E}(X(Y_{n,z})\mathbf{1}(T(t) < n)) + \mathbf{E}(X(Y_{n,t}))\mathbf{1}(T(t) \geq n) \\ &= \mathbf{E}(\exp\{z \underbrace{W(T(t))}_{=\frac{t}{\sqrt{2}}} - T(t)\frac{z^2}{2}\}\mathbf{1}(T(t) < n)) + \mathbf{E}(\exp\{zW(n) - n\frac{z^2}{2}\}\mathbf{1}(T(t) \geq n)) \end{aligned}$$

Es ist zu zeigen, dass $E(\exp\{zW(n) - n\frac{z^2}{2}\}1(T(t) \geq n)) \xrightarrow[n \rightarrow \infty]{} 0$. Daraus wird folgen, dass $1 = \lim_{n \rightarrow \infty} E(\exp\{z\frac{t}{\sqrt{2}} - T(t)\frac{z^2}{2}\}1(T(t) < n)) = E \exp\{z\frac{t}{\sqrt{2}} - T(t)\frac{z^2}{2}\}$, weil $T(t)$ eine endliche

Stoppzeit ist, d.h. $P(T(t) < \infty) = 1$ für alle $t \geq 0$.

Die obige Konvergenz gilt nach Lebesgueschem Satz über die majorisierte Konvergenz

$$\Rightarrow E \exp\{-T(t)\frac{z^2}{2}\} - e^{-t\frac{z^2}{2}} \Rightarrow E e^{-uT(t)} = e^{-t\sqrt{u}}, \quad u \geq 0.$$

Es ist noch zu zeigen, dass $E(\exp\{zW(n) - n\frac{z^2}{2}\}1(T(t) \geq n)) \xrightarrow[n \rightarrow \infty]{} 0$.

Zusätzlich gilt: $T(t) \geq n \Rightarrow W(n) \leq \frac{t}{\sqrt{2}}$.

$\exp\{zW(n) - n\frac{z^2}{2}\}1(T(t) \geq n) \leq \exp\{t\frac{t}{\sqrt{2}}\}$ für alle $n \in \mathbb{N}_0$.

\Rightarrow Die Konvergenz ergibt sich aus dem Satz von Lebesgue. □

Bemerkung 5.6.1

Falls $T(t) = \min\{s \geq 0 : W(s) + \mu s = \delta t\}$, $\mu \in \mathbb{R}$, $\delta > 0$, $t \geq 0$, dann kann die Laplace-Transformierte in $T(t)$ (vgl. Theorem 5.6.4) $E e^{-zT(t)} = \exp\{-t\delta(\sqrt{2z + \mu^2} - \mu)\}$ explizit invariantiert werden: die Dichte von $T(t)$ läßt sich schreiben als

$$f_{T(t)}(x) = \frac{\delta t}{\sqrt{2\pi}} e^{\delta t \mu} x^{-\frac{3}{2}} \exp\{-\frac{1}{2}(t^2 \delta^2 \frac{1}{x} + \mu^2 x)\} 1(x \geq 0).$$

Das ist die Dichte der *sogennanten inversen Gauß-Verteilung*.

Theorem 5.6.5

Sei $X = \{X(t), t \geq 0\}$ ein Lèvy-Prozess und sei $T = \{T(t), t \geq 0\}$ ein Subordinator, die beide auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) definiert sind. Seien X und T unabhängig. Dann ist $Y = \{Y(t), t \geq 0\}$ definiert durch $Y(t) = X(T(t))$, $t \geq 0$, ebenfalls ein Lèvy-Prozess.

Ohne Beweis

5.7 Ergänzende Aufgaben

Aufgabe 5.7.1

Seien $X, Y : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariablen über (Ω, \mathcal{F}, P) mit

$$E|X| < \infty, \quad E|Y| < \infty, \quad E|XY| < \infty,$$

und sei $\mathcal{G} \subset \mathcal{F}$ eine beliebige Teil- σ -Algebra von \mathcal{F} . Dann gilt

- (a) $E(X|\{\emptyset, \Omega\}) = EX$, $E(X|\mathcal{F}) = X$,
- (b) $E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$ für beliebige $a, b \in \mathbb{R}$,
- (c) $E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$, falls $X \leq Y$,
- (d) $E(XY|\mathcal{G}) = YE(X|\mathcal{G})$, falls Y eine $(\mathcal{G}, \mathcal{B}(\mathbb{R}))$ -messbare Zufallsvariable ist,

- (e) $E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_1)$, falls \mathcal{G}_1 und \mathcal{G}_2 Teil- σ -Algebren von \mathcal{F} sind mit $\mathcal{G}_1 \subset \mathcal{G}_2$,
- (f) $E(X|\mathcal{G}) = EX$, falls die σ -Algebra \mathcal{G} und $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$ unabhängig sind, d. h., falls $P(A \cap A') = P(A)P(A')$ für beliebige $A \in \mathcal{G}$ und $A' \in \sigma(X)$.
- (g) $E(f(X)|\mathcal{G}) \geq f(E(X|\mathcal{G}))$, falls $f: \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion ist, so dass $E|f(X)| < \infty$.

Aufgabe 5.7.2

Betrachte die zwei Zufallsvariablen X und Y über dem Wahrscheinlichkeitsraum $([-1, 1], \mathcal{B}([-1, 1]), \frac{1}{2}\nu)$ mit $E|X| < \infty$, wobei ν das Lebesguemaß auf $[-1, 1]$ bezeichnet. Bestimme für die folgenden Zufallsvariablen jeweils $\sigma(Y)$ und eine Version der bedingten Erwartung $E(X|Y)$.

- (a) $Y(\omega) = \omega^5$ (Hinweis: Zeige zunächst, dass $\sigma(Y) = \mathcal{B}([-1, 1])$)
- (b) $Y(\omega) = (-1)^k$ für $\omega \in \left[\frac{k-3}{2}, \frac{k-2}{2}\right)$, $k = 1, \dots, 4$ und $Y(1) = 1$
(Hinweis: Es gilt $E(X|B) = \frac{E(X1_B)}{P(B)}$ für $B \in \sigma(Y)$ mit $P(B) > 0$)
- (c) Berechne die Verteilung von $E(X|Y)$ in (a) und (b), falls $X \sim U[-1, 1]$.

Aufgabe 5.7.3

Seien X und Y Zufallsvariablen über einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Die bedingte Varianz $\text{var}(Y|X)$ ist definiert durch

$$\text{var}(Y|X) = E((Y - E(Y|X))^2|X).$$

Zeige, dass

$$\text{var} Y = E(\text{var}(Y|X)) + \text{var}(E(Y|X)).$$

Aufgabe 5.7.4

Für eine Stoppzeit τ definieren wir die gestoppte σ -Algebra \mathcal{F}_τ wie folgt:

$$\mathcal{F}_\tau = \{B \in \mathcal{F} : B \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ für beliebige } t \geq 0\}.$$

Seien nun S und T Stoppzeiten bzgl. der Filtration $\{\mathcal{F}_t, t \geq 0\}$. Zeige:

- (a) $A \cap \{S \leq T\} \in \mathcal{F}_T \forall A \in \mathcal{F}_S$
- (b) $\mathcal{F}_{\min\{S, T\}} = \mathcal{F}_S \cap \mathcal{F}_T$

Aufgabe 5.7.5 (a) Sei $\{X(t), t \geq 0\}$ ein Martingal. Zeige, dass $EX(t) = EX(0)$ für alle $t \geq 0$ gilt.

- (b) Sei $\{X(t), t \geq 0\}$ ein Sub- bzw. Supermartingal. Zeige, dass $EX(t) \geq EX(0)$ bzw. $EX(t) \leq EX(0)$ für alle $t \geq 0$ gilt.

Aufgabe 5.7.6

Der stochastische Prozess $X = \{X(t), t \geq 0\}$ sei adaptiert und càdlàg. Zeige, dass

$$P\left(\sup_{0 \leq v \leq t} X(v) > x\right) \leq \frac{EX(t)^2}{x^2 + EX(t)^2}$$

für beliebige $x > 0$ und $t \geq 0$ gilt, falls X ein Submartingal mit $EX(t) = 0$ und $EX(t)^2 < \infty$ ist.

Aufgabe 5.7.7 (a) Sei $g : [0, \infty) \rightarrow [0, \infty)$ eine monoton wachsende Funktion mit

$$\frac{g(x)}{x} \rightarrow \infty, \quad x \rightarrow \infty.$$

Zeige, dass die Folge X_1, X_2, \dots von Zufallsvariablen gleichgradig integrierbar ist, falls $\sup_{n \in \mathbb{N}} \mathbb{E}g(|X_n|) < \infty$.

(b) Sei $X = \{X(n), n \in \mathbb{N}\}$ ein Martingal. Zeige, dass die Folge von Zufallsvariablen $X(T \wedge 1), X(T \wedge 2), \dots$ für jede endliche Stoppzeit T gleichgradig integrierbar ist, falls $\mathbb{E}|X(T)| < \infty$ und $\mathbb{E}(|X(n)|1_{\{T > n\}}) \rightarrow 0$ für $n \rightarrow \infty$.

Aufgabe 5.7.8

Sei $S = \{S_n = a + \sum_{i=1}^n X_i, n \in \mathbb{N}\}$ eine symmetrische zufällige Irrfahrt mit $a > 0$ und $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$ für $i \in \mathbb{N}$. Die zufällige Irrfahrt wird zu demjenigen Zeitpunkt T gestoppt, bei dem sie zum ersten Mal einen der beiden Werte 0 und $K > a$ unter- bzw. überschreitet, d. h.

$$T = \min_{k \geq 0} \{S_k \leq 0 \text{ oder } S_k \geq K\}.$$

Zeige, dass $M_n = \sum_{i=0}^n S_i - \frac{1}{3}S_n^3$ ein Martingal ist und $\mathbb{E}(\sum_{i=0}^T S_i) = \frac{1}{3}(K^2 - a^2)a + a$ gilt.

Hinweis: Für die Berechnung von $\mathbb{E}(M_n | \mathcal{F}_m^M)$, $n > m$, kann $\mathbb{E}(\sum_{i=k}^l X_i)^3 = 0$, $1 \leq k \leq l$, $M_n = \sum_{r=0}^m S_r + \sum_{r=m+1}^n S_r - \frac{1}{3}S_n^3$ und $S_n = S_n - S_m + S_m$ verwendet werden.

Ein diskretes Martingal bezüglich einer Filtration $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ ist eine Folge von Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, so dass X_n bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ messbar ist und $\mathbb{E}(X_{n+1} | X_n) = X_n$ für alle $n \in \mathbb{N}$. Eine diskrete Stoppzeit bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ ist eine Zufallsvariable $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$, so dass $\{T \leq n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N} \cup \{\infty\}$, wobei $\mathcal{F}_\infty = \sigma\{\bigcup_{n=1}^\infty \mathcal{F}_n\}$.

Aufgabe 5.7.9

Seien $\{X_n\}_{n \in \mathbb{N}}$ ein diskretes Martingal und T eine diskrete Stoppzeit bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$. Zeige, dass auch $\{X_{\min\{T, n\}}\}_{n \in \mathbb{N}}$ ein Martingal bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ ist.

Aufgabe 5.7.10

Sei $\{S_n\}_{n \in \mathbb{N}}$ eine symmetrische zufällige Irrfahrt mit $S_n = \sum_{i=1}^n X_i$ für eine Folge von unabhängigen und identisch verteilten Zufallsvariablen X_1, X_2, \dots , so dass $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Sei $T = \inf\{n : |S_n| > \sqrt{n}\}$ und $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$, $n \in \mathbb{N}$.

- (a) Zeige, dass T eine Stoppzeit bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ ist.
- (b) Zeige, dass $\{G_n\}_{n \in \mathbb{N}}$ mit $G_n = S_{\min\{T, n\}}^2 - \min\{T, n\}$ ein Martingal bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ ist. (Hinweis: Verwende Aufgabe 5.7.9)
- (c) Zeige, dass $|G_n| \leq 4T$ für alle $n \in \mathbb{N}$ gilt.
(Hinweis: Es gilt $|G_n| \leq |S_{\min\{T, n\}}^2| + |\min\{T, n\}| \leq S_{\min\{T, n\}}^2 + T$)

Aufgabe 5.7.11

Sei X_1, X_2, \dots eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E}|X_1| < \infty$. Sei $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$, $n \in \mathbb{N}$, und sei T eine Stoppzeit bezüglich $\{\mathcal{F}_n\}_{n \in \mathbb{N}}$ mit $\mathbb{E}T < \infty$.

- (a) Sei T unabhängig von X_1, X_2, \dots . Leite eine Formel für die charakteristische Funktion von $S_T = \sum_{i=1}^T X_i$ her und weise damit die Waldsche Identität nach, d. h. $\mathbb{E}S_T = \mathbb{E}T \mathbb{E}X_1$.

- (b) Sei zusätzlich $EX_1 = 0$ und $T = \inf\{n : S_n < 0\}$. Verwende Theorem 2.1.3 aus der Vorlesung, um zu zeigen, dass $ET = \infty$. (Hinweis: Widerspruchsbeweis)

6 Stationäre Folgen von Zufallsvariablen

6.1 Reihen von unabhängigen Zufallsvariablen

Es ist bekannt, dass die Reihe

$$\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}} < \infty \iff \alpha > 1,$$

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^{\alpha}} < \infty \iff \alpha > 0,$$

weil die Drift der benachbarten Glieder die Ordnung $\frac{1}{n^{1+\alpha}}$ haben.

Wann (für welche $\alpha > 0$) konvergiert die Reihe $\sum_{n=1}^{\infty} \frac{\delta_n}{n^{\alpha}}$, wobei δ_n u.i.v. Zufallsvariablen sind mit $E\delta_n = 0$, z.B. $P(\delta_n = \pm 1) = \frac{1}{2}$?

Allgemeinere Frage: Unter welchen Bedingungen konvergiert (f.s.) die Reihe $\sum_{n=1}^{\infty} X_n$, wobei X_n unabhängig sind?

Man weiss, dass für eine Folge von Zufallsvariablen $\{Y_n\}$ aus $Y_n \xrightarrow[n \rightarrow \infty]{f.s.} Y$ gilt, dass $Y_n \xrightarrow[n \rightarrow \infty]{P} Y$. Das Gegenteil gilt im Allg. nicht.

Theorem 6.1.1

Seien $X_n, n \in \mathbb{N}$, unabhängige Zufallsvariablen. Falls $S_n = \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{P} S$, dann $S_n \xrightarrow[n \rightarrow \infty]{f.s.} S$.

Ohne Beweis

Folgerung 6.1.1

Falls die Folgen $X_n, n \in \mathbb{N}$, unabhängig sind, $\text{var } X_n < \infty, n \in \mathbb{N}, EX_n = 0, \sum_{n=1}^{\infty} \text{var } X_n < \infty$, dann konvergiert $\sum_{n=1}^{\infty} X_n$ f.s.

Beweis $S_n = \sum_{i=1}^n X_i, S = \sum_{i=1}^{\infty} X_i, m < n$,

$$E(S_n - S_m)^2 = \|S_n - S_m\|_{L^2}^2 = \sum_{i=m+1}^n \text{var } X_i \xrightarrow[n, m \rightarrow \infty]{} 0,$$

weil $\sum_{i=1}^{\infty} \text{var } X_i < \infty \Rightarrow \{S_n\}_{n \in \mathbb{N}}$ ist eine Cauchy-Folge in $L^2(\Omega, \mathcal{F}, P)$

$$\Rightarrow \exists S = \lim_{n \rightarrow \infty} S_n = \sum_{i=1}^{\infty} X_i \Rightarrow S_n \xrightarrow[n \rightarrow \infty]{P} S \xrightarrow{\text{Theorem 6.1.1}} S_n \xrightarrow[n \rightarrow \infty]{f.s.} S.$$

□

Folgerung 6.1.2

Falls $\sum_{n=1}^{\infty} a_n^2 < \infty$, wobei $\{a_n\}_{n \in \mathbb{N}}$ eine deterministische Folge ist, und $\{\delta_n\}$ eine Folge von u.i.v. Zufallsvariablen ist mit $E\delta_n = 0, \text{var } \delta_n = \sigma^2 < \infty, n \in \mathbb{N}$, dann konvergiert die Reihe $\sum_{n=1}^{\infty} a_n \delta_n$ f.s.

Aufgabe 6.1.1

Leiten Sie die Folgerung 6.1.2 aus dem Theorem 6.1.1 ab.

Bei uns: δ_n u.i.v., $E\delta_n = 0$, $\text{var } \delta_n = \sigma^2 > 0$ (z.B. $\delta_n \sim \text{Bernouli}(\frac{1}{2})$), $a_n = \frac{1}{n^\alpha}$, $n \in \mathbb{N}$.
 $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty$, falls $\sum_{n=1}^{\infty} \frac{1}{n^{2\alpha}} < \infty$, d.h. für $\alpha < \frac{1}{2}$.

Folgerung 6.1.3

Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $\sum_{n=1}^{\infty} EX_n$, $\sum_{n=1}^{\infty} \text{var } X_n < \infty$
 $\Rightarrow \sum_{n=1}^{\infty} X_n \stackrel{f.s.}{<} \infty$.

Beweis Sei $Y_n = X_n - EX_n$, daher ist $X_n = \underbrace{EX_n}_{=a_n} + Y_n$, $n \in \mathbb{N}$, und $EY_n = 0$, $\sum_{n=1}^{\infty} a_n < \infty$ nach Voraussetzung. $\sum_{n=1}^{\infty} Y_n \stackrel{f.s.}{<} \infty$ nach Folgerung 6.1.1, weil $\text{var } X_n = \text{var } Y_n$, $n \in \mathbb{N}$,
 $\sum_{n=1}^{\infty} \text{var } X_n < \infty \Rightarrow \sum_n X_n = \sum_n a_n + \sum_n Y_n \stackrel{f.s.}{<} \infty$. \square

6.2 Stationarität im engeren Sinne und Ergodentheorie

6.2.1 Grundbegriffe

Sei $\{X_n\}_{n \in \mathbb{N}}$ ein stationäre im engeren Sinne Folge von Zufallsvariablen, d.h. für alle $n, k \in \mathbb{N}$ die Verteilung von $(X_n, \dots, X_{n+k})^\top$ unabhängig von $n \in \mathbb{N}$ ist. Insbesondere heißt es, dass alle X_n identisch verteilt sind. In der Sprache des Theorems von Kolmogorov:

$$P((X_n, X_{n+1}, \dots) \in B) = P((X_1, X_2, \dots) \in B),$$

für alle $n \in \mathbb{N}$, für alle $B \in \mathcal{B}(\mathbb{R}^\infty)$, $\mathbb{R}^\infty = \mathbb{R} \times \mathbb{R} \times \dots \times \dots$

Beispiel 6.2.1 (von stationären Folgen von Zufallsvariablen): 1. Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von u.i.v. Zufallsvariablen, dann ist $\{X_n\}_{n \in \mathbb{N}}$ stationär.

2. Sei $Y_n = a_0 X_n + \dots + a_k X_{n+k}$, k - fixierte Zahl aus \mathbb{N} , $\{X_n\}_{n \in \mathbb{N}}$ aus 1), $a_0, \dots, a_k \in \mathbb{R}$ (fixiert), $n \in \mathbb{N}$. Y_n sind nicht mehr unabhängig, aber identisch verteilt. Die Folge $\{Y_n\}_{n \in \mathbb{N}}$ ist stationär.

3. Sei $Y_n = \sum_{j=0}^{\infty} a_j X_{n+j}$ für beliebiges $n \in \mathbb{N}$. Die Folge $\{a_j\}_{j \in \mathbb{N}}$ ist eine Zahlenfolge aus \mathbb{R} mit der Eigenschaft, dass $\sum_{j=1}^{\infty} |a_j| < \infty$ und $EX_n = 0$, $\sum_{n=1}^{\infty} \text{var } X_n < \infty$, $\sum_{j=1}^{\infty} a_j^2 < \infty$ (vgl. Folgerung 6.1.2).

Es ist offensichtlich, dass $\{Y_n\}_{n \in \mathbb{N}}$ eine stationäre Folge ist. (Diese Konstruktion ist wichtig für die autoregressiven Zeitreihen (AR-Prozesse), z.B. in der Ökonometrie).

4. Sei $Y_n = g(X_n, X_{n+1}, \dots)$, $n \in \mathbb{N}$, $g: \mathbb{R}^\infty \rightarrow \mathbb{R}$ meßbar, $\{X_n\}_{n \in \mathbb{N}}$ aus 1). Dann ist $\{Y_n\}_{n \in \mathbb{N}}$ stationär.

Bemerkung 6.2.1 1. Eine beliebige stationäre Folge von Zufallsvariablen $X = \{X_n\}_{n \in \mathbb{N}}$ kann man erweitern zu einer stationären Folge $\bar{X} = \{X_n\}_{n \in \mathbb{Z}}$. Tatsächlich kann die endlich dimensionale Verteilung von \bar{X} nach dem Satz von Kolmogorov durch die von X definiert werden:

$$(X_n, \dots, X_{n+k}) \stackrel{d}{=} (X_1, \dots, X_{k+1}), \quad n \in \mathbb{Z}, k \in \mathbb{N}.$$

Deshalb (nach dem Satz von Kolmogorov) existiert ein Wahrscheinlichkeitsraum und eine Folge $\{Y_n\}_{n \in \mathbb{Z}}$ mit der obigen Verteilung. Wir setzen $\bar{X} = \{Y_n\}_{n \in \mathbb{Z}}$ und daraus folgt, dass $\{Y_n\}_{n \in \mathbb{N}} \stackrel{d}{=} \{X_n\}_{n \in \mathbb{N}}$.

2. Wir definieren eine Koordinatenverschiebung. Sei $x \in \mathbb{R}_{-\infty}^{\infty}$, $x = (x_k, k \in \mathbb{N})$, $x = (x_k, k \in \mathbb{Z})$. Definiere die Abbildung $\theta : \mathbb{R}_{-\infty}^{\infty} \rightarrow \mathbb{R}_{-\infty}^{\infty}$, $(\theta x)_k = x_{k+1}$ (Verschiebung der Koordinaten um 1), $k \in \mathbb{N}$, $k \in \mathbb{Z}$. Falls θ auf $\mathbb{R}_{-\infty}^{\infty}$ betrachtet wird, so ist sie bijektiv und die Umkehrabbildung wäre $(\theta^{-1}x)_k = x_{k-1}$, $k \in \mathbb{Z}$.

Sei nun $X = \{X_n, n \in \mathbb{Z}\}$ eine stationäre Folge von Zufallsvariablen. Sei $\bar{X} = \theta X$. Es ist offensichtlich, dass \bar{X} wieder stationär ist und $\bar{X} \stackrel{d}{=} X$. Daraus folgt, dass

$$P(\theta X \in B) = P(X \in B), \quad B \in \mathcal{B}(\mathbb{R}_{-\infty}^{\infty}).$$

θ wird eine *maßerhaltende Abbildung* genannt. Es gibt aber auch andere Abbildungen, die maßerhaltend wirken.

Definition 6.2.1

Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum. Eine Abbildung $T : \Omega \rightarrow \Omega$ heißt *maßerhaltend*, falls

1. T meßbar ist, d.h. $T^{-1}A \in \mathcal{F}$ für alle $A \in \mathcal{F}$,
2. $P(T^{-1}A) = P(A)$, $A \in \mathcal{F}$.

Lemma 6.2.1

Sei T eine maßerhaltende Abbildung und X_0 – eine Zufallsvariable. Wir definieren eine Folge von Zufallsvariablen X_n . Sei Abbildung $UY(\omega) = Y(T(\omega))$, $\omega \in \Omega$, für eine beliebige Zufallsvariable Z auf (Ω, \mathcal{F}, P) . Definiere $X_n(\omega) = U^n X_0(\omega) = X_0(T^n(\omega))$, $\omega \in \Omega$, $n \in \mathbb{N}$. Dann ist die Folge von Zufallsvariablen $X = \{X_0, X_1, X_2, \dots\}$ stationär.

Beweis Sei $B \in \mathcal{B}(\mathbb{R}^{\infty})$, $A = \{\omega \in \Omega : X(\omega) \in B\}$, $A_1 = \{\omega \in \Omega : \theta X(\omega) \in B\}$.

$$\begin{aligned} X(\omega) &= (X_0(\omega), X_0(T(\omega)), X_0(T^2(\omega)), \dots) \\ \theta X(\omega) &= (X_0(T(\omega)), X_0(T^2(\omega)), \dots) \end{aligned}$$

Deshalb $\omega \in A_1 \Leftrightarrow T(\omega) \in A$. Weil $P(T^{-1}A) = P(A)$, gilt $P(A_1) = P(A)$. Für $A_n = \{\omega \in \Omega : \theta^n X(\omega) \in B\}$ gilt dasselbe, $P(A_n) = P(A)$, $n \in \mathbb{N}$ (Induktion). Und daraus folgt, dass die Folge X stationär ist. \square

Die Folge X in Lemma 6.2.1 wird die *Folge, die von T erzeugt wird*, genannt.

Definition 6.2.2

Eine Abbildung $T : \Omega \rightarrow \Omega$ heißt *maßerhaltend in beide Richtungen*, falls

1. T bijektiv ist und $T(\Omega) = \Omega$,
2. T und T^{-1} meßbar sind,
3. $P(T^{-1}A) = P(A)$, $A \in \mathcal{F}$, und, folglich, $P(TA) = P(A)$.

Somit können wir genau wie in Lemma 6.2.1 stationäre Folgen von Zufallsvariablen mit Zeitparameter $n \in \mathbb{Z}$ konstruieren:

$$X(\omega) = \{X_0(T^n(\omega))\}_{n \in \mathbb{N}}, \quad \omega \in \Omega,$$

wobei T eine maßerhaltende Abbildung (in beide Richtungen) ist, $X_0(T^0(\omega)) = X_0(\omega)$, ($T^0 = Id$).

Lemma 6.2.2

Für eine beliebige stationäre Folge von Zufallsvariablen $X = (X_0, X_1, \dots)$ existiert eine maßerhaltende Abbildung T und eine Zufallsvariable Y_0 , so dass $Y(\omega) = \{Y_0(T^n(\omega))\}_{n \in \mathbb{N}}$ dieselbe Verteilung wie X besitzt: $X \stackrel{d}{=} Y$. Dieselbe Aussage gilt für Folgen mit dem Zeitparameter $n \in \mathbb{Z}$.

Beweis Betrachten wir den kanonischen Wahrscheinlichkeitsraum $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), P_X)$, $Y(\omega) = \omega$, $\omega \in \mathbb{R}$, $T = \theta$. Damit ist Y konstruiert, weil $P_X(A) = P_Y(A) = P_X(Y \in A)$, $A \in \mathcal{B}(\mathbb{R}^\infty)$. \square

Beispiel 6.2.2 (Maßerhaltende Abbildungen): 1. Sei $\Omega = \{\omega_1, \dots, \omega_k\}$, $k \geq 2$, $\mathcal{F} = 2^\Omega$, $P(\omega_i) = \frac{1}{k}$, $i = 1, \dots, k$, ein Laplacescher Wahrscheinlichkeitsraum. $T\omega_i = \omega_{i+1}$ für alle $i = 1, \dots, k-1$, $T\omega_k = \omega_1$.

2. Sei $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}([0, 1))$, $P = \nu_1$ – Lebesgue-Maß auf $[0, 1)$. $T\omega = (\omega + s) \bmod 1$, $s \geq 0$. T ist maßerhaltend in beide Richtungen.

Folgen von Zufallsvariablen, die in diesen Beispielen durch die Abbildung T erzeugt werden können, sind meistens deterministisch bzw. zyklisch. Im Beispiel 1) können wir eine Zufallsvariable $X_0 : \Omega \rightarrow \mathbb{R}$ betrachten, so dass $X(\omega_i) = x_i$ alle von einander verschieden sind. Deswegen $X_n(\omega) = X_0(T^n(\omega))$ wird eindeutig den Wert von $X_{n+1}(\omega) = X_0(T^{n+1}(\omega))$ definieren, für alle $n \in \mathbb{N}$.

Bemerkung 6.2.2

Maßerhaltende Abbildungen spielen eine große Rolle auch in der Physik. Dort wird T als die Veränderung des Zustandes von einem physikalischen System interpretiert und das Maß kann z.B. das Volumen sein. (Bsp.: T – Veränderung der Temperatur, Maß P – Volumen vom Gas.) Deswegen ist die zu entwickelte Ergodentheorie auf manche physikalische Vorgänge übertragen.

Theorem 6.2.1 (Poincarè):

Falls T eine maßerhaltende Abbildung auf (Ω, \mathcal{F}, P) ist, $A \in \mathcal{F}$, dann für fast alle $\omega \in A$ die Relation $\{T^n(\omega) \in A\}$ gilt für unendlich viele $n \in \mathbb{N}$.

Das heißt, dass die Trajektorie $\{T^n(\omega), n \in \mathbb{N}\}$ kehrt unendlich oft zu A zurück, falls $\omega \in \Omega$, $P(A) > 0$.

Beweis Es ist zu zeigen, dass $A \in \mathcal{F}$, $T : \Omega \rightarrow \Omega$ maßerhaltend. Zeige, dass für fast alle $\omega \in \Omega$, $T^n(\omega) \in A$ für unendlich viele $n \in \mathbb{N}$. Sei $N = \{\omega \in A : T^n(\omega) \notin A \forall n \geq 1\}$. Es ist klar, dass $N \in \mathcal{F}$, weil $\{\omega \in \Omega : T^n(\omega) \notin A\} \in \mathcal{F}$ für alle $n \geq 1$. $N \cap T^{(-n)}N = \emptyset$ für alle $n \geq 1$. Tatsächlich, falls $\omega \in N \cap T^{(-n)}N$, dann $\omega \in A$, $T^{(n)}(\omega) \notin A$ für alle $n \geq 1$, $\omega_1 = T^n(\omega)$, $\omega_1 \in N$. Daraus folgt, dass $\omega_1 \in A$ und $T^n(\omega) \in A$. Das ist Widerspruch. $T^{(-n)}N = \{\omega \in \Omega : T^n(\omega) \in N\}$. Für beliebige $m \in \mathbb{N}$ gilt

$$T^{(-m)}N \cap T^{(-(n+m))}N = T^{(-m)}(N \cap T^{(-n)}N) = T^{(-m)}(\emptyset) = \emptyset.$$

Daraus folgt, dass die Mengen $T^{(-n)}N$, $n \in \mathbb{N}$, paarweise disjunkt sind, zu \mathcal{F} gehören und $P(T^{(-n)}N) = P(A) = a \geq 0$ haben.

$$1 \geq P(\cup_n T^{(-n)}N) = \sum P(T^{(-n)}N) = \sum_{n=0}^{\infty} a \Rightarrow a = 0 \Rightarrow P(N) = 0.$$

Daraus folgt, dass für fast alle $\omega \in A$ ($\omega \in A \setminus N$) ein $n_1 = n_1(\omega)$ existiert, so dass $T^{(m)}(\omega) \in A$. Sei nun T^k an Stelle von T , $k \in \mathbb{N}$. Es gilt $P(N_k) = 0$ und für alle $\omega \in A \setminus N_k$ existiert

$n_k = n_k(\omega)$, so dass $(T^k)^{n_k}(\omega) \in A$. Da $kn_k \geq k$ folgt für fast alle $\omega \in A$, dass $T^{(n)}(\omega) \in A$ für unendlich viele n . \square

Folgerung 6.2.1

Sei $X \geq 0$ eine Zufallsvariable, $A = \{\omega \in \Omega : X(\omega) > 0\}$. Dann für fast alle $\omega \in \Omega$ gilt $\sum_{n=0}^{\infty} X(T^{(n)}(\omega)) = +\infty$, wobei T eine maßerhaltende Abbildung ist.

Aufgabe 6.2.1

Beweisen Sie es.

Bemerkung 6.2.3

Der Beweis des Theoremes 6.2.1 gilt für Mengen $A \in \mathcal{F} : P(A) \geq 0$. Falls jedoch $P(A) = 0$, kann es sein, dass $A \cap N = \emptyset$ und somit ist die Aussage des Theoremes trivial.

Als Beispiel betrachten wir $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}([0, 1))$, $P = \nu_1$ - Lebesgue-Maß, $T(\omega) = \omega + s \bmod 1$, $s \in \mathbb{Q}$. Als Menge A betrachten wir $A = \omega_0$, $\omega_0 \in \Omega$. Dann gilt $T^n(\omega_0) \neq \omega_0$ für alle n , denn sonst existiert $k, m \in \mathbb{N}$, so dass $\omega_0 + ks - m = \omega_0$ und daraus folgt $s = \frac{m}{k} \in \mathbb{Q}$. Somit bekommen wir einen Widerspruch.

6.2.2 Mischungseigenschaften und Ergodizität

Hier studieren wir die Abhängigkeitsstruktur in einer stationären Folge von Zufallsvariablen, die durch eine maßerhaltende Abbildung T erzeugt wird.

Sei $X = \{X_n\}_{n \in \mathbb{N}}$ eine stationäre Folge (im engen Sinne) von Zufallsvariablen. Dann existiert eine maßerhaltende Abbildung $T : \Omega \rightarrow \Omega$, so dass $X_n(\omega) \stackrel{d}{=} X_0(T^{(n)}(\omega))$ und $X_n \stackrel{d}{=} X_0$, und somit gibt X_s die Randverteilung der Folge X an. Dafür ist die Abbildung T für die Abhängigkeiten innerhalb von X zuständig (sie gibt Eigenschaften von mehrdimensionalen Verteilungen an). Deshalb werden wir jetzt die Abhängigkeitseigenschaften untersuchen, die von T erzeugt werden.

Definition 6.2.3 1. Ereignis $A \in \mathcal{F}$ heißt invariant bzgl. (einer maßerhaltenden Abbildung) $T : \Omega \rightarrow \Omega$, falls $T^{-1}A = A$.

2. Ereignis $A \in \mathcal{F}$ heißt fast invariant bzgl. T , falls $P(T^{-1}A \Delta A) = 0$. Δ bedeutet die symmetrische Differenz.

Aufgabe 6.2.2

Zeigen Sie, dass die Menge aller (fast) invarianten Ereignisse bzgl. T eine σ -Algebra $J(J^*)$ ist.

Lemma 6.2.3

Sei $A \in J^*$. Dann existiert $B \in J^*$, so dass $P(A \Delta B) = 0$

Beweis Sei $B = \limsup_{n \rightarrow \infty} T^n A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k} A$. Es ist zu zeigen, dass $B \in J$, $P(A \Delta B) = 0$. Es ist klar, dass $T^{-1}(B) = \limsup_{n \rightarrow \infty} T^{-(n+1)} A = B$ und daraus folgt, dass $B \in J$.

Es ist leicht zu sehen, dass $A \Delta B \subset \bigcup_{k=0}^{\infty} (T^{-k} A \Delta T^{-(k+1)} A)$. Da $P(T^{-k} A \Delta T^{-(k+1)} A) = 0$ für alle $k \geq 1$ wegen $A \in J^*$, folgt, dass $P(A \Delta B) = 0$. \square

Definition 6.2.4 1. Die maßerhaltende Abbildung $T : \Omega \rightarrow \Omega$ heißt ergodisch, falls für jedes $A \in J$

$$P(A) = \begin{cases} 0 \\ 1 \end{cases} .$$

2. Die stationäre Folge von Zufallsvariablen $X = \{X_n\}_{n \in \mathbb{N}}$ heißt ergodisch, falls die maßerhaltende Abbildung $T : \Omega \rightarrow \Omega$, die X erzeugt, ergodisch ist.

Lemma 6.2.4

Die maßerhaltende Abbildung T ist ergodisch genau dann, wenn die Wahrscheinlichkeit beliebiger fast invarianter Mengen

$$P(A) = \begin{cases} 0 \\ 1 \end{cases} \text{ für alle } A \in J^*.$$

Beweis „ \Leftarrow “

Klar, weil beliebige invariante Menge auch fast invariant ist, d.h. $J \subset J^*$

„ \Rightarrow “

T – ergodisch. Sei $A \in J^*$. Es folgt, dass es $B \in J$ existiert, so dass $P(A \Delta B) = 0$ nach Lemma 6.2.3. T – ergodisch und daraus folgt

$$P(B) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{und} \quad P(A) = \begin{cases} 0 \\ 1 \end{cases}.$$

□

Definition 6.2.5

Eine Zufallsvariable $Y : \Omega \rightarrow \mathbb{R}$ heißt (fast) invariant bzgl. $T : \Omega \rightarrow \Omega$ (maßerhaltende Abbildung), falls $Y(\omega) = Y(T(\omega))$ für (fast) alle $\omega \in \Omega$.

Theorem 6.2.2

Sei $T : \Omega \rightarrow \mathbb{R}$ eine maßerhaltende Abbildung. Folgende Aussagen sind äquivalent:

1. T – ergodisch
2. Falls Y invariant bzgl. T ist, dann $Y = \text{const}$ f.s.
3. Falls Y fast invariant bzgl. T ist, dann $Y = \text{const}$ f.s.

Beweis 1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 1)

1) \Rightarrow 2)

T – ergodisch, Y – fast invariant. Es ist zu zeigen, dass $Y(\omega) = \text{const}$ für fast alle $\omega \in \Omega$.

$Y(T(\omega)) = Y(\omega)$ fast sicher. Sei $A_v = \{\omega \in \Omega : Y(\omega) \leq v\}$, $v \in \mathbb{R}$. Daraus folgt, dass $A_v \in J^*$ für alle $v \in \mathbb{R}$ und nach dem Lemma 6.2.4

$$P(A_v) = \begin{cases} 0 \\ 1 \end{cases} \text{ für alle } v.$$

Sei $c = \sup \{v : P(A_v) = 0\}$. Zeige, dass $P(Y = c) = 1$.

$A_v \uparrow \Omega$, $v \rightarrow \infty$, $A_v \downarrow \emptyset$, $v \rightarrow -\infty \Rightarrow |c| < \infty$.

$$P(Y < c) = P\left(\bigcup_{n=1}^{\infty} \left\{Y \leq c - \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{\infty} P\left(A_{c-\frac{1}{n}}\right) = 0.$$

Genauso $P(Y > c) = 0$ und $P(Y = c) = 1$.

2) \Rightarrow 3)

Offensichtlich.

3) \Rightarrow 1) Es ist zu zeigen, dass T ergodisch ist, d.h. für alle $A \in \mathcal{J}$ $P(A) = \begin{cases} 0 \\ 1 \end{cases}$.

Sei $Y = 1_A$ - invariant bzgl. T , folgt daraus, dass $1_A = const = \begin{cases} 0 \\ 1 \end{cases}$ und $P(A) = \begin{cases} 0 \\ 1 \end{cases}$. \square

Bemerkung 6.2.4 1. Die Aussage des Theorems 6.2.2 bleibt gültig, wenn man 3) für f.s. beschränkte Zufallsvariablen Y fordert.

2. Falls Y invariant bzgl. T ist, dann ist $Y_n = \min\{Y, n\}$, $n \in \mathbb{N}$, auch invariant bzgl. T .

Beispiel 6.2.3 1. Sei $\Omega = \{\omega_1, \dots, \omega_d\}$, $\mathcal{F} = 2^\Omega$, $P(\{\omega_i\}) = \frac{1}{d}$, $i = 1, \dots, d$. Sei $T(\omega_i) = \omega_{i+1} \pmod d$, d.h. $\omega_d \xrightarrow{T} \omega_1$. T ist offensichtlich ergodisch und jede invariante Zufallsvariable ist konstant.

2. Sei $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, $P = \nu_1$, $T(\omega) = (\omega + s) \pmod 1$. Zeige, dass T ergodisch $\iff s \notin \mathbb{Q}$.

Beweis „ \Leftarrow “

Sei $s \notin \mathbb{Q}$, Y - eine beliebige invariante Zufallsvariable. Sei $EY^2 < \infty$. Zerlegen wir die Zufallsvariable Y in eine Fourier-Reihe. Die Fourier-Reihe von Y ist $Y(\omega) = \sum_{n=0}^\infty a_n e^{2\pi i n \omega}$. Wir wollen zeigen, dass $a_n = 0$, $n > 0$, und daraus folgt dann, dass $Y(\omega) = a_0$ f.s.. Dann ist T ergodisch nach dem Theorem 6.2.2.

$$a_n = \langle Y(\omega), e^{2\pi i n \omega} \rangle_{L^2} = E(Y(\omega)e^{-2\pi i n \omega}) = E(Y(T(\omega))e^{-2\pi i n \omega})e^{-2\pi i n s} = e^{-2\pi i n s} a_n,$$

$s \notin \mathbb{Q} \Rightarrow a_n = 0$.

„ \Rightarrow “

Falls $s = \frac{m}{n} \in \mathbb{Q}$, dann ist T nicht ergodisch, d.h. existiert $A \in \mathcal{J}$, so dass $0 < P(A) < 1$.

Sei $A = \cup_{k=0}^{n-1} \left\{ \omega \in \Omega : \frac{2k}{2n} \leq \omega < \frac{2k+1}{2n} \right\}$ und $P(A) = \frac{1}{2}$. A ist invariant, weil $T(A) = \left(A + \frac{2m}{2n} \right) \pmod 1 = A$. \square

Definition 6.2.6 1. Die maerhaltende Abbildung $T : \Omega \rightarrow \Omega$ heit *mischend*, falls fr alle $A_1, A_2 \in \mathcal{F}$ gilt: $P(A_1 \cap T^{-n}A_2) \xrightarrow{n \rightarrow \infty} P(A_1)P(A_2)$, d.h. bei wiederholten Anwendungen von T auf A_2 werden A_1 und A_2 asymptotisch unabhngig.

2. Sei $X = \{X_n\}_{n \in \mathbb{N}_0}$ eine stationre Folge von Zufallsvariablen die von Zufallsvariable X_0 und einer maerhaltenden Abbildung T erzeugt wird. X heit *schwach abhngig*, falls Zufallsvariable X_k und X_{k+n} fr $n \rightarrow \infty$ asymptotisch unabhngig werden, d.h. fr alle $B_1, B_2 \in \mathcal{B}_{\mathbb{R}}$

$$P(X_k \in B_1, X_{k+n} \in B_2) \xrightarrow{n \rightarrow \infty} P(X_0 \in B_1)P(X_0 \in B_2).$$

Theorem 6.2.3

Eine stationre Folge von Zufallsvariablen $X = \{X_n\}_{n \in \mathbb{N}_0}$, erzeugt durch die maerhaltende Abbildung T , ist schwach abhngig im Mittel genau dann, wenn T mischend im Mittel ist.

Aufgabe 6.2.3

Beweisen Sie das Theorem.

Theorem 6.2.4

Sei T eine maerhaltende Abbildung. Sie ist ergodisch genau dann, wenn sie mischend im Mittel ist.

Beweis „ \Leftarrow “

Es ist zu zeigen, dass wenn T mischend im Mittel ist, folgt daraus, dass T ergodisch ist,

$$\text{d.h. für alle } A \in \mathcal{J} \text{ gilt } P(A) = \begin{cases} 0 \\ 1 \end{cases} \cdot A_1 \in \mathcal{F}, A_2 = A = J, \frac{1}{n} \sum_{k=1}^n P(A_1 \cap \underbrace{T^{-k}(A_2)}_{=A_2}) =$$

$$P(A_1 \cap A_2) \xrightarrow{n \rightarrow \infty} P(A_1)P(A_2). \quad P(A_1 \cap A_2) = P(A_1)P(A_2) \text{ für } A_1 = A, P(A) = P^2(A) \text{ und}$$

$$P(A) = \begin{cases} 0 \\ 1 \end{cases}$$

„ \Rightarrow “
Später. □

Jetzt geben wir die Motivation die Motivation an für den Begriff „mischende Abbildung“.

Theorem 6.2.5

Sei $A \in \mathcal{F}$, $P(A) > 0$. Die maßerhaltende Abbildung $T : \Omega \rightarrow \Omega$ ist ergodisch (d.h. mischend im Mittel) genau dann, wenn

$$P\left(\bigcup_{n=0}^{\infty} T^{-n}A\right) = 1.$$

D.h. die Urbilder $T^{-n}A$, $n \in \mathbb{N}_0$, decken fast das ganze Ω ab.

Beweis „ \Leftarrow “

Sei $B = \bigcup_{n=0}^{\infty} T^{-n}A$. Offensichtlich, $T^{-1}B = \bigcup_{n=1}^{\infty} T^{-n}A \subset B$. Da T maßerhaltend ist, d.h. $P(T^{-1}B) = P(B)$, folgt, dass $P(T^{-1}B \Delta B) = 0$, $B \in \mathcal{J}^*$ (B – fast invariant bzgl. T) und

$$P(B) = \begin{cases} 0 \\ 1 \end{cases} \cdot P(B) \geq P(A) > 0 \Rightarrow P(B) = 1.$$

„ \Rightarrow “

Sei T nicht ergodisch. Es ist zu zeigen, dass $P(B) < 1$.

Wenn T nicht ergodisch ist, dann existiert $A \in \mathcal{J}$, so dass $0 < P(A) < 1$. $B = \bigcup_{n=0}^{\infty} T^{-n}A = A$ und $P(B) < 1$. □

Bemerkung 6.2.5

Bisher wurde niemals explizit die Tatsache genutzt, dass die Zufallsvariablen X reellwertig sind. Deshalb kann man die obigen Betrachtungen ohne Veränderung auf Folgen von Folgelementen mit Werten in einem bel. meßbaren Raum \mathcal{M} übertragen.

6.2.3 Ergodensatz

Sei $X = \{X_n\}_{n=0}^{\infty}$ eine Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) . Falls X_n u.i.v. sind, dann

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow[n \rightarrow \infty]{f.s.} EX_0, \quad E|X_0| < \infty.$$

Wir wollen eine ähnliche Aussage über stationäre Folgen beweisen.

Theorem 6.2.6 (Ergodensatz, Birkhoff-Khinchin):

Sei $X = \{X_n\}_{n \in \mathbb{N}_0}$ eine stationäre Folge von Zufallsvariablen, erzeugt von der Zufallsvariable

X_0 und einer maßerhaltenden Abbildung $T : \Omega \rightarrow \Omega$. Sei J die σ -Algebra der invarianten Mengen von T , d.h. $E|X_0| < \infty$. Dann

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow[n \rightarrow \infty]{f.s.} E(X_0 | J).$$

Falls X schwach abhängig im Mittel ist (d.h. T – ergodisch), dann $E(X_0 | J) = E(X_0)$.

Lemma 6.2.5

Seien $\{X_n\}$, T wie oben. Sei $S_n(\omega) = \sum_{k=0}^{n-1} X_0(T^k(\omega))$, $M_n(\omega) = \max\{0, S_1(\omega), \dots, S_n(\omega)\}$. Unter der Bedingung des Theorems 6.2.6 gilt

$$E(X_0 1(M_n > 0)) \geq 0, \quad n \in \mathbb{N}.$$

Beweis Für alle $k \leq n$ gilt $\underbrace{S_k(\omega)}_{S_k(T(\omega))} \leq \underbrace{M_n(\omega)}_{M_n(T(\omega))}$. Wir können noch X_0 hinzuaddieren und

bekommen

$$X_0(\omega) + M_n(T(\omega)) \geq X_0(\omega) + S_k(T(\omega)) = S_{k+1}(\omega).$$

Für $k = 0$ gilt $X_0(\omega) \geq S_1(\omega) - M_n(T(\omega))$. Dasselbe gilt für $k = 0, \dots, n-1$. Daraus folgt, dass $X_0(\omega) \geq \underbrace{\max\{S_1(\omega), \dots, S_n(\omega)\}}_{=M_n(\omega)} - M_n(T(\omega))$. Da $M_n(\omega) > 0$, dann $M_n = \max\{S_1, \dots, S_n\}$.

Es folgt, dass

$$E(X_0 1(M_n > 0)) \geq E((M_n - M_n(T)) 1(M_n > 0)) \geq E(M_n - M_n(T)) = 0.$$

□

Beweis des Ergordensatzes Die Aussage $E(X_0 | J) = E(X_0)$ ist trivial, weil für ergodische T gilt $J = \{\emptyset, \Omega\}$. O.B.d.A. sei $E(X_0 | J) = 0$, sonst betrachte $X_0 = E(X | J)$.

Es ist zu zeigen: $\lim_{n \rightarrow \infty} \frac{S_n}{n} \stackrel{f.s.}{=} 0$, $S_n = \sum_{k=0}^{n-1} X_k$. Es genügt zu zeigen, dass

$$0 \leq \liminf_{n \rightarrow \infty} \frac{S_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq 0.$$

Zunächst zeigen wir, dass $\bar{S} = \limsup_{n \rightarrow \infty} \frac{S_n}{n} \leq 0$. Es reicht zu zeigen, dass $P(\underbrace{\bar{S} > \varepsilon}_{A_\varepsilon}) = 0$ für

alle $\varepsilon > 0$. Seien $X_0^* = (X_0 - \varepsilon) 1_{A_\varepsilon}$, $S_k^* = \sum_{j=0}^{k-1} X_0^*(T^j(\omega))$, $M_k^* = \max\{0, S_1^*, \dots, S_k^*\}$. Aus Lemma 6.2.5 folgt $E(X_0^* 1(M_n^* > 0)) \geq 0$ für alle $n \geq 1$. Aber,

$$\{M_n^* > 0\} = \left\{ \max_{1 \leq k \leq n} S_k^* > 0 \right\} \uparrow_{n \rightarrow \infty} \left\{ \sup_{k \geq 1} S_k^* > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\} \cap A_\varepsilon = A_\varepsilon,$$

weil $\left\{ \sup_{k \geq 1} \frac{S_k}{k} > \varepsilon \right\} \supset \left\{ \bar{S} > \varepsilon \right\} = A_\varepsilon$. Nach dem Lebesgue-Satz: $0 \leq E(X_0^* 1(M_n^* > 0)) \xrightarrow[n \rightarrow \infty]{} E(X_0^* 1_{A_\varepsilon})$, weil $E|X_0^*| \leq E|X_0| + \varepsilon$. Deshalb $0 \leq E(X_0^* 1_{A_\varepsilon}) = E((X_0 - \varepsilon) 1_{A_\varepsilon}) = E(X_0 1_{A_\varepsilon}) - \varepsilon P(A_\varepsilon) = E(E(X_0 1_{A_\varepsilon} | J)) - \varepsilon P(A_\varepsilon) = E(1_{A_\varepsilon} \underbrace{E(X_0 | J)}_{=0}) - \varepsilon P(A_\varepsilon) = -\varepsilon P(A_\varepsilon)$ und daraus folgt

$P(A_\varepsilon) \leq 0$ und $P(A_\varepsilon) = 0$ für alle $\varepsilon > 0$.

Um $0 \leq \liminf_{n \rightarrow \infty} \frac{S_n}{n} = \underline{S}$ zu zeigen genügt es $-X_0$ statt X_0 zu betrachten, denn $\limsup_{n \rightarrow \infty} \left(-\frac{S_n}{n}\right) = \liminf_{n \rightarrow \infty} \left(\frac{S_n}{n}\right)$. Da $P(-\underline{S} \leq 0) = 1$ gilt $P(\underline{S} \geq 0) = 1$. □

Bemerkung 6.2.6

Die Besonderheit des Ergodensatzes 6.2.6 liegt, im Vergleich zu dem üblichen Gesetz der großen Zahlen, in der Tatsache, dass der Grenzwert $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{f.s.} \mathbb{E}(X_0 | J)$ zufällig ist.

Beispiel 6.2.4

Betrachten wir den Wahrscheinlichkeitsraum aus dem Beispiel 6.2.3 a). $\Omega = \{\omega_1, \dots, \omega_d\}$, $d = 2l \in \mathbb{N}$. $T : \Omega \rightarrow \Omega$ sei definiert durch

$$\begin{cases} T(\omega_i) &= \omega_{i+2} &, \quad i = 1, \dots, d-2, \\ T(\omega_{d-1}) &= \omega_1 &, \\ T(\omega_d) &= \omega_2 & . \end{cases}$$

Seien $A_1 = \{\omega_1, \omega_3, \dots, \omega_{2l-1}\}$, $A_2 = \{\omega_2, \omega_4, \dots, \omega_{2l}\}$. Da $(\Omega, \mathcal{F}, \mathbb{P})$ ein Laplacescher Wahrscheinlichkeitsraum ($\mathbb{P}(\{\omega_i\}) = \frac{1}{d}$, für alle i) folgt, dass $\mathbb{P}(A_i) = \frac{1}{2}$, $i = 1, 2$. Andererseits, $A_1, A_2 \in J$ bzgl. T und deswegen ist T nicht ergodisch. Für eine beliebige Zufallsvariable $X_0 : \Omega \rightarrow \mathbb{R}$ gilt

$$\frac{1}{n} \sum_{k=0}^{n-1} (T^k(\omega)) \xrightarrow{n \rightarrow \infty} \begin{cases} \frac{2}{d} \sum_{j=0}^{l-1} X_0(\omega_{2j+1}), & \text{mit Wahrscheinlichkeit } \frac{1}{2}, \text{ falls } \omega \in A_1, \\ \frac{2}{d} \sum_{j=1}^l X_0(\omega_{2j}) & , \text{ mit Wahrscheinlichkeit } \frac{1}{2}, \text{ falls } \omega \in A_2. \end{cases}$$

Beweis des Theorems 6.2.4 Es ist zu zeigen: Falls $T : \Omega \rightarrow \Omega$ ergodisch, dann ist T mischend im Mittel, d.h. für alle $A_1, A_2 \in \mathcal{F}$

$$\frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(A_1 \cap T^{-k} A_2) \xrightarrow{n \rightarrow \infty} \mathbb{P}(A_1) \mathbb{P}(A_2).$$

Sei $Y_n = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}(T^{-k} A_2) \xrightarrow[n \rightarrow \infty]{\text{Theorem 6.2.6}} \mathbb{P}(A_2)$, weil T ergodisch ist, somit auch die Folge $\{\mathbf{1}(T^{-k} A_2)\}_{k \in \mathbb{N}}$. Nach dem Satz von Lebesgue aus $\mathbf{1}(A_1) Y_n \xrightarrow{n \rightarrow \infty} \mathbf{1}(A_1) \mathbb{P}(A_2)$ folgt

$$\mathbb{E}(\mathbf{1}(A_1) Y_n) = \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(A_1 \cap T^{-k} A_2) \xrightarrow{n \rightarrow \infty} \mathbb{P}(A_1) \mathbb{P}(A_2).$$

□

Lemma 6.2.6

Falls $\{X_n\}_{n \in \mathbb{N}}$ eine gleichgradig integrierbare Folge von Zufallsvariablen ist und $p_{n,i} \geq 0$, so dass $\sum_{i=1}^n p_{n,i} = 1$ für alle $n \in \mathbb{N}$, dann ist auch die Folge von Zufallsvariablen $Y_n = \sum_{i=1}^n p_{n,i} |X_i|$, $n \in \mathbb{N}$, gleichgradig integrierbar.

Ohne Beweis**Folgerung 6.2.2**

Unter den Voraussetzungen des Theorems 6.2.6 gilt

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow[n \rightarrow \infty]{L^2} \mathbb{E}(X_0 | J)$$

bzw.

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow[n \rightarrow \infty]{L^2} \mathbb{E}(X_0)$$

im ergodischen Fall.

Beweis Falls $\{X_n\}_{n \in \mathbb{N}_0}$ stationär, dann gilt $\sup_n \mathbb{E}(|X_n| \mathbf{1}(|X_n| > \varepsilon)) = \mathbb{E}(|X_0| \mathbf{1}(|X_0| > \varepsilon)) \xrightarrow{\varepsilon \rightarrow 0} 0$, weil $\mathbb{E}|X_0| < \infty$. Sei $S_n = \frac{1}{n} \sum_{k=0}^{n-1} X_k = \sum_{i=1}^n p_{n,i} X_{i-1}$, $p_{n,i} = \frac{1}{n}$, $\tilde{S}_n = \frac{1}{n} \sum_{k=0}^{n-1} X_k = \sum_{i=1}^n p_{n,i} |X_{i-1}|$. Aus dem Lemma 6.2.6 ist auch $\{\tilde{S}_n\}_{n \in \mathbb{Z}}$ gleichgradig integrierbar und nach dem Lemma 5.3.2 aus $S_k \xrightarrow[k \rightarrow \infty]{f.s.} 0$ folgt $\mathbb{E}|S_n| \leq \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}|X_k| \xrightarrow{n \rightarrow \infty} 0$. \square

6.3 Stationarität im weiteren Sinne

Sei $\{X_n\}_{n \in \mathbb{Z}}$ eine Folge von Zufallsvariablen, die stationär im weiteren Sinne ist: $\mathbb{E}|X_n|^2 < \infty$, $n \in \mathbb{N}$. $\mathbb{E}|X_n| = \text{const}$, $n \in \mathbb{N}$, $\text{cov}(X_n, X_m) = C(n - m)$, $n, m \in \mathbb{Z}$.

6.3.1 Korrelationstheorie

Theorem 6.3.1 (Herglotz):

Sei $C : \mathbb{Z} \rightarrow \mathbb{R}$ eine positiv semi-definite Funktion. Dann existiert ein endliches Maß μ auf $(-\pi, \pi)$, so dass

$$C(n) = \int_{-\pi}^{\pi} e^{inx} \mu(dx), \quad n \in \mathbb{Z}.$$

μ heißt *Spektralmaß* von C .

Bemerkung 6.3.1

Da die Kovarianzfunktion einer stationären Folge positiv semi-definit ist, gilt die obige Darstellung für eine beliebige Kovarianzfunktion C .

Definition 6.3.1

Eine Familie $\{Q_\lambda, \lambda \in \Lambda\}$ von Wahrscheinlichkeitsmaßen heißt *schwach relativ kompakt*, falls eine beliebige Folge von Maßen $\{Q_{\lambda_n}\}_{n \in \mathbb{N}}$ eine Teilfolge $\{Q_{\lambda_{n_k}}\}_{n \in \mathbb{N}}$ besitzt, die schwach konvergiert.

Definition 6.3.2

Eine Familie von Wahrscheinlichkeitsmaßen $Q = \{Q_\lambda, \lambda \in \Lambda\}$ auf $(\mathcal{S}, \mathcal{B})$, \mathcal{B} – Borelsche σ -Algebra auf einem metrischen Raum \mathcal{S} heißt *dicht*, falls für alle $\varepsilon > 0$ ein Kompaktum existiert, so dass $K_\varepsilon \in \mathcal{B}$ und $Q_\lambda(K_\varepsilon) > 1 - \varepsilon$ für alle $\lambda \in \Lambda$.

Theorem 6.3.2 (Prokhorov):

Falls die Familie von Wahrscheinlichkeitsmaßen $Q = \{Q_\lambda, \lambda \in \Lambda\}$ auf dem metrischen messbaren Raum $(\mathcal{S}, \mathcal{B})$ dicht ist, dann ist sie schwach relativ kompakt. Falls \mathcal{S} ein Banachraum ist, dann ist jede schwach relativ kompakte Familie $Q = \{Q_\lambda, \lambda \in \Lambda\}$ von Maßen auch dicht.

Ohne Beweis

Der Satz von Prokhorov wird verwendet um die schwache Konvergenz einer Folge von Wahrscheinlichkeitsmaßen zu beweisen, indem man unter anderem ihre Dichtheit prüft. Insbesondere, falls \mathcal{S} kompakt ist, ist jede Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{S}, \mathcal{B})$ dicht, denn $K_\varepsilon = \mathcal{S}$ für alle $\varepsilon > 0$.

Beweis des Theorems 6.3.2 „ \Leftarrow “

Falls $C(n) = \int_{-\pi}^{\pi} e^{inx} \mu(dx)$, $n \in \mathbb{Z}$, dann für alle $n \in \mathbb{N}$, für alle $z_1, \dots, z_n \in \mathbb{C}$ und $t_1, \dots, t_n \in \mathbb{Z}$

$$\sum_{i,j=1}^n z_j \bar{z}_i C(t_i - t_j) = \int_{-\pi}^{\pi} \left| \sum_{i=1}^n z_i e^{iz_i x} \right|^2 \mu(dx) \geq 0.$$

Daraus folgt, dass C positiv semi-definit ist.

„ \Rightarrow “

Für alle $N \geq 1$, $x \in [-\pi, \pi]$, definiere die Funktion $g_N(x) = \frac{1}{2\pi N} \sum_{k,j=1}^N C(k-j) e^{-ikx} e^{ijx} \geq 0$, die stetig in x ist, weil C positiv semi-definit ist. Es gilt

$$g_N(x) = \frac{1}{2\pi} \sum_{|n| < N} \left(1 - \frac{|n|}{N}\right) C(n) e^{-inx},$$

weil es $N - |n|$ Paare $(k, j) \in \{1, \dots, N\}^2$ gibt, so dass $k - j = n$. Definiere Maß μ_N auf $([-\pi, \pi], \mathcal{B}_{[-\pi, \pi]})$ durch $\mu_N(B) = \int g_N(x) dx$, $B \in \mathcal{B}_{[-\pi, \pi]}$.

$$\int_{-\pi}^{\pi} e^{inx} Q_N(dx) = \int_{-\pi}^{\pi} e^{inx} g_N(x) dx = \begin{cases} \left(1 - \frac{|n|}{N}\right) C(n), & |n| < N, \\ 0, & \text{sonst,} \end{cases}$$

weil $\{e^{inx}\}_{n \in \mathbb{Z}}$ ein orthogonales System in $L^2[-\pi, \pi]$ ist. Für $n = 0$ gilt $Q_N([-\pi, \pi]) = C(0) < \infty$, deshalb ist $\left\{\frac{Q_N}{C(0)}\right\}_{n \in \mathbb{N}}$ eine Familie von Wahrscheinlichkeitsmaßen, die dicht ist. Nach dem Theorem 6.3.2 existiert eine Teilfolge $\{N_k\}_{k \in \mathbb{N}}$, so dass $Q_{N_k} \xrightarrow[k \rightarrow \infty]{\omega} \mu$. μ - endliches Maß auf $[-\pi, \pi]$ und daraus folgt

$$\lim_{k \rightarrow \infty} \int_{-\pi}^{\pi} e^{inx} g_{N_k}(x) dx = \lim_{k \rightarrow \infty} \left(1 - \frac{|n|}{N_k}\right) C(n) = C(n), \quad \text{für alle } n \in \mathbb{Z}.$$

□

Sei $X = \{X_n\}_{n \in \mathbb{Z}}$ eine stationäre im weiten Sinne Folge von Zufallsvariablen. Dann gilt folgende Spektraldarstellung:

$$X_n \stackrel{d}{=} \int_{-\pi}^{\pi} e^{inx} Z(dx), \quad n \in \mathbb{Z},$$

wobei Z ein orthogonales Zufallsmaß auf $([-\pi, \pi], \mathcal{B}_{[-\pi, \pi]})$ ist. Daher soll sowohl Z als auch $I(f) = \int_{-\pi}^{\pi} f(x) Z(dx)$ für deterministische Funktionen $f: [-\pi, \pi] \rightarrow \mathbb{C}$ eingeführt werden.

6.3.2 Orthogonale Zufallsmaße

Konstruktionsschema von Z bzw. $I(\cdot)$:

1. Z wird auf einem Simiring \mathcal{K} (der Teilmengen von Λ) definiert.
2. Z wird erweitert auf die Algebra A , die von \mathcal{K} erzeugt wird.
3. Definiere das Integral I bzgl. Z für einfache Funktionen auf $\sigma(A)$, wenn das Maß $\mu(\Lambda) < \infty$, μ - gegebenes Maß.

4. Definiere I als $\lim_{n \rightarrow \infty} I(f_n)$ für beliebige meßbare Funktionen f , $f = \lim_{n \rightarrow \infty} f_n$, f_n einfach, $\mu(\Lambda) < \infty$.
5. Definiere I auf einem σ -endlichem Raum $\Lambda = \cup_n \Lambda_n$, $\mu(\Lambda_n) < \infty$, $\Lambda_n \cap \Lambda_m = \emptyset$, $n \neq m$, als $I(f) = \sum_n I(f | \Lambda_n)$, I_n - Integral bzgl. Z auf Λ_n . Dadurch sind Z auf $\{A \in \sigma(A) : \mu(A) < \infty\}$ erweitert als $Z(A) = I(1(A))$.

Schritt 1

Sei \mathcal{K} ein Semiring der Teilmengen von Λ (Λ - beliebiger Raum), d.h. für alle $A, B \in \mathcal{K}$ gilt $A \cap B \in \mathcal{K}$; falls $A \subset B$, dann existieren $A_1, \dots, A_n \in \mathcal{K}$, $A_i \cap A_j = \emptyset$, $i \neq j$, so dass $B = A \cup \cup_{i=1}^n A_i$.

Definition 6.3.3 1. Ein komplexwertiges signiertes Zufallsmaß $Z = \{Z(B), B \in \mathcal{K}\}$, gegeben auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, heißt orthogonal, wenn

- a) alle $Z(B) \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $B \in \mathcal{K}$,
 - b) $A, B \in \mathcal{K}$, $A \cap B = \emptyset \Rightarrow \langle Z(A), Z(B) \rangle_{L^2(\Omega, \mathcal{F}, \mathbb{P})} = \mathbf{E}(Z(A), \overline{Z(B)}) = 0$,
 - c) als Zufallsmaß gilt die σ -Additivität von Z : Falls $B, B_1, \dots, B_n, \dots \in \mathcal{K}$, $B = \cup_n B_n$, $B_i \cap B_j = \emptyset$, $i \neq j$, $Z(B) \stackrel{f.s.}{=} \sum_n Z(B_n)$, wobei die Konvergenz dieser Reihe im $L^2(\Omega, \mathcal{F}, \mathbb{P})$ -Sinne zu verstehen ist.
2. Die Größe $\mu = \{\mu(B), B \in \mathcal{K}\}$ definiert durch $\mu(B) = \mathbf{E}|Z(B)|^2 = \langle Z(B), Z(B) \rangle_{L^2(\Omega, \mathcal{F}, \mathbb{P})}$, $B \in \mathcal{K}$, heißt Strukturmaß von Z . Es ist leicht zu sehen, dass μ tatsächlich ein Maß auf \mathcal{K} ist. Falls $\Lambda \in \mathcal{K}$, dann ist μ endlich, ansonsten σ -endlich, $\Lambda = \cup_n \Lambda_n$, $\Lambda_n \in \mathcal{K}$, $\Lambda_n \cap \Lambda_m = \emptyset$, so dass $\mu(\Lambda_n) < \infty$.

3. Das orthogonale Zufallsmaß Z heißt *zentriert*, falls $\mathbf{E}Z(B) = 0$, $B \in \mathcal{K}$.

Beispiel 6.3.1

Sei $\Lambda = [0, \infty)$, $\mathcal{K} = \{[a, b), 0 \leq a < b < \infty\}$, $Z([a, b)) = W(b) - W(a)$, $0 \leq a < b < \infty$, wobei $W = \{W(t), t \geq 0\}$ der Wiener-Prozess ist. Z ist ein orthogonales Zufallsmaß auf \mathcal{K} , weil W unabhängige Zuwächse hat. Analog kann diese Definition auf einem beliebigen quadratisch integrierbaren stochastischen Prozess X mit unabhängigen Zuwächsen an Stelle von W übertragen werden.

Schritt 2

Theorem 6.3.3

Sei μ ein σ -endliches Maß auf der Algebra A , die von \mathcal{K} erzeugt wird (nach dem Theorem von Caratheodon wird μ eindeutig auf $\sigma(A)$ fortgesetzt). Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und ein zentriertes orthogonales Zufallsmaß Z auf $(\Omega, \mathcal{F}, \mathbb{P})$, definiert auf $\{B \in A : \mu(B) < \infty\}$, mit Strukturmaß (oder Kontrollmaß) μ .

Ohne Beweis

Zur Definition von Z auf A : für $B \in A$, $B = \cup_{i=1}^n B_i$, $B_i \in \mathcal{K}$, $B_i \cap B_j = \emptyset$, $i \neq j$, wird $Z(B) = \sum_{i=1}^n Z(B_i)$ gesetzt.

6.3.3 Integral bezüglich eines orthogonalen Zufallsmaßes

Schritt 3

Sei $f : \Lambda \rightarrow \mathbb{C}$ eine einfache Funktion, d.h. $f(x) = \sum_{i=1}^n c_i 1(x \in B_i)$, für $c_i \in \mathbb{C}$ und $B_i \in \mathcal{E}$, $i = 1, \dots, n$, so dass $\cup_{i=1}^n B_i = \Lambda$, $B_i \cap B_j = \emptyset$, $i \neq j$, und $(\Lambda, \mathcal{E}, \mu)$ ein meßbarer Raum mit $\mu(\Lambda) < \infty$.

Definition 6.3.4

Das Integral von f bezüglich eines orthogonalen Zufallsmaßes Z definiert auf $(\Omega, \mathcal{F}, \mathbb{P})$ ist gegeben durch $I(f) := \int_{\Lambda} f(x) Z(dx) = \sum_{i=1}^n c_i Z(B_i)$.

Aufgabe 6.3.1

Zeigen Sie, dass die Definition korrekt ist, d.h. $I(f)$ hängt nicht von der Darstellung von f als einfache Funktion ab.

Lemma 6.3.1 (Eigenschaften von I):

Sei $I(\cdot)$ das Integral bzgl. des orthogonalen Zufallsmaßes, definiert auf einer einfachen Funktion $\Lambda \rightarrow \mathbb{C}$ wie oben. Es gelten folgende Eigenschaften:

1. *Isometrie:* $\langle I(f), I(g) \rangle_{L^2(\Omega)} = \langle f, g \rangle_{L^2(\Lambda)}$, wobei f und g einfache Funktionen $\Lambda \rightarrow \mathbb{C}$ sind, $\langle f, g \rangle_{L^2(\Lambda)} = \int_{\Lambda} f(x) \overline{g(x)} \mu(dx)$.
2. *Linearität:* Für jede einfache Funktion $f, g : \Lambda \rightarrow \mathbb{C}$ gilt $I(f + g) \stackrel{f.s.}{=} I(f) + I(g)$.

Aufgabe 6.3.2

Beweisen Sie es.

Schritt 4

Sei nun $f \in L^2(\Omega, \mathcal{E}, \mu)$. Dann existiert eine Folge von einfachen Funktionen $f_n : \Lambda \rightarrow \mathbb{C}$, so dass $f_n \xrightarrow[n \rightarrow \infty]{L^2(\Lambda)} f$ (einfache Funktion ist dicht in $L^2(\Lambda)$). Dann definiere $I(f) = \lim_{n \rightarrow \infty} I(f_n)$, wobei dieser Grenzwert in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ -Sinne zu verstehen ist. Man kann zeigen, dass die Definition von $I(f)$ unabhängig von der Wahl der Folge $\{f_n\}$ ist.

Lemma 6.3.2

Die Aussagen des Lemmas 6.3.1 gelten im allgemeinen Fall.

Beweis Benutze die Stetigkeit $\langle \cdot, \cdot \rangle$. □

Bemerkung 6.3.2

Falls Z zentriert ist, dann gilt $\mathbf{E}I(f) = 0$ für beliebige Funktionen $f \in L^2(\Lambda, \mathcal{E}, \mu)$.

Schritt 5

Sei nun Λ σ -endlich, d.h. $\Lambda = \cup_n \Lambda_n$, $\mu(\Lambda_n) < \infty$, $\Lambda_n \cap \Lambda_m = \emptyset$, $n \neq m$. Dann für alle $f \in L^2(\Lambda, \mathcal{E}, \mu)$ gilt $f = \sum_n f|_{\Lambda_n}$. Auf $L^2(\Lambda_n, \mathcal{E} \cap \Lambda_n, \mu)$ wird das Integral I_n bzgl. Z wie im 1)-4) definiert. Dann setze $I(f) := \sum_n I_n(f|_{\Lambda_n})$.

Theorem 6.3.4

Die Abbildung $g : L^2(\Lambda, \mathcal{E}, \mu) \rightarrow L^2(\Omega, \mathcal{F}, \mathbb{P})$ ist eine Isometrie. Insbesondere kann dadurch das Zufallsmaß Z auf $\{B \in \mathcal{E} : \mu(B) < \infty\}$ fortgesetzt werden als $Z(B) := I(1_B)$, $B \in \mathcal{E} : \mu(B) < \infty$.

6.3.4 Spektraldarstellung

Sei $X = \{X(t), t \in T\}$ ein beliebiger komplexwertiger stochastischer Prozess auf $(\Omega, \mathcal{F}, \mathbb{P})$, T – eine beliebige Indexmenge, $\mathbf{E}|X(t)|^2 < \infty$, $t \in T$, $\mathbf{E}X(t) = 0$, $t \in T$ (o.B.d.A., sonst betrachte $\tilde{X}(t) = X(t) - \mathbf{E}X(t)$), $t \in T$, mit $C(s, t) = \mathbf{E}(X(s), \overline{X(t)})$, $s, t \in T$.

Theorem 6.3.5 (Karhunen):

X hat die Spektraldarstellung $X(t) = \int_{\Lambda} f(t, x)Z(dx)$, $t \in T$ (d.h., es existiert ein zentriertes orthogonales Zufallsmaß auf $\{B \in \mathcal{E} : \mu(B) < \infty\}$, wobei $L^2(\Lambda, \mathcal{E}, \mu)$ ein wie oben definierter Raum ist), genau dann, wenn es ein System der Funktionen $f(t, \cdot) \in L^2(\Lambda, \mathcal{E}, \mu)$, $t \in T$, existiert, so dass $C(s, t) = \int_{\Lambda} f(s, x)f(t, x)\mu(dx)$, $s, t \in T$, und dieses System F vollständig in $L^2(\Lambda, \mathcal{E}, \mu)$ ist (d.h. $\langle f(t, \cdot), \psi \rangle_{L^2(\Omega)} = 0$, $\psi \in L^2(\Omega, \mathcal{E}, \mu)$, für alle $t \in T$ und $\psi \equiv 0$, μ fast überall).

Ohne Beweis

Theorem 6.3.6

Sei $\{X_n, n \in \mathbb{Z}\}$ eine zentrierte komplexwertige stationäre im weiten Sinne Folge von Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann existiert ein orthogonales zentriertes Zufallsmaß auf $([-\pi, \pi], \mathcal{B}([-\pi, \pi]))$ (definiert auf $(\Omega, \mathcal{F}, \mathbb{P})$), so dass $X_n \stackrel{f.s.}{=} \int_{-\pi}^{\pi} e^{inx} Z(dx)$, $n \in \mathbb{Z}$.

Beweis Sei $F = \{e^{inx}, x \in [-\pi, \pi], n \in \mathbb{Z}\}$. Dieses System ist vollständig auf $L^2([-\pi, \pi])$ (vgl. die Theorie der Fourier-Reihen). Aus dem Theorem von Herglotz folgt, dass

$$C(n, m) = \mathbb{E}(X_n \bar{X}_m) = \int_{-\pi}^{\pi} e^{inx} e^{imx} \mu(dx),$$

wobei μ das Spektralmaß von X ist, also ein endliches Maß auf $([-\pi, \pi], \mathcal{B}([-\pi, \pi]))$. Nach dem Theorem 6.3.5 existiert ein orthogonales Zufallsmaß auf $(\Omega, \mathcal{F}, \mathbb{P})$, so dass $X_n \stackrel{f.s.}{=} \int_{-\pi}^{\pi} e^{inx} Z(dx)$, $n \in \mathbb{Z}$. \square

Theorem 6.3.7 (Ergodensatz für stationäre (im weiten Sinne) Folgen von Zufallsvariablen):

Unter den Voraussetzungen des Theorems 6.3.6 gilt

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{L^2(\Omega)} Z(\{0\}).$$

Insbesondere wenn X nicht zentriert ist, d.h. $\mathbb{E}X_n = a$, $n \in \mathbb{Z}$, dann konvergiert $\frac{1}{n} \sum_{k=0}^{n-1} X_k \xrightarrow{L^2(\Omega)} a$ dann, wenn $\underbrace{\mathbb{E}|Z(\{0\})|^2}_{\mu(\{0\})} = 0$, also Z und somit μ hat kein Atom im Null.

Beweis $S_n = \frac{1}{n} \sum_{k=0}^{n-1} X_k = \int \underbrace{\frac{1}{n} \sum_{k=0}^{n-1} e^{ikx}}_{\psi_n(x)} Z(dx)$. $\psi_n(x) = \begin{cases} \frac{1}{n} \frac{1-e^{inx}}{1-e^{ix}}, & x \neq 0 \\ 1, & x = 0 \end{cases}$, für alle $n \in \mathbb{N}$.

\mathbb{N} . $S_n - Z(\{0\}) = \int_{-\pi}^{\pi} \underbrace{(\psi_n(x) - 1(x=0))}_{\varphi_n(x)} Z(dx) = \int_{-\pi}^{\pi} \varphi_n(x) Z(dx)$. $\|S_n - Z(\{0\})\|_{L^2(\Omega)}^2 = \|\varphi_n(x)\|_{L^2([-\pi, \pi], \mu)}^2 = \int_{-\pi}^{\pi} |\varphi_n(x)|^2 \mu(dx) \xrightarrow{n \rightarrow \infty} 0$ nach dem Theorem von Lebesgue, weil $|\varphi_n(x)| \leq \frac{2}{n|1-e^{ix}|} \xrightarrow{n \rightarrow \infty} 0$ für alle $x \in [-\pi, \pi]$. \square

6.4 Ergänzende Aufgaben**Aufgabe 6.4.1**

Sei Z_1, Z_2, \dots eine Folge von Zufallsvariablen, so dass die Reihe $\sum_{i=1}^{\infty} Z_i$ fast sicher konvergiert.

Sei a_1, a_2, \dots eine monoton wachsende Folge positiver (deterministischer) Zahlen mit $a_n \rightarrow \infty$, $n \rightarrow \infty$. Zeige, dass

$$\frac{1}{a_n} \sum_{k=1}^n a_k Z_k \xrightarrow{f.s.} 0, \quad n \rightarrow \infty.$$

Aufgabe 6.4.2

Sei X eine nicht-negative Zufallsvariable über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und $T : \Omega \rightarrow \Omega$ eine maßerhaltende Abbildung. Zeige, dass

$$\sum_{k=0}^{\infty} X(T^k(\omega)) = \infty \quad f.s.$$

für fast alle $\omega \in \Omega$ mit $X(\omega) > 0$.

Aufgabe 6.4.3

Sei X eine Zufallsvariable über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und $T : \Omega \rightarrow \Omega$ eine maßerhaltende Abbildung. Zeige, dass $\mathbb{E}X = \mathbb{E}(X \circ T)$, d. h.

$$\int_{\Omega} X(T(\omega)) \mathbb{P}(d\omega) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

(Hinweis: algebraische Induktion)

Aufgabe 6.4.4

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, wobei $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}([0, 1))$ und \mathbb{P} das Lebesguemaß ist. Sei $\lambda \in (0, 1)$.

- Zeige, dass $T(x) = (x + \lambda) \bmod 1$ eine maßerhaltende Abbildung ist, wobei $a \bmod m = a - \lfloor \frac{a}{m} \rfloor m$ für $a \in \mathbb{R}$ und $b \in \mathbb{Z}$ und $\lfloor \cdot \rfloor$ die Gaußklammer bezeichnet.
- Zeige, dass $T(x) = \lambda x$ und $T(x) = x^2$ keine maßerhaltende Abbildungen sind.

Literaturverzeichnis

- [1] D. Applebaum. *Lévy processes and stochastic calculus*. Cambridge University Press, Cambridge, 2004.
- [2] A. A. Borovkov. *Wahrscheinlichkeitstheorie. Eine Einführung*. Birkhäuser, Basel, 1976.
- [3] W. Feller. *An introduction to probability theory and its applications. Vol. I/II*. John Wiley & Sons Inc., New York, 1970/71.
- [4] P. Gänszler and W. Stute. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1977.
- [5] G. Grimmett and D. Stirzaker. *Probability and random processes*. Oxford University Press, New York, 2001.
- [6] C. Hesse. *Angewandte Wahrscheinlichkeitstheorie*. Vieweg, Braunschweig, 2003.
- [7] O. Kallenberg. *Foundations of modern probability*. Springer, New York, 2002.
- [8] J. F. C. Kingman. *Poisson processes*. Oxford University Press, New York, 1993.
- [9] N. Krylov. *Introduction to the theory of random processes*. American Mathematical Society, Providence, RI, 2002.
- [10] S. Resnick. *Adventures in stochastic processes*. Birkhäuser, Boston, MA, 1992.
- [11] G. Samorodnitsky and M. Taqqu. *Stable non-Gaussian random processes*. Chapman & Hall, New York, 1994.
- [12] K.-I. Sato. *Lévy Processes and Infinite Divisibility*. Cambridge University Press, Cambridge, 1999.
- [13] A. N. Shiryaev. *Probability*. Springer, New York, 1996.
- [14] A. V. Skorokhod. *Basic principles and applications of probability theory*. Springer, Berlin, 2005.
- [15] J. Stoyanov. *Counterexamples in probability*. John Wiley & Sons Ltd., Chichester, 1987.
- [16] A. D. Wentzel. *Theorie zufälliger Prozesse*. Birkhäuser, Basel, 1979.



ulm university universität
uulm

Stochastik III

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

ULM
2012

Inhaltsverzeichnis

1	Tests statistischer Hypothesen	3
1.1	Allgemeine Philosophie des Testens	3
1.2	Nichtrandomisierte Tests	12
1.2.1	Parametrische Signifikanztests	12
1.3	Randomisierte Tests	17
1.3.1	Grundlagen	17
1.3.2	Neyman-Pearson-Tests bei einfachen Hypothesen	19
1.3.3	Einseitige Neyman-Pearson-Tests	24
1.3.4	Unverfälschte zweiseitige Tests	30
1.4	Anpassungstests	36
1.4.1	χ^2 -Anpassungstest	36
1.4.2	χ^2 -Anpassungstest von Pearson-Fisher	42
1.4.3	Anpassungstest von Shapiro	48
1.5	Weitere, nicht parametrische Tests	50
1.5.1	Binomialtest	50
1.5.2	Iterationstests auf Zufälligkeit	52
2	Lineare Regression	55
2.1	Multivariate Normalverteilung	55
2.1.1	Eigenschaften der multivariaten Normalverteilung	59
2.1.2	Lineare und quadratische Formen von normalverteilten Zufallsvariablen	60
2.2	Multivariate lineare Regressionsmodelle mit vollem Rang	68
2.2.1	Methode der kleinsten Quadrate	68
2.2.2	Schätzer der Varianz σ^2	73
2.2.3	Maximum-Likelihood-Schätzer für β und σ^2	75
2.2.4	Tests für Regressionsparameter	78
2.2.5	Konfidenzbereiche	81
2.3	Multivariate lineare Regression mit $\text{Rang}(X) < m$	84
2.3.1	Verallgemeinerte Inverse	85
2.3.2	MKQ-Schätzer für β	86
2.3.3	Erwartungstreu schätzbare Funktionen	89
2.3.4	Normalverteilte Störgrößen	92
2.3.5	Hypothesentests	96
2.3.6	Konfidenzbereiche	98
2.3.7	Einführung in die Varianzanalyse	100

3	Verallgemeinerte lineare Modelle	103
3.1	Exponentialfamilie von Verteilungen	103
3.2	Linkfunktion	107
3.3	Maximum-Likelihood-Schätzung von β	109
3.4	Asymptotische Tests für β	116
3.5	Kriterien zur Modellwahl bzw. Modellanpassung	122
4	Hauptkomponentenanalyse	125
4.1	Einführung	125
4.2	Hauptkomponentenanalyse auf Modellebene	126
4.3	Hauptkomponentenanalyse auf Datenebene	134
4.4	Asymptotische Verteilung von HK bei normalverteilten Stichproben	138
4.5	Ausreißerererkennung	140
4.6	Hauptkomponentenanalyse und Regression	142
4.7	Numerische Berechnung der Hauptkomponenten	148
	Literatur	151

Vorwort

Dieses Skript entstand aus dem Zyklus der Vorlesungen über Statistik, die ich in den Jahren 2006-2010 an der Universität Ulm gehalten habe. Dabei handelt es sich um die aufbauende Vorlesung Stochastik III, die auf der Vorlesung *Stochastik I* basiert.

Ich möchte gerne meinen Kollegen aus dem Institut für Stochastik, Herrn Prof. Volker Schmidt und Herrn Dipl.-Math. Malte Spiess, für ihre Unterstützung und anregenden Diskussionen während der Entstehung des Skriptes danken. Herr Marco Baur hat eine hervorragende Arbeit beim Tippen des Skriptes und bei der Erstellung zahlreicher Abbildungen, die den Text begleiten, geleistet. Dafür gilt ihm mein herzlicher Dank.

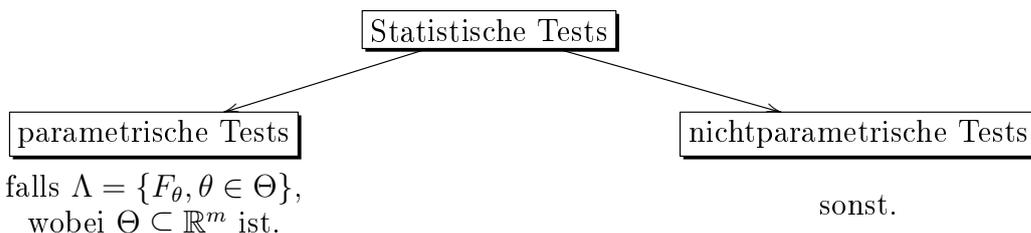
Ulm, den 20.10.2011
Evgeny Spodarev

1 Tests statistischer Hypothesen

In der Vorlesung Stochastik I haben wir schon Beispiele von statistischen Tests kennengelernt, wie etwa den Kolmogorow-Smirnow-Test (vergleiche Bemerkung 3.3.38, 3), Skript Stochastik I). Jetzt sollen statistische Signifikanztests formal eingeführt und ihre Eigenschaften untersucht werden.

1.1 Allgemeine Philosophie des Testens

Es sei eine Zufallsstichprobe (X_1, \dots, X_n) von unabhängigen, identisch verteilten Zufallsvariablen X_i gegeben, mit Verteilungsfunktion $F \in \Lambda$, wobei Λ eine Klasse von Verteilungsfunktionen ist. Es sei (x_1, \dots, x_n) eine konkrete Stichprobe, die als Realisierung von (X_1, \dots, X_n) interpretiert wird. In der Theorie des statistischen Testens werden Hypothesen über die Beschaffenheit der (unbekannten) Verteilungsfunktion F gestellt und geprüft. Dabei unterscheidet man



Bei parametrischen Tests prüft man, ob der Parameter θ bestimmte Werte annimmt (zum Beispiel $\theta = 0$). Bekannte Beispiele von nichtparametrischen Tests sind Anpassungstests, bei denen man prüft, ob die Verteilungsfunktion F gleich einer vorgegebenen Funktion F_0 ist.

Formalisieren wir zunächst den Begriff *Hypothese*. Die Menge Λ von zulässigen Verteilungsfunktionen F wird in zwei disjunkte Teilmengen Λ_0 und Λ_1 zerlegt, $\Lambda_0 \cup \Lambda_1 = \Lambda$. Die Aussage

„Man testet die *Haupthypothese* $H_0 : F \in \Lambda_0$ gegen die *Alternative* $H_1 : F \in \Lambda_1$,“

bedeutet, daß man an Hand der konkreten Stichprobe (x_1, \dots, x_n) versucht, eine Entscheidung zu fällen, ob die Verteilungsfunktion der Zufallsvariable X_i zu Λ_0 oder zu Λ_1 gehört. Dies passiert auf Grund einer statistischen *Entscheidungsregel*

$$\varphi : \mathbb{R}^n \rightarrow [0, 1],$$

die eine Statistik mit folgender Interpretation ist:

Der Stichprobenraum \mathbb{R}^n wird in drei disjunkte Bereiche K_0, K_{01} und K_1 unterteilt, sodaß $\mathbb{R}^n = K_0 \cup K_{01} \cup K_1$, wobei

$$\begin{aligned} K_0 &= \varphi^{-1}(\{0\}) &= \{x \in \mathbb{R}^n : \varphi(x) = 0\}, \\ K_1 &= \varphi^{-1}(\{1\}) &= \{x \in \mathbb{R}^n : \varphi(x) = 1\}, \\ K_{01} &= \varphi^{-1}((0, 1)) &= \{x \in \mathbb{R}^n : 0 < \varphi(x) < 1\}. \end{aligned}$$

Dementsprechend wird $H_0 : F \in \Lambda_0$

- verworfen, falls $\varphi(x) = 1$, also $x \in K_1$,
- nicht verworfen, falls $\varphi(x) = 0$, also $x \in K_0$;
- falls $\varphi(x) \in (0, 1)$, also $x \in K_{01}$, wird $\varphi(x)$ als Bernoulli-Wahrscheinlichkeit interpretiert, und es wird eine Zufallsvariable $Y \sim \text{Bernoulli}(\varphi(x))$ generiert, für die gilt:

$$Y = \begin{cases} 1 & \implies H_0 \text{ wird verworfen} \\ 0 & \implies H_0 \text{ wird nicht verworfen} \end{cases}$$

Falls $K_{01} \neq \emptyset$, wird eine solche Entscheidungsregel *randomisiert* genannt. Bei $K_{01} = \emptyset$, also $\mathbb{R}^n = K_0 \cup K_1$ spricht man dagegen von *nicht-randomisierten* Tests. Dabei heißt K_0 bzw. K_1 *Annahmereich* bzw. *Ablehnungsbereich (kritischer Bereich)* von H_0 . K_{01} heißt *Randomisierungsbereich*.

Bemerkung 1.1.1. 1. Man sagt absichtlich „ H_0 wird nicht verworfen“, statt „ H_0 wird akzeptiert“, weil die schließende Statistik generell keine positiven, sondern nur negative Entscheidungen treffen kann. Dies ist generell ein philosophisches Problem der Falsifizierbarkeit von Hypothesen oder wissenschaftlichen Theorien, von denen aber keiner behaupten kann, daß sie der Wahrheit entsprechen (vergleiche die *wissenschaftliche Erkenntnistheorie von Karl Popper (1902-1994)*).

2. Die randomisierten Tests sind hauptsächlich von theoretischem Interesse (vergleiche Abschnitt 2.3). In der Praxis werden meistens nichtrandomisierte Regeln verwendet, bei denen man aus der Stichprobe (x_1, \dots, x_n) allein die Entscheidung über H_0 treffen kann. Hier gilt $\varphi(x) = \mathbb{1}_{K_1}, x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

In diesem und in folgendem Abschnitt betrachten wir ausschließlich nichtrandomisierte Tests, um in Abschnitt 2.3 zu der allgemeinen Situation zurückzukehren.

Definition 1.1.1. Man sagt, daß die nicht-randomisierte Testregel $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$ einen (*nichtrandomisierten*) *statistischen Test zum Signifikanzniveau α* angibt, falls für $F \in \Lambda_0$ gilt

$$\mathbb{P}_F(\varphi(X_1, \dots, X_n) = 1) = P(H_0 \text{ verwerfen} \mid H_0 \text{ richtig}) \leq \alpha.$$

Definition 1.1.2. 1. Wenn man H_0 verwirft, obwohl H_0 richtig ist, begeht man den sogenannten *Fehler 1. Art*. Die Wahrscheinlichkeit

$$\alpha_n(F) = \mathbb{P}_F(\varphi(x_1, \dots, x_n) = 1), \quad F \in \Lambda_0$$

heißt die *Wahrscheinlichkeit des Fehlers 1. Art* und soll unter dem Niveau α bleiben.

2. Den *Fehler 2. Art* begeht man, wenn man die falsche Hypothese H_0 nicht verwirft. Dabei ist

$$\beta_n(F) = \mathbb{P}_F(\varphi(x_1, \dots, x_n) = 0), \quad F \in \Lambda_1$$

die *Wahrscheinlichkeit des Fehlers 2. Art*.

Eine Zusammenfassung aller Möglichkeiten wird in folgender Tabelle festgehalten:

	H_0 richtig	H_0 falsch
H_0 verwerfen	Fehler 1. Art, Wahrscheinlichkeit $\alpha_n(F) \leq \alpha$	richtige Entscheidung
H_0 nicht verwerfen	richtige Entscheidung	Fehler 2. Art mit Wahrscheinlichkeit $\beta_n(F)$

Dabei sollen α_n und β_n möglichst klein sein, was gegenläufige Tendenzen darstellt, weil beim Kleinwerden von α die Wahrscheinlichkeit des Fehlers 2. Art notwendigerweise wächst.

Definition 1.1.3. 1. Die Funktion

$$G_n(F) = \mathbb{P}_F(\varphi(X_1, \dots, X_n) = 1), \quad F \in \Lambda$$

heißt *Gütefunktion* eines Tests φ .

2. Die Einschränkung von G_n auf Λ_1 heißt *Stärke*, *Schärfe* oder *Macht* (englisch *power*) des Tests φ .

Es gilt

$$\begin{cases} G_n(F) = \alpha_n(F) \leq \alpha, & F \in \Lambda_0 \\ G_n(F) = 1 - \beta_n(F), & F \in \Lambda_1 \end{cases}$$

Beispiel 1.1.1. Parametrische Tests. Wie sieht ein parametrischer Test aus? Der Parameterraum Θ wird als $\Theta_0 \cup \Theta_1$ dargestellt, wobei $\Theta_0 \cap \Theta_1 = \emptyset$. Es gilt $\Lambda_0 = \{F_\theta : \theta \in \Theta_0\}$, $\Lambda_1 = \{F_\theta : \theta \in \Theta_1\}$. P_F wird zu P_θ , α_n , G_n und β_n werden statt auf Λ auf Θ definiert.

Welche Hypothesen H_0 und H_1 kommen oft bei parametrischen Tests vor? Zur Einfachheit betrachten wir den Spezialfall $\Theta = \mathbb{R}$.

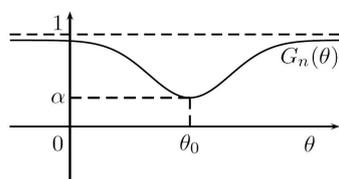
1. $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$

2. $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$
3. $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
4. $H_0 : \theta \in [a, b]$ vs. $H_1 : \theta \notin [a, b]$

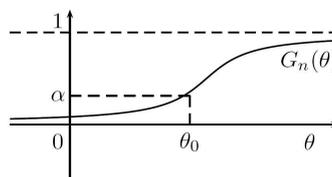
Im Fall (1) heißt der parametrische Test *zweiseitig*, in den Fällen (2) und (3) *einseitig* (*rechts-* bzw. *linksseitig*). In Fall (4) spricht man von der *Intervallhypothese* H_0 .

Bei einem zweiseitigen bzw. einseitigen Test kann die Gütefunktion wie in Abbildung 1.1 (a) bzw. 1.1 (b) aussehen,

Abbildung 1.1: Gütefunktion



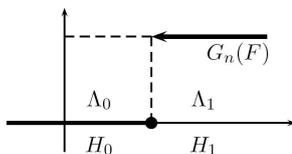
(a) eines zweiseitigen Tests



(b) eines einseitigen Tests

Bei einem allgemeinen (nicht notwendigerweise parametrischen) Modell kann man die ideale Gütefunktion wie in Abbildung 1.2 schematisch darstellen.

Abbildung 1.2: Schematische Darstellung der idealen Gütefunktion



- Man sieht aus Definition 1.1.2, dem Fehler 1. und 2. Art und der Ablehnungsregel, daß die Hypothesen H_0 und H_1 nicht symmetrisch behandelt werden, denn nur die

Wahrscheinlichkeit des Fehlers 1. Art wird kontrolliert. Dies ist der Grund dafür, daß Statistiker die eigentlich interessierende Hypothese nicht als H_0 , sondern als H_1 formulieren, damit, wenn man sich für H_1 entscheidet, man mit Sicherheit sagen kann, daß die Wahrscheinlichkeit der Fehlentscheidung unter dem Niveau α liegt.

- Wie wird ein statistischer, nicht randomisierter Test praktisch konstruiert? Die Konstruktion der Ablehnungsregel φ ähnelt sich sehr der von Konfidenzintervallen:
 1. Finde eine Teststatistik $T : \mathbb{R}^n \rightarrow \mathbb{R}$, die unter H_0 eine (möglicherweise asymptotisch für $n \rightarrow \infty$) bestimmte Prüfverteilung hat.
 2. Definiere $B_0 = [t_{\alpha_1}, t_{1-\alpha_2}]$, wobei t_{α_1} und $t_{1-\alpha_2}$ Quantile der Prüfverteilung von T sind, $\alpha_1 + \alpha_2 = \alpha \in [0, 1]$.
 3. Falls $T(X_1, \dots, X_n) \in \mathbb{R} \setminus B_0 = B_1$, setze $\varphi(X_1, \dots, X_n) = 1$. H_0 wird verworfen. Ansonsten setze $\varphi(X_1, \dots, X_n) = 0$.
- Falls die Verteilung von T nur asymptotisch bestimmt werden kann, so heißt φ *asymptotischer Test*.
- Sehr oft aber ist auch die asymptotische Verteilung von T nicht bekannt. Dann verwendet man sogenannte *Monte-Carlo Tests*, in denen dann Quantile t_α näherungsweise aus sehr vielen Monte-Carlo-Simulationen von T (unter H_0) bestimmt werden: Falls t^i , $i = 1, \dots, m$ die Werte von T in m unabhängigen Simulationsvorgängen sind, das heißt $t^i = T(x_1^i, \dots, x_n^i)$, x_j^i sind unabhängige Realisierungen von $X_j \sim F \in \Lambda_0$, $j = 1, \dots, n$, $i = 1, \dots, m$ dann bildet man ihre Ordnungsstatistiken $t^{(1)}, \dots, t^{(m)}$ und setzt $t_\alpha \approx t^{(\lfloor \alpha \cdot m \rfloor)}$, $\alpha \in [0, 1]$, wobei $t^{(0)} = -\infty$.

Bemerkung 1.1.2. Man sieht deutlich, daß aus einem beliebigen Konfidenzintervall

$$I_\theta = \left[I_1^\theta(X_1, \dots, X_n), I_2^\theta(X_1, \dots, X_n) \right]$$

zum Niveau $1 - \alpha$ für einen Parameter $\theta \in \mathbb{R}$ ein Test für θ konstruierbar ist. Die Hypothese $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ wird mit folgender Entscheidungsregel getestet:

$$\varphi(X_1, \dots, X_n) = 1, \text{ falls } \theta_0 \notin \left[I_1^{\theta_0}(X_1, \dots, X_n), I_2^{\theta_0}(X_1, \dots, X_n) \right].$$

Das Signifikanzniveau des Tests ist α .

Beispiel 1.1.2. *Normalverteilung, Test des Erwartungswertes bei bekannter Varianz.* Es seien

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

mit bekannter Varianz σ^2 . Ein Konfidenzintervall für μ ist

$$I^\mu = \left[I_1^\mu(X_1, \dots, X_n), I_2^\mu(X_1, \dots, X_n) \right] = \left[\bar{X}_n - \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}} \right]$$

(vergleiche Stochastik I, 4.2.1) H_0 wird verworfen, falls $|\mu_0 - \bar{X}_n| > \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}$. In der Testsprache bedeutet es, dass

$$\varphi(x_1, \dots, x_n) = \mathbb{I}((x_1, \dots, x_n) \in K_1),$$

wobei

$$K_1 = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : |\mu_0 - \bar{x}_n| > \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right\}$$

der Ablehnungsbereich ist. Für die Teststatistik $T(X_1, \dots, X_n)$ gilt:

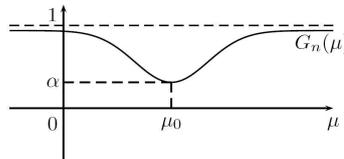
$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n} \sim N(0, 1) \mid \text{unter } H_0,$$

$$\alpha_n(\mu) = \alpha.$$

Berechnen wir nun die Gütefunktion (vergleiche Abbildung 1.3).

$$\begin{aligned} G_n(\mu) &= \mathbb{P}_\mu \left(|\mu_0 - \bar{X}_n| > \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) = 1 - \mathbb{P}_\mu \left(|\bar{X}_n - \mu_0| \leq \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right) \\ &= 1 - \mathbb{P}_\mu \left(\left| \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} + \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right| \leq z_{1-\alpha/2} \right) \\ &= 1 - \mathbb{P}_\mu \left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right) \\ &= 1 - \Phi \left(z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right) + \Phi \left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right) \\ &= \Phi \left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right) + \Phi \left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma} \sqrt{n} \right). \end{aligned}$$

Abbildung 1.3: Gütefunktion für den zweiseitigen Test des Erwartungswertes einer Normalverteilung bei bekannter Varianz



Die „Ja-Nein“-Entscheidung des Testens wird oft als zu grob empfunden. Deswegen versucht man, ein feineres Maß der Verträglichkeit der Daten mit den Hypothesen H_0 und H_1 zu bestimmen. Dies ist der sogenannte p -Wert, der von den meisten Statistik-Softwarepaketen ausgegeben wird.

Definition 1.1.4. Es sei (x_1, \dots, x_n) die konkrete Stichprobe von Daten, die als Realisierung von (X_1, \dots, X_n) interpretiert wird und $T(X_1, \dots, X_n)$ die Teststatistik, mit deren Hilfe die Entscheidungsregel φ konstruiert wurde. Der p -Wert des statistischen Tests φ ist das kleinste Signifikanzniveau, zu dem der Wert $t = T(x_1, \dots, x_n)$ zur Verwerfung der Hypothese H_0 führt.

Im Beispiel eines einseitigen Tests mit dem Ablehnungsbereich $B_1 = (t, \infty)$ sagt man grob, daß

$$p = \text{„}\mathbb{P}(T(X_1, \dots, X_n) \geq t \mid H_0)\text{“},$$

wobei die Anführungszeichen bedeuten, daß dies keine klassische, sondern eine bedingte Wahrscheinlichkeit ist, die später präzise angegeben wird.

Bei der Verwendung des p -Wertes verändert sich die Ablehnungsregel: die Hypothese H_0 wird zum Signifikanzniveau α abgelehnt, falls $\alpha \geq p$. Früher hat man die Signifikanz der Testentscheidung (Ablehnung von H_0) an Hand folgender Tabelle festgesetzt:

p -Wert	Interpretation
$p \leq 0,001$	sehr stark signifikant
$0,001 < p \leq 0,01$	stark signifikant
$0,01 < p \leq 0,05$	schwach signifikant
$0,05 < p$	nicht signifikant

Da aber heute der p -Wert an sich verwendet werden kann, kann der Anwender der Tests bei vorgegebenem p -Wert selbst entscheiden, zu welchem Niveau er seine Tests durchführen will.

- Bemerkung 1.1.3.**
1. Das Signifikanzniveau darf nicht in Abhängigkeit von p festgelegt werden. Dies würde die allgemeine Testphilosophie zerstören!
 2. Der p -Wert ist keine Wahrscheinlichkeit, sondern eine Zufallsvariable, denn er hängt von (X_1, \dots, X_n) ab. Der Ausdruck $p = \mathbb{P}(T(X_1, \dots, X_n) \geq t \mid H_0)$, der in Definition 1.1.4 für den p -Wert eines einseitigen Tests mit Teststatistik T gegeben wurde, soll demnach als *Überschreitungswahrscheinlichkeit* interpretiert werden, daß bei Wiederholung des Zufallsexperiments unter H_0 der Wert $t = T(x_1, \dots, x_n)$ oder extremere Werte in Richtung der Hypothese H_1 betrachtet werden:

$$p = \mathbb{P}(T(X'_1, \dots, X'_n) \geq T(x_1, \dots, x_n) \mid H_0),$$

wobei $(X'_1, \dots, X'_n) \stackrel{d}{=} (X_1, \dots, X_n)$. Falls wir von einer konkreten Realisierung (x_1, \dots, x_n) zur Zufallsstichprobe (X_1, \dots, X_n) übergehen, erhalten wir

$$p = p(X_1, \dots, X_n) = \mathbb{P}(T(X'_1, \dots, X'_n) \geq T(X_1, \dots, X_n) \mid H_0)$$

3. Für andere Hypothesen H_0 wird der p -Wert auch eine andere Form haben. Zum Beispiel für

a) einen symmetrischen zweiseitigen Test ist

$$B_0 = [-t_{1-\alpha/2}, t_{1-\alpha/2}]$$

der Akzeptanzbereich für H_0 .

$$\Rightarrow p = P(|T(X'_1, \dots, X'_n)| \geq t \mid H_0), t = T(X_1, \dots, X_n)$$

b) einen linksseitigen Test mit $B_0 = [t_\alpha, \infty]$ gilt

$$p = P(T(X'_1, \dots, X'_n) \leq t \mid H_0), t = T(X_1, \dots, X_n)$$

c) Das Verhalten des p -Wertes kann folgendermaßen untersucht werden:

Lemma 1.1.1. Falls die Verteilungsfunktion F von X_i stetig und streng monoton steigend ist (die Verteilung von T ist absolut stetig mit zum Beispiel stetiger Dichte), dann ist $p \sim U[0, 1]$.

Beweis. Wir zeigen es am speziellen Beispiel des rechtsseitigen Tests.

$$\begin{aligned} \mathbb{P}(p \leq \alpha \mid H_0) &= \mathbb{P}(\bar{F}_T(T(X_1, \dots, X_n)) \leq \alpha \mid H_0) \\ &= \mathbb{P}(F_T(T(X_1, \dots, X_n)) \geq 1 - \alpha \mid H_0) \\ &= \mathbb{P}(U \geq 1 - \alpha) = 1 - (1 - \alpha) = \alpha, \quad \alpha \in [0, 1], \end{aligned}$$

da $F_T(T(X_1, \dots, X_n)) \stackrel{d}{=} U \sim U[0, 1]$ und F_T absolut stetig ist. □

Übung 1.1.1. Zeigen Sie, daß für eine beliebige Zufallsvariable X mit absolut stetiger Verteilung und streng monoton steigender Verteilungsfunktion F_X gilt:

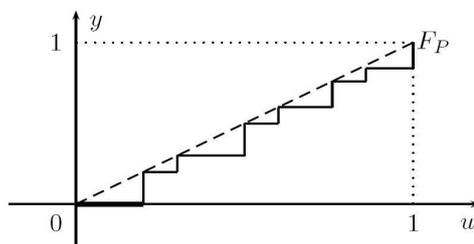
$$F_X(X) \sim U[0, 1]$$

Falls die Verteilung von T diskret ist, mit dem Wertebereich $\{t_1, \dots, t_n\}$, $t_i < t_j$ für $i < j$, so ist auch die Verteilung von p diskret, somit gilt nicht $p \sim U[0, 1]$. In diesem Fall ist $F_T(x)$ eine Treppenfunktion, die die Gerade $y = u$ in den Punkten $u = \sum_{i=1}^k \mathbb{P}(T(X_1, \dots, X_n) = t_i)$, $k = 1 \dots n$ berührt (vgl. Abbildung 1.4).

Definition 1.1.5. 1. Falls die Macht $G_n(\cdot)$ eines Tests φ zum Niveau α die Ungleichung

$$G_n(F) \geq \alpha, \quad F \in \Lambda_1$$

erfüllt, dann heißt der Test *unverfälscht*.

Abbildung 1.4: Verteilung von p für diskrete T 

2. Es seien φ und φ^* zwei Tests zum Niveau α mit Gütefunktionen $G_n(\cdot)$ und $G_n^*(\cdot)$. Man sagt, daß der Test φ *besser* als φ^* ist, falls er eine größere Macht besitzt:

$$G_n(F) \geq G_n^*(F) \quad \forall F \in \Lambda_1$$

3. Der Test φ heißt konsistent, falls $G_n(F) \xrightarrow{n \rightarrow \infty} 1$ für alle $F \in \Lambda_1$.

Bemerkung 1.1.4. 1. Die einseitigen Tests haben oft eine größere Macht als ihre zweiseitigen Versionen.

Beispiel 1.1.3. Betrachten wir zum Beispiel den Gauß-Test des Erwartungswertes der Normalverteilung bei bekannter Varianz. Beim zweiseitigen Test

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0.$$

erhalten wir die Gütefunktion

$$G_n(\mu) = \Phi \left(-z_{1-\alpha/2} + \sqrt{n} \frac{\mu - \mu_0}{\sigma} \right) + \Phi \left(-z_{1-\alpha/2} - \sqrt{n} \frac{\mu - \mu_0}{\sigma} \right).$$

Beim einseitigen Test φ^* der Hypothesen

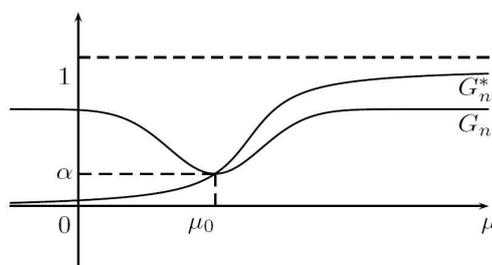
$$H_0^* : \mu \leq \mu_0 \text{ vs. } H_1^* : \mu > \mu_0$$

ist seine Gütefunktion gleich

$$G_n^*(\mu) = \Phi \left(-z_{1-\alpha} + \sqrt{n} \frac{\mu - \mu_0}{\sigma} \right)$$

Beide Tests sind offensichtlich konsistent, denn $G_n(\mu) \xrightarrow{n \rightarrow \infty} 1$, $G_n^*(\mu) \xrightarrow{n \rightarrow \infty} 1$. Dabei ist φ^* besser als φ . Beide Tests sind unverfälscht (vergleiche Abbildung 1.5).

Abbildung 1.5: Gütefunktionen eines ein- bzw. zweiseitigen Tests der Erwartungswertes einer Normalverteilung



2. Beim Testen einer Intervallhypothese $H_0 : \theta \in [a, b]$ vs. $H_1 : \theta \notin [a, b]$ zum Niveau α kann man wie folgt vorgehen: Teste

a) $H_0^a : \theta \geq a$ vs. $H_1^a : \theta < a$ zum Niveau $\alpha/2$.

b) $H_0^b : \theta \leq b$ vs. $H_1^b : \theta > b$ zum Niveau $\alpha/2$.

H_0 wird nicht abgelehnt, falls H_0^a und H_0^b nicht abgelehnt werden. Die Wahrscheinlichkeit des Fehlers 1. Art ist hier α . Die Macht dieses Tests ist im Allgemeinen schlecht.

3. Je mehr Parameter für den Aufbau der Teststatistik T geschätzt werden müssen, desto kleiner wird in der Regel die Macht.

1.2 Nichtrandomisierte Tests

1.2.1 Parametrische Signifikanztests

In diesem Abschnitt geben wir Beispiele einiger Tests, die meistens aus den entsprechenden Konfidenzintervallen für die Parameter von Verteilungen entstehen. Deshalb werden wir sie nur kurz behandeln.

1. Tests für die Parameter der Normalverteilung $N(\mu, \sigma^2)$

a) Test von μ bei unbekannter Varianz

- Hypothesen: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.

- Teststatistik:

$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{S_n} \sim t_{n-1} \quad | H_0$$

- Entscheidungsregel:

$$\varphi(X_1, \dots, X_n) = 1, \text{ falls } |T(X_1, \dots, X_n)| > t_{n-1, 1-\alpha/2}.$$

b) Test von σ^2 bei bekanntem μ

- Hypothesen: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$.
- Teststatistik:

$$T(X_1, \dots, X_n) = \frac{n\tilde{S}_n^2}{\sigma_0^2} \sim \chi_n^2 \quad | H_0$$

$$\text{mit } \tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

- Entscheidungsregel:

$$\varphi(X_1, \dots, X_n) = 1, \text{ falls } T(X_1, \dots, X_n) \notin [\chi_{n,\alpha/2}^2, \chi_{n,1-\alpha/2}^2].$$

- Gütefunktion:

$$\begin{aligned} G_n(\sigma^2) &= 1 - \mathbb{P}_{\sigma^2} \left(\chi_{n,\alpha/2}^2 \leq \frac{n\tilde{S}_n^2}{\sigma_0^2} \leq \chi_{n,1-\alpha/2}^2 \right) \\ &= 1 - \mathbb{P}_{\sigma^2} \left(\frac{\chi_{n,\alpha/2}^2 \sigma_0^2}{\sigma^2} \leq \frac{n\tilde{S}_n^2}{\sigma^2} \leq \frac{\chi_{n,1-\alpha/2}^2 \sigma_0^2}{\sigma^2} \right) \\ &= 1 - F_{\chi_n^2} \left(\chi_{n,1-\alpha/2}^2 \frac{\sigma_0^2}{\sigma^2} \right) + F_{\chi_n^2} \left(\chi_{n,\alpha/2}^2 \frac{\sigma_0^2}{\sigma^2} \right) \end{aligned}$$

c) Test von σ^2 bei unbekanntem μ

- Hypothesen: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$.
- Teststatistik:

$$T(X_1, \dots, X_n) = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi_{n-1}^2 \quad | H_0,$$

$$\text{wobei } S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

- Entscheidungsregel:

$$\varphi(X_1, \dots, X_n) = 1, \text{ falls } T(X_1, \dots, X_n) \notin [\chi_{n-1,\alpha/2}^2, \chi_{n-1,1-\alpha/2}^2].$$

Übung 1.2.1. (i) Finden Sie $G_n(\cdot)$ für die einseitige Version der obigen Tests.

(ii) Zeigen Sie, daß diese einseitigen Tests unverfälscht sind, die zweiseitigen aber nicht.

2. Asymptotische Tests

Bei asymptotischen Tests ist die Verteilung der Teststatistik nur näherungsweise (für große n) bekannt. Ebenso asymptotisch wird das Konfidenzniveau α erreicht. Ihre Konstruktion basiert meistens auf Verwendung der Grenzwertsätze.

Die allgemeine Vorgehensweise wird im sogenannten *Wald-Test* (genannt nach dem Statistiker Abraham Wald (1902-1980)) fixiert:

- Sei (X_1, \dots, X_n) eine Zufallsstichprobe, X_i seien unabhängig und identisch verteilt für $i = 1, \dots, n$, mit $X_i \sim F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$.
- Wir testen $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Es sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ ein erwartungstreuer, asymptotisch normalverteilter Schätzer für θ .

$$\frac{\hat{\theta}_n - \theta_0}{\hat{\sigma}_n} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1) \quad | H_0,$$

wobei $\hat{\sigma}_n^2$ ein konsistenter Schätzer für die Varianz von $\hat{\theta}_n$ sei.

Die Teststatistik ist

$$T(X_1, \dots, X_n) = \frac{\hat{\theta}_n(X_1, \dots, X_n) - \theta_0}{\hat{\sigma}_n}.$$

- Die Entscheidungsregel lautet: H_0 wird abgelehnt, wenn $|T(X_1, \dots, X_n)| > z_{1-\alpha/2}$, wobei $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. Diese Entscheidungsregel soll nur bei großen n verwendet werden. Die Wahrscheinlichkeit des Fehlers 1. Art ist asymptotisch gleich α , denn $\mathbb{P}(|T(X_1, \dots, X_n)| > z_{1-\alpha/2} | H_0) \xrightarrow[n \rightarrow \infty]{} \alpha$ wegen der asymptotischen Normalverteilung von T .

Die Gütefunktion des Tests ist asymptotisch gleich

$$\lim_{n \rightarrow \infty} G_n(\theta) = 1 - \Phi\left(z_{1-\alpha/2} + \frac{\theta_0 - \theta}{\sigma}\right) + \Phi\left(-z_{1-\alpha/2} + \frac{\theta_0 - \theta}{\sigma}\right),$$

wobei $\hat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma^2$.

Spezialfälle des Wald-Tests sind asymptotische Tests der Erwartungswerte bei einer Poisson- oder Bernoulliverteilten Stichprobe.

Beispiel 1.2.1. a) Bernoulliverteilung

Es seien $X_i \sim \text{Bernoulli}(p)$, $p \in [0, 1]$ unabhängige, identisch verteilte Zufallsvariablen.

- Hypothesen: $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$.
- Teststatistik:

$$T(X_1, \dots, X_n) = \begin{cases} \sqrt{n} \frac{\bar{X}_n - p_0}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}, & \text{falls } \bar{X}_n \neq 0, 1, \\ 0, & \text{sonst.} \end{cases}$$

Unter H_0 gilt: $T(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1)$.

b) Poissonverteilung

Es seien $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$ unabhängige, identisch verteilte Zufallsvariablen.

- Hypothesen: $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$
- Teststatistik:

$$T(X_1, \dots, X_n) = \begin{cases} \sqrt{n} \frac{\bar{X}_n - \lambda_0}{\sqrt{\bar{X}_n}}, & \text{falls } \bar{X}_n > 0, \\ 0, & \text{sonst.} \end{cases}$$

Unter H_0 gilt: $T(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1)$

3. Zwei-Stichproben-Probleme

Gegeben seien zwei Zufallsstichproben

$$Y_1 = (X_{11}, \dots, X_{1n_1}), Y_2 = (X_{21}, \dots, X_{2n_2}), n = \max\{n_1, n_2\}.$$

X_{ij} seien unabhängig für $j = 1, \dots, n_i$, $X_{ij} \sim F_{\theta_i}$, $i = 1, 2$.

a) Test der Gleichheit zweier Erwartungswerte bei normalverteilten Stichproben

- bei bekannten Varianzen

Es seien $X_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, $j = 1, \dots, n$. Dabei seien σ_1^2, σ_2^2 bekannt, X_{ij} seien unabhängig voneinander für alle i, j .

Die Hypothesen sind $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Wir betrachten die Teststatistik:

$$T(Y_1, Y_2) = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Unter H_0 gilt: $T(Y_1, Y_2) \sim N(0, 1)$. Als Entscheidungsregel gilt: H_0 wird abgelehnt, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$.

- **bei unbekanntem (jedoch gleichen) Varianzen**

Es seien $X_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, $j = 1, \dots, n$. Dabei seien σ_1^2, σ_2^2 unbekannt, $\sigma_1^2 = \sigma_2^2$ und X_{ij} seien unabhängig voneinander für alle i, j .

Die Hypothesen sind: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Wir betrachten die Teststatistik

$$T(Y_1, Y_2) = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2}}{S_{n_1 n_2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

wobei

$$S_{n_1 n_2}^2 = \frac{1}{n_1 + n_2 - 2} \cdot \left(\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_{1n_1})^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_{2n_2})^2 \right).$$

Man kann zeigen, daß unter H_0 gilt: $T(Y_1, Y_2) \sim t_{n_1+n_2-2}$. Die Entscheidungsregel lautet: H_0 ablehnen, falls $|T(Y_1, Y_2)| > t_{n_1+n_2-2, 1-\alpha/2}$.

b) **Test der Gleichheit von Erwartungswerten bei verbundenen Stichproben**

Es seien $Y_1 = (X_{11}, \dots, X_{1n})$ und $Y_2 = (X_{21}, \dots, X_{2n})$, $n_1 = n_2 = n$,

$$Z_j = X_{1j} - X_{2j} \sim N(\mu_1 - \mu_2, \sigma^2), \quad j = 1, \dots, n$$

unabhängig und identisch verteilt mit $\mu_i = \mathbb{E} X_{ij}$, $i = 1, 2$. Die Hypothesen sind: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$ bei unbekannter Varianz σ^2 . Als Teststatistik verwenden wir

$$T(Z_1, \dots, Z_n) = \sqrt{n} \frac{\bar{Z}_n}{S_n},$$

wobei

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (Z_j - \bar{Z}_n)^2.$$

Unter H_0 gilt dann: $T(Z_1, \dots, Z_n) \sim t_{n-1}$. Die Entscheidungsregel lautet: H_0 wird abgelehnt, falls $|T(z_1, \dots, z_n)| > t_{n-1, 1-\alpha/2}$.

c) **Test der Gleichheit von Varianzen bei unabhängigen Gaußschen Stichproben**

Es seien $Y_1 = (X_{11}, \dots, X_{1n_1})$ und $Y_2 = (X_{21}, \dots, X_{2n_2})$ unabhängig und identisch verteilt mit $X_{ij} \sim N(\mu_i, \sigma_i^2)$, wobei μ_i und σ_i^2 beide unbekannt sind. Die Hypothesen sind: $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_1 : \sigma_1^2 \neq \sigma_2^2$. Als Teststatistik verwenden wir

$$T(Y_1, Y_2) = \frac{S_{2n_2}^2}{S_{1n_1}^2},$$

wobei

$$S_{in_i}^2 = \frac{1}{n_i - 1} \sum_{j=1}^n (X_{ij} - \bar{X}_{in_i})^2, \quad i = 1, 2.$$

Unter H_0 gilt: $T(Y_1, Y_2) \sim F_{n_2-1, n_1-1}$. Die Entscheidungsregel lautet: H_0 wird abgelehnt, falls $T(Y_1, Y_2) \notin [F_{n_2-1, n_1-1, \alpha/2}, F_{n_2-1, n_1-1, 1-\alpha/2}]$.

d) Asymptotische Zwei-Stichproben-Tests

- bei Bernoulli-verteilten Stichproben

Es gilt $X_{ij} \sim \text{Bernoulli}(p_i)$, $j = 1, \dots, n_i$, $i = 1, 2$. Die Hypothesen sind $H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$. Als Teststatistik verwenden wir

$$T(Y_1, Y_2) = \frac{(\bar{X}_{1n_1} - \bar{X}_{2n_2})(1 - \mathbb{I}(\bar{X}_{1n_1} = 0, \bar{X}_{2n_2} = 0))}{\sqrt{\frac{\bar{X}_{1n_1}(1-\bar{X}_{1n_1})}{n_1} + \frac{\bar{X}_{2n_2}(1-\bar{X}_{2n_2})}{n_2}}}$$

Unter H_0 gilt: $T(Y_1, Y_2) \xrightarrow[n_1, n_2 \rightarrow \infty]{d} Y \sim N(0, 1)$. Die Entscheidungsregel lautet: H_0 wird verworfen, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$. Dies ist ein Test zum asymptotischen Signifikanzniveau α .

- bei Poisson-verteilten Stichproben

Es seien X_{ij} unabhängig, $X_{ij} \sim \text{Poisson}(\lambda_i)$, $i = 1, 2$. Die Hypothesen sind: $H_0 : \lambda_1 = \lambda_2$ vs. $H_1 : \lambda_1 \neq \lambda_2$. Als Teststatistik verwenden wir:

$$T(Y_1, Y_2) = \frac{\bar{X}_{1n_1} - \bar{X}_{2n_2}}{\sqrt{\frac{\bar{X}_{1n_1}}{n_1} + \frac{\bar{X}_{2n_2}}{n_2}}}$$

Die Entscheidungsregel lautet: H_0 ablehnen, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$. Dies ist ein Test zum asymptotischen Niveau α .

Bemerkung 1.2.1. Asymptotische Tests dürfen nur für große Stichprobenumfänge verwendet werden. Bei ihrer Verwendung für kleine Stichproben kann das asymptotische Signifikanzniveau nicht garantiert werden.

1.3 Randomisierte Tests

In diesem Abschnitt werden wir klassische Ergebnisse von Neyman-Pearson über die besten Tests präsentieren. Dabei werden randomisierte Tests eine wichtige Rolle spielen.

1.3.1 Grundlagen

Gegeben sei eine Zufallsstichprobe (X_1, \dots, X_n) von unabhängigen und identisch verteilten Zufallsvariablen X_i mit konkreter Ausprägung (x_1, \dots, x_n) . Sei unser Stichprobenraum (B, \mathcal{B}) entweder $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ oder $(\mathbb{N}_0^n, \mathcal{B}_{\mathbb{N}_0^n})$, je nachdem, ob die Stichprobenvariablen X_i , $i = 1, \dots, n$ absolut stetig oder diskret verteilt sind.

Hier wird zur Einfachheit im Falle einer diskret verteilten Zufallsvariable X_i ihr diskreter Wertebereich mit $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ gleichgesetzt. Der Wertebereich sei mit einem Maß μ versehen, wobei

$$\mu = \begin{cases} \text{Lebesgue-Ma\ss auf } \mathbb{R}, & \text{falls } B = \mathbb{R}^n, \\ \text{Z\ahlnma\ss auf } \mathbb{N}_0, & \text{falls } B = \mathbb{N}_0^n. \end{cases}$$

Dementsprechend gilt

$$\int g(x)\mu(dx) = \begin{cases} \int_{\mathbb{R}} g(x)dx, & \text{im absolut stetigen Fall,} \\ \sum_{x \in \mathbb{N}_0} g(x), & \text{im diskreten Fall.} \end{cases}$$

Es sei zus\atztzlich $X_i \sim F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^m$, $i = 1, \dots, n$ (parametrisches Modell). F\ur $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$ formulieren wir die Hypothesen $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, die mit Hilfe eines randomisierten Tests

$$\varphi(x) = \begin{cases} 1, & x \in K_1, \\ \gamma \in (0, 1), & x \in K_{01} \\ 0, & x \in K_0 \end{cases} \quad x = (x_1, \dots, x_n),$$

getestet werden.

Im Falle $x \in K_{01}$ wird mit Hilfe einer Zufallsvariable $Y \sim \text{Bernoulli}(\varphi(x))$ entschieden, ob H_0 verworfen wird ($Y = 1$) oder nicht ($Y = 0$).

Definition 1.3.1. 1. Die *G\utefunktion* eines randomisierten Tests φ sei

$$G_n(\theta) = G_n(\varphi, \theta) = \mathbb{E}_\theta \varphi(X_1, \dots, X_n), \theta \in \Theta.$$

2. Der Test φ hat das *Signifikanzniveau* $\alpha \in [0, 1]$, falls $G_n(\varphi, \theta) \leq \alpha$, $\forall \theta \in \Theta_0$ ist. Die Zahl

$$\sup_{\theta \in \Theta_0} G_n(\varphi, \theta)$$

wird *Umfang* des Tests φ genannt. Offensichtlich ist der Umfang eines Niveau- α -Tests kleiner gleich α .

3. Sei $\Psi(\alpha)$ die Menge aller Tests zum Niveau α . Der Test $\varphi_1 \in \Psi(\alpha)$ ist (*gleichm\afsig*) *besser* als Test $\varphi_2 \in \Psi(\alpha)$, falls $G_n(\varphi_1, \theta) \geq G_n(\varphi_2, \theta)$, $\theta \in \Theta_1$, also falls φ_1 eine gr\o\ssere Macht besitzt.
4. Ein Test $\varphi^* \in \Psi(\alpha)$ ist (*gleichm\afsig*) *bester Test* in $\Psi(\alpha)$, falls

$$G_n(\varphi^*, \theta) \geq G_n(\varphi, \theta), \text{ f\ur alle Tests } \varphi \in \Psi(\alpha), \theta \in \Theta_1.$$

Bemerkung 1.3.1. 1. Definition 1.3.1 1) ist eine offensichtliche Verallgemeinerung der Definition 1.1.3 der Gütefunktion eines nicht-randomisierten Tests φ . Nämlich, für $\varphi(x) = \mathbb{I}(x \in K_1)$ gilt:

$$\begin{aligned} G_n(\varphi, \theta) &= \mathbb{E}_\theta \varphi(X_1, \dots, X_n) \\ &= \mathbb{P}_\theta((X_1, \dots, X_n) \in K_1) \\ &= \mathbb{P}_\theta(H_0 \text{ ablehnen}), \theta \in \Theta. \end{aligned}$$

2. Ein bester Test φ^* in $\Psi(\alpha)$ existiert nicht immer, sondern nur unter gewissen Voraussetzungen an $\mathbb{P}_\theta, \Theta_0, \Theta_1$ und $\Psi(\alpha)$.

1.3.2 Neyman-Pearson-Tests bei einfachen Hypothesen

In diesem Abschnitt betrachten wir einfache Hypothesen

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \quad (1.3.1)$$

wobei $\theta_0, \theta_1 \in \Theta, \theta_1 \neq \theta_0$.

Dementsprechend sind $\Theta_0 = \{\theta_0\}, \Theta_1 = \{\theta_1\}$. Wir setzen voraus, daß F_{θ_i} eine Dichte $g_i(x)$ bezüglich μ besitzt, $i = 0, 1$. Führen wir einige abkürzende Bezeichnungen $\mathbb{P}_0 = \mathbb{P}_{\theta_0}, \mathbb{P}_1 = \mathbb{P}_{\theta_1}, \mathbb{E}_0 = \mathbb{E}_{\theta_0}, \mathbb{E}_1 = \mathbb{E}_{\theta_1}$ ein. Sei $f_i(x) = \prod_{j=1}^n g_i(x_j), x = (x_1, \dots, x_n), i = 0, 1$ die Dichte der Stichprobe unter H_0 bzw. H_1 .

Definition 1.3.2. Ein *Neyman-Pearson-Test (NP-Test)* der einfachen Hypothesen in (1.3.1) ist gegeben durch die Regel

$$\varphi(x) = \varphi_K(x) = \begin{cases} 1, & \text{falls } f_1(x) > K f_0(x), \\ \gamma, & \text{falls } f_1(x) = K f_0(x), \\ 0, & \text{falls } f_1(x) < K f_0(x) \end{cases} \quad (1.3.2)$$

für Konstanten $K > 0$ und $\gamma \in [0, 1]$.

Bemerkung 1.3.2. 1. Manchmal werden $K = K(x)$ und $\gamma = \gamma(x)$ als Funktionen von x und nicht als Konstanten betrachtet.

2. Der *Ablehnungsbereich* des Neyman-Pearson-Tests φ_K ist

$$K_1 = \{x \in B : f_1(x) > K f_0(x)\}.$$

3. Der *Umfang* des Neyman-Pearson-Tests φ_K ist

$$\begin{aligned} \mathbb{E}_0 \varphi_K(X_1, \dots, X_n) &= \mathbb{P}_0(f_1(X_1, \dots, X_n) > K f_0(X_1, \dots, X_n)) \\ &\quad + \gamma \mathbb{P}_0(f_1(X_1, \dots, X_n) = K f_0(X_1, \dots, X_n)) \end{aligned}$$

4. Die Definition 1.3.2 kann man äquivalent folgendermaßen geben: Wir definieren eine Teststatistik

$$T(x) = \begin{cases} \frac{f_1(x)}{f_0(x)}, & x \in B : f_0(x) > 0, \\ \infty, & x \in B : f_0(x) = 0. \end{cases}$$

Dann wird der neue Test

$$\tilde{\varphi}_K(x) = \begin{cases} 1, & \text{falls } T(x) > K, \\ \gamma, & \text{falls } T(x) = K, \\ 0, & \text{falls } T(x) < K \end{cases}$$

eingeführt, der für P_0 - und P_1 - fast alle $x \in B$ äquivalent zu φ_K ist. In der Tat gilt $\varphi_K(x) = \tilde{\varphi}_K(x) \forall x \in B \setminus C$, wobei $C = \{x \in B : f_0(x) = f_1(x) = 0\}$ das \mathbb{P}_0 - bzw. \mathbb{P}_1 -Maß Null besitzt.

In der neuen Formulierung ist der Umfang von φ bzw. $\tilde{\varphi}_K$ gleich

$$\mathbb{E}_0 \tilde{\varphi}_K = \mathbb{P}_0(T(X_1, \dots, X_n) > K) + \gamma \cdot \mathbb{P}_0(T(X_1, \dots, X_n) = K).$$

Satz 1.3.1. Optimalitätssatz

Es sei φ_K ein Neyman-Pearson-Test für ein $K > 0$ und $\gamma \in [0, 1]$. Dann ist φ_K der beste Test zum Niveau $\alpha = \mathbb{E}_0 \varphi_K$ seines Umfangs.

Beweis. Sei $\varphi \in \Psi(\alpha)$, also $\mathbb{E}_0(\varphi(X_1, \dots, X_n)) \leq \alpha$. Um zu zeigen, daß φ_K besser als φ ist, genügt es bei einfachen Hypothesen H_0 und H_1 zu zeigen, daß $\mathbb{E}_1 \varphi_K(X_1, \dots, X_n) \geq \mathbb{E}_1 \varphi(X_1, \dots, X_n)$. Wir führen dazu die folgenden Mengen ein:

$$M^+ = \{x \in B : \varphi_K(x) > \varphi(x)\}$$

$$M^- = \{x \in B : \varphi_K(x) < \varphi(x)\}$$

$$M^= = \{x \in B : \varphi_K(x) = \varphi(x)\}$$

Es gilt offensichtlich $x \in M^+ \Rightarrow \varphi_K(x) > 0 \Rightarrow f_1(x) \geq K f_0(x)$,

$$x \in M^- \Rightarrow \varphi_K(x) < 1 \Rightarrow f_1(x) \leq K f_0(x) \text{ und } B = M^+ \cup M^- \cup M^=.$$

Als Folgerung erhalten wir

$$\begin{aligned} \mathbb{E}_1(\varphi_K(X_1, \dots, X_n) - \varphi(X_1, \dots, X_n)) &= \int_B (\varphi_K(x) - \varphi(x)) f_1(x) \mu(dx) \\ &= \left(\int_{M^+} + \int_{M^-} + \int_{M^=} \right) (\varphi_K(x) - \varphi(x)) f_1(x) \mu(dx) \\ &\geq \int_{M^+} (\varphi_K(x) - \varphi(x)) K f_0(x) \mu(dx) \\ &\quad + \int_{M^-} (\varphi_K(x) - \varphi(x)) K f_0(x) \mu(dx) \\ &= \int_B (\varphi_K(x) - \varphi(x)) K f_0(x) \mu(dx) \\ &= K [\mathbb{E}_0 \varphi_K(X_1, \dots, X_n) - \mathbb{E}_0 \varphi(X_1, \dots, X_n)] \\ &\geq K(\alpha - \alpha) = 0, \end{aligned}$$

weil beide Tests das Niveau α haben. Damit ist die Behauptung bewiesen. \square

Bemerkung 1.3.3. 1. Da im Beweis γ nicht vorkommt, wird derselbe Beweis im Falle von $\gamma(x) \neq \text{const}$ gelten.

2. Aus dem Beweis folgt die Gültigkeit der Ungleichung

$$\int_B (\varphi_K(x) - \varphi(x)) (f_1(x) - K f_0(x)) \mu(dx) \geq 0$$

im Falle des konstanten K , bzw.

$$\mathbb{E}_1 (\varphi_K(X_1, \dots, X_n) - \varphi(X_1, \dots, X_n)) \geq \int_B (\varphi_K(x) - \varphi(x)) K f_0(x) \mu(dx)$$

im allgemeinen Fall.

Satz 1.3.2. (Fundamentallemma von Neyman-Pearson)

1. Zu einem beliebigen $\alpha \in (0, 1)$ gibt es einen Neyman-Pearson-Test φ_K mit Umfang α , der dann nach Satz 1.3.1 der beste Niveau- α -Test ist.
2. Ist φ ebenfalls bester Test zum Niveau α , so gilt $\varphi(x) = \varphi_K(x)$ für μ -fast alle $x \in K_0 \cup K_1 = \{x \in B : f_1(x) \neq K f_0(x)\}$ und φ_K aus Teil 1).

Beweis. 1. Für $\varphi_K(x)$ gilt

$$\varphi_K(x) = \begin{cases} 1, & \text{falls } x \in K_1 = \{x : f_1(x) > K \cdot f_0(x)\}, \\ \gamma, & \text{falls } x \in K_{01} = \{x : f_1(x) = K \cdot f_0(x)\}, \\ 0, & \text{falls } x \in K_0 = \{x : f_1(x) < K \cdot f_0(x)\}. \end{cases}$$

Der Umfang von φ_K ist

$$\mathbb{P}_0(T(X_1, \dots, X_n) > K) + \gamma \mathbb{P}_0(T(X_1, \dots, X_n) = K) = \alpha, \quad (1.3.3)$$

wobei

$$T(x_1, \dots, x_n) = \begin{cases} \frac{f_1(x_1, \dots, x_n)}{f_0(x_1, \dots, x_n)}, & \text{falls } f_0(x_1, \dots, x_n) > 0, \\ \infty, & \text{sonst.} \end{cases}$$

Nun suchen wir ein $K > 0$ und ein $\gamma \in [0, 1]$, sodaß Gleichung (1.3.3) stimmt. Es sei $\tilde{F}_0(x) = \mathbb{P}_0(T(X_1, \dots, X_n) \leq x)$, $x \in \mathbb{R}$ die Verteilungsfunktion von T . Da $T \geq 0$ ist, gilt $\tilde{F}_0(x) = 0$, falls $x < 0$. Außerdem ist $\mathbb{P}_0(T(X_1, \dots, X_n) < \infty) = 1$, das heißt $\tilde{F}_0^{-1}(\alpha) \in [0, \infty)$, $\alpha \in (0, 1)$. Die Gleichung (1.3.3) kann dann folgendermaßen umgeschrieben werden:

$$1 - \tilde{F}_0(K) + \gamma (\tilde{F}_0(K) - \tilde{F}_0(K-)) = \alpha, \quad (1.3.4)$$

wobei $\tilde{F}_0(K-) = \lim_{x \rightarrow K-0} \tilde{F}_0(x)$.

Sei $K = \tilde{F}_0^{-1}(1 - \alpha)$, dann gilt:

- a) Falls K ein Stetigkeitspunkt von \tilde{F}_0 ist, ist Gleichung (1.3.4) erfüllt für alle $\gamma \in [0, 1]$, zum Beispiel $\gamma = 0$.
- b) Falls K kein Stetigkeitspunkt von \tilde{F}_0 ist, dann ist $\tilde{F}_0(K) - \tilde{F}_0(K-) > 0$, woraus folgt

$$\gamma = \frac{\alpha - 1 + \tilde{F}_0(K)}{\tilde{F}_0(K) - \tilde{F}_0(K-)}$$

\Rightarrow es gibt einen Neyman-Pearson-Test zum Niveau α .

2. Wir definieren $M^\neq = \{x \in B : \varphi(x) \neq \varphi_K(x)\}$. Es muss gezeigt werden, daß

$$\mu\left((K_0 \cup K_1) \cap M^\neq\right) = 0.$$

Dazu betrachten wir

$$\begin{aligned} \mathbb{E}_1 \varphi(X_1, \dots, X_n) - \mathbb{E}_1 \varphi_K(X_1, \dots, X_n) &= 0 && (\varphi \text{ und } \varphi_K \text{ sind beste Tests}) \\ \mathbb{E}_0 \varphi(X_1, \dots, X_n) - \mathbb{E}_0 \varphi_K(X_1, \dots, X_n) &\leq 0 && (\varphi \text{ und } \varphi_K \text{ sind } \alpha\text{-Tests} \\ &&& \text{mit Umfang von } \varphi_K = \alpha) \end{aligned}$$

$$\Rightarrow \int_B (\varphi - \varphi_K) \cdot (f_1 - K \cdot f_0) \mu(dx) \geq 0.$$

In Bemerkung 1.3.3 wurde bewiesen, daß

$$\begin{aligned} \int_B (\varphi - \varphi_K)(f_1 - K \cdot f_0) d\mu &\leq 0 \\ \Rightarrow \int_B (\varphi - \varphi_K)(f_1 - K \cdot f_0) d\mu &= 0 = \int_{M^\neq \cap (K_0 \cup K_1)} (\varphi - \varphi_K)(f_1 - K f_0) d\mu. \end{aligned}$$

Es gilt $\mu(M^\neq \cap (K_0 \cup K_1)) = 0$, falls der Integrand $(\varphi_K - \varphi)(f_1 - K f_0) > 0$ auf M^\neq ist. Wir zeigen, daß

$$(\varphi_K - \varphi)(f_1 - K f_0) > 0 \text{ für } x \in M^\neq \quad (1.3.5)$$

ist. Es gilt

$$\begin{aligned} f_1 - K f_0 > 0 &\Rightarrow \varphi_K - \varphi > 0, \\ f_1 - K f_0 < 0 &\Rightarrow \varphi_K - \varphi < 0, \end{aligned}$$

weil

$$\begin{aligned} f_1(x) > K f_0(x) &\Rightarrow \varphi_K(x) = 1 \\ &\text{und mit } \varphi(x) < 1 \Rightarrow \varphi_K(x) - \varphi(x) > 0 \text{ auf } M^\neq. \\ f_1(x) < K f_0(x) &\Rightarrow \varphi_K(x) = 0 \\ &\text{und mit } \varphi(x) > 0 \Rightarrow \varphi_K(x) - \varphi(x) < 0 \text{ auf } M^\neq. \end{aligned}$$

Daraus folgt die Gültigkeit der Ungleichung (1.3.5) und somit

$$\mu \left((K_0 \cup K_1) \cap M^\neq \right) = 0.$$

□

Bemerkung 1.3.4. Falls φ und φ_K beste α -Tests sind, dann sind sie P_0 - bzw. P_1 - fast sicher gleich.

Beispiel 1.3.1 (Neyman-Pearson-Test für den Parameter der Poissonverteilung). Es sei (X_1, \dots, X_n) eine Zufallsstichprobe mit $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$, wobei X_i unabhängig und identisch verteilt sind für $i = 1, \dots, n$. Wir testen die Hypothesen $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda = \lambda_1$. Dabei ist

$$g_i(x) = e^{-\lambda_i} \frac{\lambda_i^x}{x!}, \quad x \in \mathbb{N}_0, \quad i = 0, 1,$$

$$f_i(x) = f_i(x_1, \dots, x_n) = \prod_{j=1}^n g_i(x_j) = \prod_{j=1}^n e^{-\lambda_i} \frac{\lambda_i^{x_j}}{x_j!} = e^{-n\lambda_i} \cdot \frac{\lambda_i^{\sum_{j=1}^n x_j}}{(x_1! \cdot \dots \cdot x_n!)}$$

für $i = 0, 1$. Die Neyman-Pearson-Teststatistik ist

$$T(x_1, \dots, x_n) = \begin{cases} \frac{f_1(x)}{f_0(x)} = e^{-n(\lambda_1 - \lambda_0)} \cdot (\lambda_1/\lambda_0)^{\sum_{j=1}^n x_j}, & \text{falls } x_1, \dots, x_n \in \mathbb{N}_0, \\ \infty, & \text{sonst.} \end{cases}$$

Die Neyman-Pearson-Entscheidungsregel lautet

$$\varphi_K(x_1, \dots, x_n) = \begin{cases} 1, & \text{falls } T(x_1, \dots, x_n) > K, \\ \gamma, & \text{falls } T(x_1, \dots, x_n) = K, \\ 0, & \text{falls } T(x_1, \dots, x_n) < K. \end{cases}$$

Wir wählen $K > 0$, $\gamma \in [0, 1]$, sodaß φ_K den Umfang α hat. Dazu lösen wir

$$\alpha = \mathbb{P}_0(T(X_1, \dots, X_n) > K) + \gamma \mathbb{P}_0(T(X_1, \dots, X_n) = K)$$

bezüglich γ und K auf.

$$\begin{aligned} \mathbb{P}_0(T(X_1, \dots, X_n) > K) &= \mathbb{P}_0(\log T(X_1, \dots, X_n) > \log K) \\ &= \mathbb{P}_0 \left(-n(\lambda_1 - \lambda_0) + \sum_{j=1}^n X_j \cdot \log \left(\frac{\lambda_1}{\lambda_0} \right) > \log K \right) = \mathbb{P}_0 \left(\sum_{j=1}^n X_j > A \right) \end{aligned}$$

$$\text{wobei } A := \left\lceil \frac{\log K + n \cdot (\lambda_1 - \lambda_0)}{\log \frac{\lambda_1}{\lambda_0}} \right\rceil,$$

falls zum Beispiel $\lambda_1 > \lambda_0$. Im Falle $\lambda_1 < \lambda_0$ ändert sich das $>$ auf $<$ in der Wahrscheinlichkeit.

Wegen der Faltungsstabilität der Poissonverteilung ist unter H_0

$$\sum_{j=1}^n X_j \sim \text{Poisson}(n\lambda_0),$$

also wählen wir K als minimale, nichtnegative Zahl, für die gilt: $\mathbb{P}_0\left(\sum_{j=1}^n X_j > A\right) \leq \alpha$, und setzen

$$\gamma = \frac{\alpha - \mathbb{P}_0(\sum_{j=1}^n X_j > A)}{\mathbb{P}_0(\sum_{j=1}^n X_j = A)},$$

wobei

$$\begin{aligned} \mathbb{P}_0\left(\sum_{j=1}^n X_j > A\right) &= 1 - \sum_{j=0}^A e^{-\lambda_0 n} \frac{(\lambda_0 n)^j}{j!}, \\ \mathbb{P}_0\left(\sum_{j=1}^n X_j = A\right) &= e^{-\lambda_0 n} \frac{(\lambda_0 n)^A}{A!}. \end{aligned}$$

Somit haben wir die Parameter K und γ gefunden und damit einen Neyman-Pearson-Test φ_K konstruiert.

1.3.3 Einseitige Neyman-Pearson-Tests

Bisher betrachteten wir Neyman-Pearson-Tests für einfache Hypothesen der Form $H_i : \theta = \theta_i$, $i = 0, 1$. In diesem Abschnitt wollen wir einseitige Neyman-Pearson-Tests einführen, für Hypothesen der Form $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$.

Zunächst konstruieren wir einen Test für diese Hypothesen: Sei (X_1, \dots, X_n) eine Zufallsstichprobe, X_i seien unabhängig und identisch verteilt mit

$$X_i \sim F_\theta \in \Lambda = \{F_\theta : \theta \in \Theta\},$$

wobei $\Theta \subset \mathbb{R}$ offen ist und Λ eindeutig parametrisiert, das heißt

$$\theta \neq \theta' \Rightarrow F_\theta \neq F_{\theta'}.$$

Ferner besitze F_θ eine Dichte g_θ bezüglich des Lebesgue-Maßes (bzw. Zählmaßes) auf \mathbb{R} (bzw. \mathbb{N}_0). Dann ist

$$f_\theta(x) = \prod_{j=1}^n g_\theta(x_j), \quad x = (x_1, \dots, x_n)$$

eine Dichte von (X_1, \dots, X_n) bezüglich μ auf B .

Definition 1.3.3. Eine Verteilung auf B mit Dichte f_θ gehört zur Klasse von *Verteilungen mit monotonen Dichtekoeffizienten* in T , falls es für alle $\theta < \theta'$ eine Funktion $h : \mathbb{R} \times \Theta^2 \rightarrow \mathbb{R} \cup \infty$, die monoton wachsend in $t \in \mathbb{R}$ ist und eine Statistik $T : B \rightarrow \mathbb{R}$ gibt, mit der Eigenschaft

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = h(T(x), \theta, \theta'),$$

wobei

$$h(T(x), \theta, \theta') = \infty \quad \text{für alle } x \in B : f_\theta(x) = 0, f_{\theta'}(x) > 0.$$

Der Fall $f_\theta(x) = f_{\theta'}(x) = 0$ tritt mit \mathbb{P}_θ - bzw. $\mathbb{P}_{\theta'}$ -Wahrscheinlichkeit 0 auf.

Definition 1.3.4. Es sei Q_θ eine Verteilung auf (B, \mathcal{B}) mit der Dichte f_θ bzgl. μ . Q_θ gehört zur *einparametrischen Exponentialklasse* ($\theta \in \Theta \subset \mathbb{R}$ offen), falls die Dichte folgende Form hat:

$$f_\theta(x) = \exp \{c(\theta) \cdot T(x) + a(\theta)\} \cdot l(x), \quad x = (x_1, \dots, x_n) \in B,$$

wobei $c(\theta)$ eine monoton steigende Funktion ist, und $\text{Var}_\theta T(X_1, \dots, X_n) > 0$, $\theta \in \Theta$.

Lemma 1.3.1. Verteilungen aus der einparametrischen Exponentialfamilie besitzen einen monotonen Dichtekoeffizienten.

Beweis. Es sei Q_θ aus der einparametrischen Exponentialfamilie mit der Dichte

$$f_\theta(x) = \exp \{c(\theta) \cdot T(x) + a(\theta)\} \cdot l(x).$$

Für $\theta < \theta'$ ist dann

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = \exp \{(c(\theta') - c(\theta)) \cdot T(x) + a(\theta') - a(\theta)\}$$

monoton bezüglich T , weil $c(\theta') - c(\theta) > 0$ wegen der Monotonie von $c(\theta)$. Also besitzt f_θ einen monotonen Dichtekoeffizienten. \square

Beispiel 1.3.2. 1. *Normalverteilte Stichprobenvariablen*

Es seien $X_i \sim N(\mu, \sigma_0^2)$, $i = 1, \dots, n$, unabhängige, identisch verteilte Zufallsvariablen, mit unbekanntem Parameter μ und bekannter Varianz σ_0^2 (Hier wird μ für die Bezeichnung des Erwartungswertes von X_i und nicht des Maßes auf \mathbb{R}^n verwendet.

(wie früher)). Die Dichte des Zufallsvektors $X = (X_1, \dots, X_n)^\top$ ist gleich

$$\begin{aligned}
 f_\mu(x) &= \prod_{i=1}^n g_\mu(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma_0^2}} \\
 &= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \\
 &= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + \mu^2 n\right)\right\} \\
 &= \exp\left(\underbrace{\frac{\mu}{\sigma_0^2} \cdot \sum_{i=1}^n x_i}_{c(\mu)} - \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{a(\mu)}\right) \cdot \underbrace{\frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_0^2}\right)}_{l(x)}.
 \end{aligned}$$

Also gehört $N(\mu, \sigma_0^2)$ zur einparametrischen Exponentialklasse mit $c(\mu) = \frac{\mu}{\sigma_0^2}$ und $T(x) = \sum_{i=1}^n x_i$.

2. Binomialverteilte Stichprobenvariablen

Es seien $X_i \sim \text{Bin}(k, p)$ unabhängig und identisch verteilt, $i = 1, \dots, n$. Der Parameter p sei unbekannt. Die Zähldichte des Zufallsvektors $X = (X_1, \dots, X_n)^\top$ ist

$$\begin{aligned}
 f_p(x) &= \mathbb{P}_p(X_i = x_i, i = 1, \dots, n) \\
 &= \prod_{i=1}^n \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i} = p^{\sum_{i=1}^n x_i} \cdot \frac{(1-p)^{nk}}{(1-p)^{\sum_{i=1}^n x_i}} \cdot \prod_{i=1}^n \binom{k}{x_i} \\
 &= \exp\left\{\underbrace{\left(\sum_{i=1}^n x_i\right)}_{T(x)} \cdot \underbrace{\log\left(\frac{p}{1-p}\right)}_{c(p)} + \underbrace{nk \cdot \log(1-p)}_{a(p)}\right\} \cdot \underbrace{\prod_{i=1}^n \binom{k}{x_i}}_{l(x)},
 \end{aligned}$$

also gehört $\text{Bin}(n, p)$ zur einparametrischen Exponentialklasse mit

$$c(p) = \log\left(\frac{p}{1-p}\right)$$

und

$$T(x) = \sum_{i=1}^n x_i.$$

Lemma 1.3.2. Falls φ_K der Neyman-Pearson-Test der Hypothesen $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ ist, dann gilt:

$$\mu(\underbrace{\{x \in B : f_1(x) \neq K f_0(x)\}}_{K_0 \cup K_1}) > 0.$$

Beweis. Wegen $\theta_0 \neq \theta_1$ und der eindeutigen Parametrisierung gilt $f_0 \neq f_1$ auf einer Menge mit μ -Maß > 0 .

Nun sei $\mu(K_0 \cup K_1) = 0$. Daraus folgt, daß $f_1(x) = K \cdot f_0(x)$ μ -fast sicher. Das heißt

$$1 = \int_B f_1(x) dx = K \cdot \int_B f_0(x) dx,$$

woraus folgt, daß $K = 1$ und $f_1(x) = f_0(x)$ μ -fast sicher, was aber ein Widerspruch zur eindeutigen Parametrisierung ist. \square

Im Folgenden sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen mit $X_i \sim$ Dichte g_θ , $i = 1, \dots, n$ und

$$(X_1, \dots, X_n) \sim \text{Dichte } f_\theta(x) = \prod_{i=1}^n g_\theta(x_i)$$

aus der Klasse der Verteilungen mit monotonen Dichtekoeffizienten und einer Statistik $T(X_1, \dots, X_n)$.

Wir betrachten die Hypothesen $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ und den Neyman-Pearson-Test:

$$\varphi_{K^*}^*(x) = \begin{cases} 1, & \text{falls } T(x) > K^*, \\ \gamma^*, & \text{falls } T(x) = K^*, \\ 0, & \text{falls } T(x) < K^* \end{cases} \quad (1.3.6)$$

für $K^* \in \mathbb{R}$ und $\gamma^* \in [0, 1]$. Die Gütefunktion von $\varphi_{K^*}^*$ bei θ_0 ist

$$G_n(\theta_0) = \mathbb{E}_0 \varphi_{K^*}^* = \mathbb{P}_0(T(X_1, \dots, X_n) > K^*) + \gamma^* \cdot \mathbb{P}_0(T(X_1, \dots, X_n) = K^*)$$

Satz 1.3.3. 1. Falls $\alpha = \mathbb{E}_0 \varphi_{K^*}^* > 0$, dann ist der soeben definierte Neyman-Pearson-Test ein bester Test der einseitigen Hypothesen H_0 vs. H_1 zum Niveau α .

2. Zu jedem Konfidenzniveau $\alpha \in (0, 1)$ gibt es ein $K^* \in \mathbb{R}$ und $\gamma^* \in [0, 1]$, sodaß $\varphi_{K^*}^*$ ein bester Test zum Umfang α ist.

3. Die Gütefunktion $G_n(\theta)$ von $\varphi_{K^*}^*(\theta)$ ist monoton wachsend in θ . Falls $0 < G_n(\theta) < 1$, dann ist sie sogar streng monoton wachsend.

Beweis. 1. Wähle $\theta_1 > \theta_0$ und betrachte die einfachen Hypothesen $H_0' : \theta = \theta_0$ und $H_1' : \theta = \theta_1$. Sei

$$\varphi_K(x) = \begin{cases} 1, & f_1(x) > K f_0(x), \\ \gamma, & f_1(x) = K f_0(x), \\ 0, & f_1(x) < K f_0(x) \end{cases}$$

der Neyman-Pearson-Test für H'_0, H'_1 mit $K > 0$. Da f_θ den monotonen Dichtekoeffizienten mit Statistik T besitzt,

$$\frac{f_1(x)}{f_0(x)} = h(T(x), \theta_0, \theta_1),$$

existiert ein $K > 0$, so dass

$$\left\{ x : \frac{f_1(x)}{f_0(x)} > K \right\} \subset \left\{ T(x) > K^* \right\} \quad \text{mit } K = h(K^*, \theta_0, \theta_1).$$

φ_K ist ein bester Neyman-Pearson-Test zum Niveau $\alpha = \mathbb{E}_0 \varphi_K = \mathbb{E}_0 \varphi_{K^*}$. Aus $\alpha > 0$ folgt $K < \infty$, denn aus $K = \infty$ würde folgen

$$\begin{aligned} 0 < \alpha = \mathbb{E}_0 \varphi_K &\leq \mathbb{P}_0(T(X_1, \dots, X_n) \geq K^*) \leq \mathbb{P}_0\left(\frac{f_1(X_1, \dots, X_n)}{f_0(X_1, \dots, X_n)} = \infty\right) \\ &= \mathbb{P}_0(f_1(X_1, \dots, X_n) > 0, f_0(X_1, \dots, X_n) = 0) \\ &= \int_B \mathbb{I}(f_1(x) > 0, f_0(x) = 0) \cdot f_0(x) \mu(dx) = 0. \end{aligned}$$

Für den Test φ_{K^*} aus (1.3.6) gilt dann

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } f_1(x)/f_0(x) > K, \\ \gamma^*(x), & \text{falls } f_1(x)/f_0(x) = K, \\ 0, & \text{falls } f_1(x)/f_0(x) < K, \end{cases}$$

wobei $\gamma^*(x) \in \{\gamma^*, 0, 1\}$. Daraus folgt, daß φ_{K^*} ein bester Neyman-Pearson-Test ist für H'_0 vs. H'_1 (vergleiche Bemerkung 1.3.2, 1.) und Bemerkung 1.3.3) für beliebige $\theta_1 > \theta_0$. Deshalb ist φ_{K^*} ein bester Neyman-Pearson-Test für $H''_0 : \theta = \theta_0$ vs. $H''_1 : \theta > \theta_0$ ist.

Die selbe Behauptung erhalten wir aus dem Teil 3. des Satzes für $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, weil dann $G_n(\theta) \leq G_n(\theta_0) = \alpha$ für alle $\theta < \theta_0$.

2. Siehe Beweis zu Satz 1.3.2, 1.).

3. Wir müssen zeigen, daß $G_n(\theta)$ monoton ist. Dazu wählen wir $\theta_1 < \theta_2$ und zeigen, daß $\alpha_1 = G_n(\theta_1) \leq G_n(\theta_2)$. Wir betrachten die neuen, einfachen Hypothesen $H''_0 : \theta = \theta_1$ vs. $H''_1 : \theta = \theta_2$. Der Test φ_{K^*} kann genauso wie in 1. als Neyman-Pearson-Test dargestellt werden (für die Hypothesen H''_0 und H''_1), der ein bester Test zum Niveau α_1 ist. Betrachten wir einen weiteren konstanten Test $\varphi(x) = \alpha_1$. Dann ist $\alpha_1 = \mathbb{E}_{\theta_2} \varphi \leq \mathbb{E}_{\theta_2} \varphi_{K^*} = G_n(\theta_2)$. Daraus folgt, daß $G_n(\theta_1) \leq G_n(\theta_2)$.

Nun zeigen wir, daß für $G_n(\theta) \in (0, 1)$ gilt: $G_n(\theta_1) < G_n(\theta_2)$. Wir nehmen an, daß $\alpha_1 = G_n(\theta_1) = G_n(\theta_2)$ und $\theta_1 < \theta_2$ für $\alpha \in (0, 1)$. Es folgt, daß $\varphi(x) = \alpha_1$ auch ein bester Test für H''_0 und H''_1 ist. Aus Satz 1.3.2, 2.) folgt

$$\mu(\underbrace{\{x \in B : \varphi(x) \neq \varphi_{K^*}(x)\}}_{=\alpha_1}) = 0 \text{ auf } K_0 \cup K_1 = \{f_1(x) \neq K f_0(x)\},$$

was ein Widerspruch zur Bauart des Tests φ_{K^*} ist, der auf $K_0 \cup K_1$ nicht gleich $\alpha_1 \in (0, 1)$ sein kann. □

Bemerkung 1.3.5. 1. Der Satz 1.3.3 ist genauso auf Neyman-Pearson-Tests der einseitigen Hypothesen

$$H_0 : \theta \geq \theta_0 \text{ vs. } H_1 : \theta < \theta_0$$

anwendbar, mit dem entsprechenden Unterschied

$$\begin{aligned} \theta &\mapsto -\theta \\ T &\mapsto -T \end{aligned}$$

Somit existiert der beste α -Test auch in diesem Fall.

2. Man kann zeigen, daß die Gütefunktion $G_n(\varphi_{K^*}^*, \theta)$ des besten Neyman-Pearson-Tests auf $\Theta_0 = (-\infty, \theta_0)$ folgende Minimalitätseigenschaft besitzt:

$$G_n(\varphi_{K^*}^*, \theta) \leq G_n(\varphi, \theta) \quad \forall \varphi \in \Psi(\alpha), \theta \leq \theta_0$$

Beispiel 1.3.3. Wir betrachten eine normalverteilte Stichprobe (X_1, \dots, X_n) von unabhängigen und identisch verteilten Zufallsvariablen X_i , wobei $X_i \sim N(\mu, \sigma_0^2)$ und σ_0^2 sei bekannt. Es werden die Hypothesen

$$H_0 : \mu \leq \mu_0 \text{ vs. } H_1 : \mu > \mu_0,$$

getestet. Aus Beispiel 1.1.2 kennen wir die Testgröße

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma_0},$$

wobei unter H_0 gilt: $T(X_1, \dots, X_n) \sim N(0, 1)$. H_0 wird verworfen, falls

$$T(X_1, \dots, X_n) > z_{1-\alpha}, \quad \text{wobei } \alpha \in (0, 1).$$

Wir zeigen jetzt, daß dieser Test der beste Neyman-Pearson-Test zum Niveau α ist. Aus Beispiel 1.3.2 ist bekannt, daß die Dichte f_n von (X_1, \dots, X_n) aus der einparametrischen Exponentialklasse ist, mit

$$\tilde{T}(X_1, \dots, X_n) = \sum_{i=1}^n X_i.$$

Dann gehört f_μ von (x_1, \dots, x_n) zur einparametrischen Exponentialklasse auch bezüglich der Statistik

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0}$$

Es gilt nämlich

$$\begin{aligned} f_\mu(x) &= \exp\left(\underbrace{\frac{\mu}{\sigma_0^2}}_{\tilde{c}(\mu)} \cdot \underbrace{\sum_{i=1}^n x_i}_{\tilde{T}} - \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{\tilde{a}(\mu)}\right) \cdot l(x) \\ &= \exp\left(\underbrace{\frac{\mu\sqrt{n}}{\sigma_0}}_{c(\mu)} \cdot \underbrace{\sqrt{n}\bar{x}_n - \mu}_{T} + \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{a(\mu)}\right) \cdot l(x). \end{aligned}$$

Die Statistik T kann also in der Konstruktion des Neyman-Pearson-Tests (Gleichung (1.3.6)) verwendet werden:

$$\varphi_{K^*}(x) = \begin{cases} 1, & \text{falls } T(x) > z_{1-\alpha}, \\ 0, & \text{falls } T(x) = z_{1-\alpha}, \\ 0, & \text{falls } T(x) < z_{1-\alpha} \end{cases}$$

(mit $K^* = z_{1-\alpha}$ und $\gamma^* = 0$). Nach Satz 1.3.3 ist dieser Test der beste Neyman-Pearson-Test zum Niveau α für unsere Hypothesen:

$$\begin{aligned} G_n(\varphi_{K^*}, \mu_0) &= \mathbb{P}_0(T(X_1, \dots, X_n) > z_{1-\alpha}) + 0 \cdot \mathbb{P}_0(T(X_1, \dots, X_n) \leq z_{1-\alpha}) \\ &= 1 - \Phi(z_{1-\alpha}) = 1 - (1 - \alpha) = \alpha. \end{aligned}$$

1.3.4 Unverfälschte zweiseitige Tests

Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen mit der Dichte

$$f_\theta(x) = \prod_{i=1}^n g_\theta(x_i).$$

Es wird ein zweiseitiger Test der Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

betrachtet. Für alle $\alpha \in (0, 1)$ kann es jedoch keinen besten Test φ zum Niveau α für H_0 vs. H_1 geben. Denn, nehmen wir an, φ wäre der beste Test zum Niveau α für H_0 vs. H_1 , dann wäre φ der beste Test für die Hypothesen

1. $H'_0 : \theta = \theta_0$ vs. $H'_1 : \theta > \theta_0$
2. $H''_0 : \theta = \theta_0$ vs. $H''_1 : \theta < \theta_0$.

Dann ist nach Satz 1.3.3, 3. die Gütefunktion

1. $G_n(\varphi, \theta) < \alpha$ auf $\theta < \theta_0$, bzw.

2. $G_n(\varphi, \theta) > \alpha$ auf $\theta < \theta_0$,

was ein Widerspruch ist!

Darum werden wir die Klasse aller möglichen Tests auf unverfälschte Tests (Definition 1.1.5) eingrenzen. Der Test φ ist unverfälscht genau dann, wenn

$$\begin{aligned} G_n(\varphi, \theta) &\leq \alpha \text{ für } \theta \in \Theta_0 \\ G_n(\varphi, \theta) &\geq \alpha \text{ für } \theta \in \Theta_1 \end{aligned}$$

Beispiel 1.3.4. 1. $\varphi(x) \equiv \alpha$ ist unverfälscht.

2. Der zweiseitige Gauß-Test ist unverfälscht, vergleiche Beispiel 1.1.2: $G_n(\varphi, \mu) \geq \alpha$ für alle $\mu \in \mathbb{R}$.

Im Folgenden seien X_i unabhängig und identisch verteilt. Die Dichte f_θ von (X_1, \dots, X_n) gehöre zur einparametrischen Exponentialklasse:

$$f_\theta(x) = \exp \{c(\theta) \cdot T(x) + a(\theta)\} \cdot l(x), \quad (1.3.7)$$

wobei $c(\theta)$ und $a(\theta)$ stetig differenzierbar auf Θ sein sollen, mit

$$c'(\theta) > 0 \quad \text{und} \quad \text{Var}_\theta T(X_1, \dots, X_n) > 0$$

für alle $\theta \in \Theta$. Sei $f_\Phi(x)$ stetig in (x, Θ) auf $B \times \Theta$.

Übungsaufgabe 1.3.1. Zeigen Sie, daß folgende Relation gilt:

$$a'(\theta) = -c'(\theta) \mathbb{E}_\theta T(X_1, \dots, X_n).$$

Lemma 1.3.3. Es sei φ ein unverfälschter Test zum Niveau α für

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

Dann gilt:

1. $\alpha = \mathbb{E}_0 \varphi(X_1, \dots, X_n) = G_n(\varphi, \theta_0)$
2. $\mathbb{E}_0 [T(X_1, \dots, X_n) \varphi(X_1, \dots, X_n)] = \alpha \cdot \mathbb{E}_0 T(X_1, \dots, X_n)$

Beweis. 1. Die Gütefunktion von φ ist

$$G_n(\varphi, \theta) = \int_B \varphi(x) f_\theta(x) \mu(dx)$$

Da f_θ aus der einparametrischen Exponentialklasse ist, ist $G_n(\varphi, \theta)$ differenzierbar (unter dem Integral) bezüglich θ . Wegen der Unverfälschtheit von φ gilt

$$G_n(\varphi, \theta_0) \leq \alpha, \quad G_n(\varphi, \theta) \geq \alpha, \quad \theta \neq \theta_0$$

und daraus folgt $G_n(\varphi, \theta_0) = \alpha$ und θ_0 ist ein Minimumpunkt von G_n . Somit ist 1) bewiesen.

2. Da θ_0 der Minimumpunkt von G_n ist, gilt

$$\begin{aligned} 0 &= G'_n(\varphi, \theta_0) = \int_B \varphi(x)(c'(\theta_0)T(x) + a'(\theta_0))f_0(x)\mu(dx) \\ &= c'(\theta_0) \cdot \mathbb{E}_0 [\varphi(X_1, \dots, X_n)T(X_1, \dots, X_n)] + a'(\theta) \cdot G_n(\varphi, \theta_0) \\ &= c'(\theta_0) \cdot \mathbb{E}_0 [\varphi(X_1, \dots, X_n)T(X_1, \dots, X_n)] + \alpha a'(\theta_0) \\ &\stackrel{\text{(Übung 1.3.1)}}{=} c'(\theta_0) (\mathbb{E}_0 (\varphi \cdot T) - \alpha \mathbb{E}_0 T) \end{aligned}$$

Daraus folgt $\mathbb{E}_0 (\varphi T) = \alpha \mathbb{E}_0 T$ und damit ist das Lemma bewiesen. \square

Wir definieren jetzt die modifizierten Neyman-Pearson-Tests für einfache Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1' : \theta = \theta_1, \quad \theta_1 \neq \theta_0.$$

Für $\lambda, K \in \mathbb{R}$, $\gamma : B \rightarrow [0, 1]$ definieren wir

$$\varphi_{K,\lambda}(x) = \begin{cases} 1, & \text{falls } f_1(x) > (K + \lambda T(x))f_0(x), \\ \gamma(x), & \text{falls } f_1(x) = (K + \lambda T(x))f_0(x), \\ 0, & \text{falls } f_1(x) < (K + \lambda T(x))f_0(x), \end{cases} \quad (1.3.8)$$

wobei $T(x)$ die Statistik aus der Darstellung (1.3.7) ist.

Es sei $\tilde{\Psi}(\alpha)$ die Klasse aller Tests, die Aussagen 1) und 2) des Lemmas 1.3.3 erfüllen. Aus Lemma 1.3.3 folgt dann, daß die Menge der unverfälschten Tests zum Niveau α eine Teilmenge von $\tilde{\Psi}(\alpha)$ ist.

Satz 1.3.4. Der modifizierte Neyman-Pearson-Test $\varphi_{K,\lambda}$ ist der beste α -Test in $\tilde{\Psi}(\alpha)$ für Hypothesen H_0 vs. H_1' zum Niveau $\alpha = \mathbb{E}_0 \varphi_{K,\lambda}$, falls $\varphi_{K,\lambda} \in \tilde{\Psi}(\alpha)$.

Beweis. Es ist zu zeigen, daß $\mathbb{E}_1 \varphi_{K,\lambda} \geq \mathbb{E}_1 \varphi$ für alle $\varphi \in \tilde{\Psi}(\alpha)$, bzw. $\mathbb{E}_1 (\varphi_{K,\lambda} - \varphi) \geq 0$. Es gilt

$$\begin{aligned} \mathbb{E}_1 (\varphi_{K,\lambda} - \varphi) &= \int_B (\varphi_{K,\lambda}(x) - \varphi(x))f_1(x)\mu(dx) \\ &\stackrel{\text{(Bem. 1.3.3, 2.)}}{\geq} \int_B (\varphi_{K,\lambda}(x) - \varphi(x))(K + \lambda T(x))f_0(x)\mu(dx) \\ &= K \left(\underbrace{\mathbb{E}_0 \varphi_{K,\lambda}}_{=\alpha} - \underbrace{\mathbb{E}_0 \varphi}_{=\alpha} \right) + \lambda \left(\underbrace{\mathbb{E}_0 (\varphi_{K,\lambda} \cdot T)}_{\alpha \mathbb{E}_0 T} - \underbrace{\mathbb{E}_0 (\varphi \cdot T)}_{=\alpha \mathbb{E}_0 T} \right) \\ &= 0, \end{aligned}$$

weil $\varphi, \varphi_{K,\lambda} \in \tilde{\Psi}(\alpha)$. \square

Wir definieren folgende Entscheidungsregel, die später zum Testen der zweiseitigen Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

verwendet wird:

$$\varphi_c(x) = \begin{cases} 1, & \text{falls } T(x) \notin (c_1, c_2), \\ \gamma_1, & \text{falls } T(x) = c_1, \\ \gamma_2, & \text{falls } T(x) = c_2, \\ 0, & \text{falls } T(x) \in (c_1, c_2), \end{cases} \quad (1.3.9)$$

für $c_1 \leq c_2 \in \mathbb{R}$, $\gamma_1, \gamma_2 \in [0, 1]$ und die Statistik $T(x)$, $x = (x_1, \dots, x_n) \in B$, die in der Dichte (1.3.7) vorkommt. Zeigen wir, daß φ_c sich als modifizierter Neyman-Pearson-Test schreiben lässt.

Für die Dichte

$$f_\theta(x) = \exp\{c(\theta)T(x) + a(\theta)\} \cdot l(x)$$

wird (wie immer) vorausgesetzt, daß $l(x) > 0$, $c'(\theta) > 0$ und $a'(\theta)$ existiert für $\theta \in \Theta$.

Lemma 1.3.4. Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen mit gemeinsamer Dichte $f_\theta(x)$, $x \in B$, die zur einparametrischen Exponentialfamilie gehört. Sei $T(x)$ die dazugehörige Statistik, die im Exponenten der Dichte f_θ vorkommt. Für beliebige reelle Zahlen $c_1 \leq c_2$, $\gamma_1, \gamma_2 \in [0, 1]$ und Parameterwerte $\theta_0, \theta_1 \in \Theta : \theta_0 \neq \theta_1$ läßt sich der Test φ_c aus (1.3.9) als modifizierter Neyman-Pearson-Test $\varphi_{K,\lambda}$ aus (1.3.8) mit gegebenen $K, \lambda \in \mathbb{R}$, $\gamma(x) \in [0, 1]$ schreiben.

Beweis. Falls wir die Bezeichnung

$$f_{\theta_i}(x) = f_i(x), \quad i = 0, 1$$

verwenden, dann gilt

$$\frac{f_1(x)}{f_0(x)} = \exp \left\{ \underbrace{(c(\theta_1) - c(\theta_0))}_{c} T(x) + \underbrace{a(\theta_1) - a(\theta_0)}_a \right\},$$

und somit

$$\{x \in B : f_1(x) > (K + \lambda T(x)) f_0(x)\} = \{x \in B : \exp(cT(x) + a) > K + \lambda T(x)\}.$$

Finden wir solche K und λ aus \mathbb{R} , für die die Gerade $K + \lambda t$, $t \in \mathbb{R}$ die konvexe Kurve $\exp\{ct + a\}$ genau an den Stellen c_1 und c_2 schneidet (falls $c_1 \neq c_2$) bzw. an der Stelle $t = c_1$ berührt (falls $c_1 = c_2$). Dies ist immer möglich, siehe Abbildung 1.1.

Ferner setzen wir $\gamma(x) = \gamma_i$ für $\{x \in B : T(x) = c_i\}$. Insgesamt gilt dann

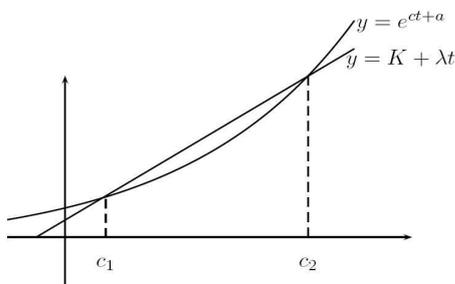
$$\{x : \exp(cT(x) + a) > K + \lambda T(x)\} = \{x : T(x) \notin [c_1, c_2]\}$$

und

$$\{x : \exp(cT(x) + a) < K + \lambda T(x)\} = \{x : T(x) \in (c_1, c_2)\}.$$

Damit ist das Lemma bewiesen. □

Abbildung 1.1:



Bemerkung 1.3.6. 1. Die Umkehrung des Lemmas stimmt nicht, denn bei vorgegebenen Kurven $y = K + \lambda t$ und $y = \exp\{ct + a\}$ muss es die Schnittpunkte c_1 und c_2 nicht unbedingt geben. So kann die Gerade vollständig unter der Kurve $y = \exp\{ct + a\}$ liegen.

2. Der Test φ_c macht von den Werten θ_0 und θ_1 nicht explizit Gebrauch. Dies unterscheidet ihn vom Test $\varphi_{K,\lambda}$, für den die Dichten f_0 und f_1 gebraucht werden.

Jetzt sind wir bereit, den Hauptsatz über zweiseitige Tests zum Prüfen der Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

zu formulieren und zu beweisen.

Satz 1.3.5. (Hauptsatz über zweiseitige Tests)

Unter den Voraussetzungen des Lemmas 1.3.4 sei φ_c ein Test aus (1.3.9), für den $\varphi_c \in \tilde{\Psi}(\alpha)$ gilt. Dann ist φ_c bester unverfälschter Test zum Niveau α (und dadurch bester Test in $\tilde{\Psi}(\alpha)$) der Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

Beweis. Wählen wir ein beliebiges $\theta_1 \in \Theta$, $\theta_1 \neq \theta_0$. Nach Lemma 1.3.4 ist φ_c ein modifizierter Neyman-Pearson-Test $\varphi_{K,\lambda}$ für eine spezielle Wahl von K und $\lambda \in \mathbb{R}$. $\varphi_{K,\lambda}$ ist aber nach Satz 1.3.4 bester Test in $\tilde{\Psi}(\alpha)$ für $H_0 : \theta = \theta_0$ vs. $H_1' : \theta = \theta_1$. Da φ_c nicht von θ_1 abhängt, ist es bester Test in $\tilde{\Psi}(\alpha)$ für $H_1 : \theta \neq \theta_0$. Da unverfälschte Niveau- α -Tests in $\tilde{\Psi}(\alpha)$ liegen, müssen wir nur zeigen, daß φ_c unverfälscht ist. Da φ_c der beste Test ist, ist er nicht schlechter als der konstante unverfälschte Test $\varphi = \alpha$, das heißt

$$G_n(\varphi_c, \theta) \geq G_n(\varphi, \theta) = \alpha, \quad \theta \neq \theta_0.$$

Somit ist auch φ_c unverfälscht. Der Beweis ist beendet. \square

Bemerkung 1.3.7. Wir haben gezeigt, daß φ_c der beste Test seines Umfangs ist. Es wäre jedoch noch zu zeigen, daß für beliebiges $\alpha \in (0, 1)$ Konstanten $c_1, c_2, \gamma_1, \gamma_2$ gefunden werden, für die $\mathbb{E}_0 \varphi_c = \alpha$ gilt. Da der Beweis schwierig ist, wird er hier ausgelassen. Im folgenden Beispiel jedoch wird es klar, wie die Parameter $c_1, c_2, \gamma_1, \gamma_2$ zu wählen sind.

Beispiel 1.3.5 (Zweiseitiger Gauß-Test). Im Beispiel 1.1.2 haben wir folgenden Test des Erwartungswertes einer normalverteilten Stichprobe (X_1, \dots, X_n) mit unabhängigen und identisch verteilten X_i und $X_i \sim N(\mu, \sigma_0^2)$ bei bekannten Varianzen σ_0^2 betrachtet. Getestet werden die Hypothesen

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0.$$

Der Test $\varphi(x)$ lautet

$$\varphi(x) = \mathbb{I}(x \in \mathbb{R}^n : |T(x)| > z_{1-\alpha/2}),$$

wobei

$$T(x) = \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0}.$$

Zeigen wir, daß φ der beste Test zum Niveau α in $\tilde{\Psi}(\alpha)$ (und somit bester unverfälschter Test) ist. Nach Satz 1.3.5 müssen wir lediglich prüfen, daß φ als φ_c mit (1.3.9) dargestellt werden kann, weil die n -dimensionale Normalverteilung mit Dichte f_μ (siehe Beispiel 1.3.3) zu der einparametrischen Exponentialfamilie mit Statistik

$$T(x) = \sqrt{n} \frac{\bar{x}_n - \mu}{\sigma_0}$$

gehört. Setzen wir $c_1 = z_{1-\alpha/2}$, $c_2 = -z_{1-\alpha/2}$, $\gamma_1 = \gamma_2 = 0$. Damit ist

$$\varphi(x) = \varphi_c(x) = \begin{cases} 1, & \text{falls } |T(x)| > z_{1-\alpha/2}, \\ 0, & \text{falls } |T(x)| \leq z_{1-\alpha/2}. \end{cases}$$

und die Behauptung ist bewiesen, weil aus der in Beispiel 1.1.2 ermittelten Gütefunktion $G_n(\varphi, \theta)$ von φ ersichtlich ist, daß φ ein unverfälschter Test zum Niveau α ist (und somit $\varphi \in \tilde{\Psi}(\alpha)$).

Bemerkung 1.3.8. Bisher haben wir immer vorausgesetzt, daß nur *ein* Parameter der Verteilung der Stichprobe (X_1, \dots, X_n) unbekannt ist, um die Theorie des Abschnittes 1.3 über die besten (Neyman-Pearson-) Tests im Fall der einparametrischen Exponentialfamilie aufstellen zu können. Um jedoch den Fall weiterer unbekannter Parameter betrachten zu können (wie im Beispiel der zweiseitigen Tests des Erwartungswertes der normalverteilten Stichprobe bei unbekannter Varianz (der sog. t -Test, vergleiche Abschnitt 1.2.1, 1 (a)), bedarf es einer tiefergehenden Theorie, die aus Zeitgründen in dieser Vorlesung nicht behandelt wird. Der interessierte Leser findet das Material dann im Buch [14].

1.4 Anpassungstests

Sei eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen (X_1, \dots, X_n) gegeben mit $X_i \sim F$ (Verteilungsfunktion) für $i = 1, \dots, n$. Bei den Anpassungstests wird die Hypothese

$$H_0 : F = F_0 \text{ vs. } H_1 : F \neq F_0$$

überprüft, wobei F_0 eine vorgegebene Verteilungsfunktion ist.

Einen Test aus dieser Klasse haben wir bereits in der Vorlesung Stochastik I kennengelernt: den Kolmogorow-Smirnov-Test (vergleiche Bemerkung 3.3.8. 3), Vorlesungsskript Stochastik I).

Jetzt werden weitere nichtparametrische Anpassungstests eingeführt. Der erste ist der χ^2 -Anpassungstest von K. Pearson.

1.4.1 χ^2 -Anpassungstest

Der Test von Kolmogorov-Smirnov basiert auf dem Abstand

$$D_n = \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F_0(x) |$$

zwischen der empirischen Verteilungsfunktion der Stichprobe (X_1, \dots, X_n) und der Verteilungsfunktion F_0 . In der Praxis jedoch erscheint dieser Test zu feinfühlig, denn er ist zu sensibel gegenüber Unregelmäßigkeiten in den Stichproben und verwirft H_0 zu oft. Einen Ausweg aus dieser Situation stellt die Vergrößerung der Haupthypothese H_0 dar, auf welcher der folgende χ^2 -Anpassungstest beruht.

Man zerlegt den Wertebereich der Stichprobenvariablen X_i in r Klassen $(a_j, b_j]$, $j = 1, \dots, r$ mit der Eigenschaft

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = \dots = a_r < b_r \leq \infty.$$

Anstelle von $X_i, i = 1, \dots, n$ betrachten wir die sogenannten *Klassenstärken* $Z_j, j = 1, \dots, r$, wobei

$$Z_j = \#\{i : a_j < X_i \leq b_j, 1 \leq i \leq n\}.$$

Lemma 1.4.1. Der Zufallsvektor $Z = (Z_1, \dots, Z_r)^\top$ ist *multinomialverteilt* mit Parametervektor

$$p = (p_1, \dots, p_{r-1})^\top \in [0, 1]^{r-1},$$

wobei

$$p_j = \mathbb{P}(a_j < X_1 \leq b_j) = F(b_j) - F(a_j), \quad j = 1, \dots, r-1, \quad p_r = 1 - \sum_{j=1}^{r-1} p_j.$$

Schreibweise:

$$Z \sim M_{r-1}(n, p)$$

Beweis. Es ist zu zeigen, daß für alle Zahlen $k_1, \dots, k_r \in \mathbb{N}_0$ mit $k_1 + \dots + k_r = n$ gilt:

$$\mathbb{P}(Z_i = k_i, i = 1, \dots, r) = \frac{n!}{k_1! \cdot \dots \cdot k_r!} p_1^{k_1} \cdot \dots \cdot p_r^{k_r}. \quad (1.4.1)$$

Da X_i unabhängig und identisch verteilt sind, gilt

$$\mathbb{P}(X_j \in (a_{i_j}, b_{i_j}], j = 1, \dots, n) = \prod_{j=1}^n \mathbb{P}(a_{i_j} < X_1 \leq b_{i_j}) = p_1^{k_1} \cdot \dots \cdot p_r^{k_r},$$

falls die Folge von Intervallen $(a_{i_j}, b_{i_j}]_{j=1, \dots, n}$ das Intervall $(a_i, b_i]$ k_i Mal enthält, $i = 1, \dots, r$. Die Formel (1.4.1) ergibt sich aus dem Satz der totalen Wahrscheinlichkeit als Summe über die Permutationen von Folgen $(a_{i_j}, b_{i_j}]_{j=1, \dots, n}$ dieser Art. \square

Im Sinne des Lemmas 1.4.1 werden neue Hypothesen über die Beschaffenheit von F geprüft.

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0,$$

wobei $p = (p_1, \dots, p_{r-1})^\top$ der Parametervektor der Multinomialverteilung von Z ist, und $p_0 = (p_{01}, \dots, p_{0,r-1})^\top \in (0, 1)^{r-1}$ mit $\sum_{i=1}^{r-1} p_{0i} < 1$. In diesem Fall ist

$$\Lambda_0 = \{F \in \Lambda : F(b_j) - F(a_j) = p_{0j}, \quad j = 1, \dots, r-1\},$$

$\Lambda_1 = \Lambda \setminus \Lambda_0$, wobei Λ die Menge aller Verteilungsfunktionen ist. Um H_0 vs. H_1 zu testen, führen wir die *Pearson-Teststatistik*

$$T_n(x) = \sum_{j=1}^r \frac{(z_j - np_{0j})^2}{np_{0j}}$$

ein, wobei $x = (x_1, \dots, x_n)$ eine konkrete Stichprobe der Daten ist und $z_j, j = 1, \dots, r$ ihre Klassenstärken sind.

Unter H_0 gilt

$$\mathbb{E} Z_j = np_{0j}, \quad j = 1, \dots, r,$$

somit soll H_0 abgelehnt werden, falls $T_n(X)$ ungewöhnlich große Werte annimmt.

Im nächsten Satz zeigen wir, daß $T(X_1, \dots, X_n)$ asymptotisch (für $n \rightarrow \infty$) χ_{r-1}^2 -verteilt ist, was zu folgendem Anpassungstest (χ^2 -Anpassungstest) führt:

$$H_0 \text{ wird verworfen, falls } T_n(x_1, \dots, x_n) > \chi_{r-1, 1-\alpha}^2.$$

Dieser Test ist nach seinem Entdecker *Karl Pearson* (1857-1936) benannt worden.

Satz 1.4.1. Unter H_0 gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p_0}(T_n(X_1, \dots, X_n) > \chi_{r-1, 1-\alpha}^2) = \alpha, \quad \alpha \in (0, 1),$$

das heißt, der χ^2 -Pearson-Test ist ein asymptotischer Test zum Niveau α .

Beweis. Führen wir die Bezeichnung $Z_{nj} = Z_j(X_1, \dots, X_n)$ der Klassenstärken ein, die aus der Stichprobe (X_1, \dots, X_n) entstehen. Nach Lemma 1.4.1 ist

$$Z_n = (Z_{n1}, \dots, Z_{nr}) \sim M_{r-1}(n, p_0) \text{ unter } H_0.$$

Insbesondere soll $\mathbb{E} Z_{nj} = np_{0j}$ und

$$\text{Cov}(Z_{ni}, Z_{nj}) = \begin{cases} np_{0j}(1 - p_{0j}), & i = j, \\ -np_{0i}p_{0j}, & i \neq j \end{cases}$$

für alle $i, j = 1, \dots, r$ gelten. Da

$$Z_{nj} = \sum_{i=1}^n \mathbb{I}(a_j < X_i \leq b_j), \quad j = 1, \dots, r,$$

ist $Z_n = (Z_{n1}, \dots, Z_{n,r-1})$ eine Summe von n unabhängigen und identisch verteilten Zufallsvektoren $Y_i \in \mathbb{R}^{r-1}$ mit Koordinaten $Y_{ij} = \mathbb{I}(a_j < X_i \leq b_j)$, $j = 1, \dots, r-1$. Daher gilt nach dem multivariaten Grenzwertsatz (der in Lemma 1.4.2 bewiesen wird), daß

$$Z'_n = \frac{Z_n - \mathbb{E} Z_n}{\sqrt{n}} = \frac{\sum_{i=1}^n Y_i - n\mathbb{E} Y_1}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, K),$$

mit $N(0, K)$ eine $(r-1)$ -dimensionale multivariate Normalverteilung (vergleiche Vorlesungsskript WR, Beispiel 3.4.1. 3.) mit Erwartungswertvektor Null und Kovarianzmatrix $K = (\sigma_{ij}^2)$, wobei

$$\sigma_{ij}^2 = \begin{cases} -p_{0i}p_{0j}, & i \neq j, \\ p_{0i}(1 - p_{0j}), & i = j \end{cases}$$

für $i, j = 1, \dots, r-1$ ist. Diese Matrix K ist invertierbar mit $K^{-1} = A = (a_{ij})$,

$$a_{ij} = \begin{cases} \frac{1}{p_{0r}}, & i \neq j, \\ \frac{1}{p_{0i}} + \frac{1}{p_{0r}}, & i = j. \end{cases}$$

Außerdem ist K (als Kovarianzmatrix) symmetrisch und positiv definit. Aus der linearen Algebra ist bekannt, daß es eine invertierbare $(r-1) \times (r-1)$ -Matrix $A^{1/2}$ gibt, mit der Eigenschaft $A = A^{1/2}(A^{1/2})^\top$. Daraus folgt,

$$K = A^{-1} = ((A^{1/2})^\top)^{-1} \cdot (A^{1/2})^{-1}.$$

Wenn wir $(A^{1/2})^\top$ auf Z'_n anwenden, so bekommen wir

$$(A^{1/2})^\top \cdot Z'_n \xrightarrow[n \rightarrow \infty]{d} (A^{1/2})^\top \cdot Y,$$

wobei

$$(A^{1/2})^\top \cdot Y \sim N\left(0, (A^{1/2})^\top \cdot K \cdot A^{1/2}\right) = N(0, \mathcal{I}_{r-1})$$

nach der Eigenschaft der multivariaten Normalverteilung, die im Kapitel 2, Satz 2.1.3 behandelt wird. Des Weiteren wurde hier der Stetigkeitssatz aus der Wahrscheinlichkeitsrechnung benutzt, daß

$$Y_n \xrightarrow[n \rightarrow \infty]{d} Y \implies \varphi(Y_n) \xrightarrow[n \rightarrow \infty]{d} \varphi(Y)$$

für beliebige Zufallsvektoren $\{Y_n\}$, $Y \in \mathbb{R}^m$ und stetige Abbildungen $\varphi: \mathbb{R} \rightarrow \mathbb{R}$. Diesen Satz haben wir in WR für Zufallsvariablen bewiesen (Satz 6.4.3, Vorlesungsskript WR). Die erneute Anwendung des Stetigkeitssatzes ergibt

$$\left| (A^{1/2})^\top Z'_n \right|^2 \xrightarrow[n \rightarrow \infty]{d} |Y|^2 = R \sim \chi_{r-1}^2.$$

Zeigen wir, daß

$$T_n(X_1, \dots, X_n) = \left| (A^{1/2})^\top Z'_n \right|^2.$$

Es gilt:

$$\begin{aligned} \left| (A^{1/2})^\top Z'_n \right|^2 &= ((A^{1/2})^\top Z'_n)^\top ((A^{1/2})^\top Z'_n) = Z_n'^\top \cdot \underbrace{A^{1/2} \cdot (A^{1/2})^\top}_A Z'_n = Z_n'^\top A Z'_n \\ &= n \sum_{j=1}^{r-1} \frac{1}{p_{0j}} \left(\frac{Z_{nj}}{n} - p_{0j} \right)^2 + \frac{n}{p_{0r}} \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} \left(\frac{Z_{ni}}{n} - p_{0i} \right) \left(\frac{Z_{nj}}{n} - p_{0j} \right) \\ &= \sum_{j=1}^{r-1} \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} + \frac{n}{p_{0r}} \left(\sum_{j=1}^{r-1} \left(\frac{Z_{nj}}{n} - p_{0j} \right) \right)^2 \\ &= \sum_{j=1}^{r-1} \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} + \frac{n}{p_{0r}} \left(\frac{Z_{nr}}{n} - p_{0r} \right)^2 \\ &= \sum_{j=1}^r \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} = T_n(X_1, \dots, X_n), \end{aligned}$$

weil

$$\begin{aligned} \sum_{j=1}^{r-1} Z_{nj} &= n - Z_{nr}, \\ \sum_{j=1}^{r-1} p_{0j} &= 1 - p_{0r}. \end{aligned}$$

□

Lemma 1.4.2 (Multivariater zentraler Grenzwertsatz). Sei $\{Y_n\}_{n \in \mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvektoren, mit $\mathbb{E} Y_1 = \mu$ und Kovarianzmatrix K . Dann gilt

$$\frac{\sum_{i=1}^n Y_i - n\mu}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, K). \quad (1.4.2)$$

Beweis. Sei $Y_j = (Y_{j1}, \dots, Y_{jm})^\top$. Nach dem Stetigkeitssatz für charakteristische Funktionen ist die Konvergenz (1.4.2) äquivalent zu

$$\varphi_n(t) \xrightarrow[n \rightarrow \infty]{} \varphi(t) \quad t \in \mathbb{R}^m, \quad (1.4.3)$$

wobei

$$\varphi_n(t) = \mathbb{E} e^{itS_n} = \mathbb{E} \exp \left\{ i \sum_{j=1}^m t_j \frac{Y_{1j} + \dots + Y_{nj} - n\mu_j}{\sqrt{n}} \right\}$$

die charakteristische Funktion vom Zufallsvektor

$$S_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sqrt{n}}$$

und

$$\varphi(t) = e^{-t^\top K t / 2}$$

die charakteristische Funktion der $N(0, K)$ -Verteilung ist. Die Funktion $\varphi_n(t)$ kann in der Form

$$\varphi_n(t) = \mathbb{E} \exp \left\{ i \sum_{i=1}^n \frac{\sum_{j=1}^m t_j (Y_{ij} - \mu_j)}{\sqrt{n}} \right\}, \quad t = (t_1, \dots, t_m)^\top \in \mathbb{R}^m$$

umgeschrieben werden, wobei für die Zufallsvariable

$$L_i := \sum_{j=1}^m t_j (Y_{ij} - \mu_j)$$

gilt:

$$\begin{aligned} \mathbb{E} L_i &= 0, \\ \text{Var } L_i &= \mathbb{E} \left[\sum_{k,j=1}^m t_j (Y_{ij} - \mu_j) (Y_{ik} - \mu_k) t_k \right] = t^\top K t, \quad i \in \mathbb{N}. \end{aligned}$$

Falls $t^\top Kt = 0$, dann gilt $L_i = 0$ fast sicher, für alle $i \in \mathbb{N}$. Hieraus folgt $\varphi_n(t) = \varphi(t) = 1$, also gilt die Konvergenz 1.4.2.

Falls jedoch $t^\top Kt > 0$, dann kann $\varphi_n(t)$ als charakteristische Funktion der Zufallsvariablen

$$\sum_{i=1}^n L_i / \sqrt{n}$$

an Stelle 1, und $\varphi(t)$ als charakteristische Funktion der eindimensionalen Normalverteilung $N(0, t^\top Kt)$ an Stelle 1 interpretiert werden. Aus dem zentralen Grenzwertsatz für eindimensionale Zufallsvariablen (vergleiche Satz 7.2.1, Vorlesungsskript WR) gilt

$$\sum_{i=1}^n \frac{L_i}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} L \sim N(0, t^\top Kt)$$

und somit

$$\varphi_n(t) = \varphi\left(\sum_{i=1}^n L_i / \sqrt{n}\right)(1) \xrightarrow[n \rightarrow \infty]{} \varphi_L(1) = \varphi(t).$$

Somit ist die Konvergenz (1.4.2) bewiesen. \square

Bemerkung 1.4.1. 1. Die im letzten Beweis verwendete Methode der Reduktion einer mehrdimensionalen Konvergenz auf den eindimensionalen Fall mit Hilfe von Linearkombinationen von Zufallsvariablen trägt den Namen von *Cramér-Wold*.

2. Der χ^2 -Pearson-Test ist asymptotisch, also für große Stichprobenumfänge, anzuwenden. Aber welches n ist groß genug? Als „Faustregel“ gilt: np_{0j} soll größer gleich a sein, $a \in (2, \infty)$. Für eine größere Klassenanzahl $r \geq 10$ kann sogar $a = 1$ verwendet werden. Wir zeigen jetzt, daß der χ^2 -Anpassungstest konsistent ist.

Lemma 1.4.3. Der χ^2 -Pearson-Test ist konsistent, das heißt

$$\forall p \in [0, 1]^{r-1}, p \neq p_0 \text{ gilt: } \lim_{n \rightarrow \infty} \mathbb{P}_p(T_n(X_1, \dots, X_n) > \chi_{r-1, 1-\alpha}^2) = 1$$

Beweis. Unter H_1 gilt

$$Z_{nj}/n = \frac{\sum_{i=1}^n \mathbb{I}(a_j < X_i \leq b_j)}{n} \xrightarrow[n \rightarrow \infty]{f.s.} \underbrace{\mathbb{E} \mathbb{I}(a_j < X_1 \leq b_j)}_{=p_j}$$

nach dem starken Gesetz der großen Zahlen. Wir wählen j so, daß $p_j \neq p_{0j}$. Es gilt

$$T_n(X_1, \dots, X_n) \geq \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} \geq \underbrace{n \left(\frac{Z_{nj}}{n} - p_{0j} \right)^2}_{\sim n(p_j - p_{0j})^2} \xrightarrow[n \rightarrow \infty]{f.s.} \infty.$$

Somit ist auch

$$\mathbb{P}_p(T_n(X_1, \dots, X_n) > \chi_{r-1, 1-\alpha}^2) \xrightarrow[n \rightarrow \infty]{f.s.} 1.$$

\square

1.4.2 χ^2 -Anpassungstest von Pearson-Fisher

Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen X_i , $i = 1, \dots, n$. Wir wollen testen, ob die Verteilungsfunktion F von X_i zu einer parametrischen Familie

$$\Lambda_0 = \{F_\theta : \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}^m$$

gehört. Seien die Zahlen a_i, b_i , $i = 1, \dots, r$ vorgegeben mit der Eigenschaft

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = \dots = a_r < b_r \leq \infty$$

und

$$\begin{aligned} Z_j &= \#\{X_i, i = 1, \dots, n : a_j < X_i \leq b_j\}, \quad j = 1, \dots, r, \\ Z &= (Z_1, \dots, Z_r)^\top. \end{aligned}$$

Nach Lemma 1.4.1 gilt: $Z \sim M_{r-1}(n, p)$, $p = (p_0, \dots, p_{r-1})^\top \in [0, 1]^{r-1}$. Unter der Hypothese $H_0 : F \in \Lambda_0$ gilt: $p = p(\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$. Wir vergrößern die Hypothese H_0 und wollen folgende neue Hypothese testen:

$$H_0 : p \in \{p(\theta) : \theta \in \Theta\} \text{ vs. } H_1 : p \notin \{p(\theta) : \theta \in \Theta\}.$$

Um dieses Hypothesenpaar zu testen, wird der χ^2 -Pearson-Fisher-Test wie folgt aufgebaut:

1. Ein (schwach konsistenter) Maximum-Likelihood-Schätzer $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ für θ wird gefunden: $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{P} \theta$. Dabei muß $\{\hat{\theta}_n\}_{n \in \mathbb{N}}$ asymptotisch normalverteilt sein.
2. Es wird der Plug-In-Schätzer $p(\hat{\theta}_n)$ für $p(\theta)$ gebildet.
3. Die Testgröße

$$\hat{T}_n(X_1, \dots, X_n) = \sum_{j=1}^r \frac{(Z_{nj} - np_j(\hat{\theta}))^2}{np_j(\hat{\theta})} \xrightarrow[n \rightarrow \infty]{P} \eta \sim \chi_{r-m-1}^2$$

unter H_0 und gewissen Voraussetzungen.

4. H_0 wird verworfen, falls $\hat{T}_n(X_1, \dots, X_n) > \chi_{r-m-1, 1-\alpha}^2$. Dies ist ein asymptotischer Test zum Niveau α .

Bemerkung 1.4.2. 1. Bei einem χ^2 -Pearson-Fisher-Test wird vorausgesetzt, daß die Funktion $p(\theta)$ explizit bekannt ist, θ jedoch unbekannt. Das bedeutet, daß für jede Klasse von Verteilungen Λ_0 die Funktion $p(\cdot)$ berechnet werden soll.

2. Warum kann \hat{T}_n die Hypothese H_0 von H_1 unterscheiden? Nach dem Gesetz der großen Zahlen gilt

$$\frac{1}{n}Z_{nj} - p_j(\hat{\theta}_n) = \underbrace{\frac{1}{n}Z_{nj} - p_j(\theta)}_{\xrightarrow{P} 0} - \underbrace{(p_j(\hat{\theta}_n) - p_j(\theta))}_{\xrightarrow{P} 0} \xrightarrow[n \rightarrow \infty]{P} 0,$$

falls $\hat{\theta}_n$ schwach konsistent ist und $p_j(\cdot)$ eine stetige Funktion für alle $j = 1, \dots, r$ ist.

Das heißt, unter H_0 soll $\hat{T}_n(X_1, \dots, X_n)$ relativ kleine Werte annehmen. Eine signifikante Abweichung von diesem Verhalten soll zur Ablehnung von H_0 führen, vergleiche Punkt 4.

Für die Verteilung F_θ von X_i gelten folgende Regularitätsvoraussetzungen (vergleiche Satz 3.4.2, Vorlesungsskript Stochastik I).

1. Die Verteilungsfunktion F_θ ist entweder diskret oder absolut stetig für alle $\theta \in \Theta$.
2. Die Parametrisierung ist eindeutig, das heißt: $\theta \neq \theta_1 \Leftrightarrow F_\theta \neq F_{\theta_1}$.
3. Der Träger der Likelihood-Funktion

$$L(x, \theta) = \begin{cases} P_\theta(X_1 = x), & \text{im Falle von diskreten } F_\theta, \\ f_\theta(x), & \text{im absolut stetigen Fall.} \end{cases}$$

$\text{Supp}L(x, \theta) = \{x \in \mathbb{R} : L(x, \theta) > 0\}$ hängt nicht von θ ab.

4. $L(x, \theta)$ sei 3 Mal stetig differenzierbar, und es gelte für $k = 1, \dots, 3$ und $i_1, \dots, i_k \in \{1 \dots m\}$, daß

$$\left(\sum\right) \int \frac{\partial^k L(x, \theta)}{\partial \theta_{i_1} \dots \partial \theta_{i_k}} dx = \frac{\partial^k}{\partial \theta_{i_1} \dots \partial \theta_{i_k}} \left(\sum\right) \int L(x, \theta) dx = 0.$$

5. Für alle $\theta_0 \in \Theta$ gibt es eine Konstante c_{θ_0} und eine messbare Funktion $g_{\theta_0} : \text{Supp}L \rightarrow \mathbb{R}_+$, sodaß

$$\left| \frac{\partial^3 \log L(x, \theta)}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \right| \leq g_{\theta_0}(x), \quad |\theta - \theta_0| < c_{\theta_0}$$

und

$$\mathbb{E}_{\theta_0} g_{\theta_0}(X_1) < \infty.$$

Wir definieren die *Informationsmatrix von Fisher* durch

$$I(\theta) = \left(\mathbb{E} \left[\frac{\partial \log L(X_1, \theta)}{\partial \theta_i} \frac{\partial \log L(X_1, \theta)}{\partial \theta_j} \right] \right)_{i,j=1,\dots,m}. \quad (1.4.4)$$

Satz 1.4.2 (asymptotische Normalverteiltheit von konsistenten ML-Schätzern $\hat{\theta}_n$, multivariater Fall $m > 1$). Es seien X_1, \dots, X_n unabhängig und identisch verteilt mit Likelihood-Funktion L , die den Regularitätsbedingungen 1-5 genügt. Sei $I(\theta)$ positiv definit für alle $\theta \in \Theta \subset \mathbb{R}^m$. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ eine Folge von schwach konsistenten Maximum-Likelihood-Schätzern für θ . Dann gilt:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} N(0, I^{-1}(\theta)).$$

Ohne Beweis; siehe den Beweis des Satzes 3.4.2, Vorlesungsskript Stochastik I.

Für unsere vergrößerte Hypothese $H_0 : p \in \{p(\theta), \theta \in \Theta\}$ stellen wir folgende, stückweise konstante, Likelihood-Funktion auf:

$$L(x, \theta) = p_j(\theta), \text{ falls } x \in (a_j, b_j].$$

Dann ist die Likelihood-Funktion der Stichprobe (x_1, \dots, x_n) gleich

$$\begin{aligned} L(x_1, \dots, x_n, \theta) &= \prod_{j=1}^r p_j(\theta)^{Z_j(x_1, \dots, x_n)} \\ \Rightarrow \log L(x_1, \dots, x_n, \theta) &= \sum_{j=1}^r Z_j(x_1, \dots, x_n) \cdot \log p_j(\theta). \\ \hat{\theta}_n = \hat{\theta}(x_1, \dots, x_n) &= \operatorname{argmax}_{\theta \in \Theta} \log L(x_1, \dots, x_n, \theta) \\ \Rightarrow \sum_{j=1}^r Z_j(x_1, \dots, x_n) \frac{\partial p_j(\theta)}{\partial \theta_i} \cdot \frac{1}{p_j(\theta)} &= 0, \quad i = 1, \dots, m. \end{aligned}$$

Aus $\sum_{j=1}^r p_j(\theta) = 1$ folgt

$$\sum_{j=1}^r \frac{\partial p_j(\theta)}{\partial \theta_i} = 0 \Rightarrow \sum_{j=1}^r \frac{Z_j(x_1, \dots, x_n) - np_j(\theta)}{p_j(\theta)} \cdot \frac{\partial p_j(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, m.$$

Lemma 1.4.4. Im obigen Fall gilt $I(\theta) = C^\top(\theta) \cdot C(\theta)$, wobei $C(\theta)$ eine $(r \times m)$ -Matrix mit Elementen

$$c_{ij}(\theta) = \frac{\partial p_i(\theta)}{\partial \theta_j} \cdot \frac{1}{\sqrt{p_i(\theta)}} \quad \text{ist.}$$

Beweis.

$$\begin{aligned} \mathbb{E}_0 \left[\frac{\partial \log L(X_1, \theta)}{\partial \theta_i} \cdot \frac{\partial \log L(X_1, \theta)}{\partial \theta_j} \right] &= \sum_{k=1}^r \frac{\partial \log p_k(\theta)}{\partial \theta_i} \cdot \frac{\partial \log p_k(\theta)}{\partial \theta_j} \cdot p_k(\theta) \\ &= \sum_{k=1}^r \frac{\partial p_k(\theta)}{\partial \theta_i} \frac{1}{p_k(\theta)} \cdot \frac{\partial p_k(\theta)}{\partial \theta_j} \cdot \frac{1}{p_k(\theta)} \cdot p_k(\theta) \\ &= \left(C^\top(\theta) \cdot C(\theta) \right)_{ij}, \end{aligned}$$

$$\text{denn } \log L(X_1, \theta) = \sum_{i=1}^r \log p_j(\theta) \cdot \mathbb{I}(x \in (a_j, b_j]).$$

□

Deshalb gilt die Folgerung aus Satz 1.4.2:

Folgerung 1.4.1. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ ein Maximum-Likelihood-Schätzer von θ im vergrößerten Modell, der schwach konsistent ist und den obigen Regularitätsbedingungen genügt. Sei die Informationsmatrix von Fisher $I(\theta) = C^\top(\theta) \cdot C(\theta)$ für alle $\theta \in \Theta$ positiv definit. Dann ist $\hat{\theta}$ asymptotisch normalverteilt:

$$\sqrt{n} \left(\hat{\theta}_n - \theta \right) \xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, I^{-1}(\theta))$$

Satz 1.4.3. Es sei $\hat{\theta}_n$ ein Maximum-Likelihood-Schätzer im vergrößerten Modell für θ , für den alle Voraussetzungen der Folgerung 1.4.1 erfüllt sind. Die Teststatistik

$$\hat{T}_n(X_1, \dots, X_n) = \sum_{j=1}^r \frac{(Z_j(X_1, \dots, X_n) - np_j(\hat{\theta}_n))^2}{np_j(\hat{\theta}_n)}$$

ist unter H_0 asymptotisch χ_{r-m-1}^2 -verteilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(\hat{T}_n(X_1, \dots, X_n) > \chi_{r-m-1, 1-\alpha}^2 \right) = \alpha.$$

ohne Beweis (siehe [15]).

Aus diesem Satz folgt, daß der χ^2 -Pearson-Fisher-Test ein asymptotischer Test zum Niveau α ist.

Beispiel 1.4.1. 1. χ^2 -Pearson-Fisher-Test der Normalverteilung

Sei (X_1, \dots, X_n) eine Zufallsstichprobe. Es soll geprüft werden, ob $X_i \sim N(\mu, \sigma^2)$. Es gilt

$$\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+.$$

Sei $(a_j, b_j]_{j=1, \dots, r}$ eine beliebige Aufteilung von \mathbb{R} in r disjunkte Intervalle. Sei

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

die Dichte der $N(\mu, \sigma^2)$ -Verteilung.

$$p_j(\theta) = \mathbb{P}_0(a_j < X_1 \leq b_j) = \int_{a_j}^{b_j} f_\theta(x) dx, \quad j = 1, \dots, r$$

mit den Klassenstärken

$$Z_j = \#\{i : X_i \in (a_j, b_j]\}.$$

Wir suchen den Maximum-Likelihood-Schätzer im vergrößerten Modell:

$$\begin{aligned} \frac{\partial p_j(\theta)}{\partial \mu} &= \int_{a_j}^{b_j} \frac{\partial}{\partial \mu} f_\theta(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{a_j}^{b_j} \frac{x-\mu}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ \frac{\partial p_j(\theta)}{\partial \sigma^2} &= \int_{a_j}^{b_j} \frac{\partial}{\partial \sigma^2} f_\theta(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{a_j}^{b_j} \left[-\frac{1}{2} \cdot \frac{1}{(\sigma^2)^{3/2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(\frac{(x-\mu)^2}{2(\sigma^2)^2} \right) \right] dx \\ &= -\frac{1}{2} \frac{1}{\sigma^2} \int_{a_j}^{b_j} f_\theta(x) dx + \frac{1}{2(\sigma^2)^2} \int_{a_j}^{b_j} (x-\mu)^2 f_\theta(x) dx \end{aligned}$$

Die notwendigen Bedingungen des Maximums sind:

$$\begin{aligned} \sum_{i=1}^r Z_j \frac{\int_{a_j}^{b_j} x f_\theta(x) dx}{\int_{a_j}^{b_j} f_\theta(x) dx} - \mu \underbrace{\sum_{j=1}^r Z_j}_{=n} &= 0, \\ \frac{1}{\sigma^2} \sum_{j=1}^r Z_j \frac{\int_{a_j}^{b_j} (x-\mu)^2 f_\theta(x) dx}{\int_{a_j}^{b_j} f_\theta(x) dx} - \underbrace{\sum_{j=1}^r Z_j}_{=n} &= 0. \end{aligned}$$

Daraus folgen die Maximum-Likelihood-Schätzer $\hat{\mu}$ und $\hat{\sigma}^2$ für μ und σ^2 :

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^r Z_j \frac{\int_{a_j}^{b_j} x f_{\theta}(x) dx}{\int_{a_j}^{b_j} f_{\theta}(x) dx},$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^r Z_j \frac{\int_{a_j}^{b_j} (x - \mu)^2 f_{\theta}(x) dx}{\int_{a_j}^{b_j} f_{\theta}(x) dx}.$$

Wir konstruieren eine Näherung zu $\hat{\mu}$ und $\hat{\sigma}^2$ für $r \rightarrow \infty$. Falls $r \rightarrow \infty$ (und somit auch $n \rightarrow \infty$), dann ist $b_j - a_j$ klein und nach der einfachen Quadraturregel gilt:

$$\int_{a_j}^{b_j} x f_{\theta}(x) dx \approx (b_j - a_j) y_j f_{\theta}(y_j),$$

$$\int_{a_j}^{b_j} f_{\theta}(x) dx \approx (b_j - a_j) f_{\theta}(y_j),$$

wobei $y_1 = b_1$, $y_r = b_{r-1} = a_r$,

$$y_j = (b_{j+1} + b_j)/2, \quad j = 2, \dots, r-1.$$

Daraus folgen für die Maximum-Likelihood-Schätzer $\hat{\mu}$ und $\hat{\sigma}^2$:

$$\hat{\mu} \approx \frac{1}{n} \sum_{j=1}^r y_j \cdot Z_j = \tilde{\mu}$$

$$\hat{\sigma}^2 \approx \frac{1}{n} \sum_{j=1}^r (y_j - \tilde{\mu})^2 Z_j = \tilde{\sigma}^2,$$

$$\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}^2).$$

Der χ^2 -Pearson-Fisher-Test lautet dann: H_0 wird abgelehnt, falls

$$\hat{T}_n = \frac{\sum_{j=1}^r \left(Z_j - np_j(\tilde{\theta}) \right)^2}{np_j(\tilde{\theta})} > \chi_{r-3, 1-\alpha}^2.$$

2. χ^2 -Pearson-Fisher-Test der Poissonverteilung

Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen. Wir wollen testen, ob $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Es gilt $\theta = \lambda$ und $\Theta = (0, +\infty)$. Die Vergrößerung von Θ hat die Form

$$-\infty = a_1 < \underbrace{b_1}_{=0} = a_2 < \underbrace{b_2}_{=1} = a_3 < \dots < \underbrace{b_{r-1}}_{=r-2} = a_r < b_r = +\infty.$$

Dann ist

$$\begin{aligned}
 p_j(\lambda) &= \mathbb{P}_\lambda(X_1 = j - 1) = e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!}, \quad j = 1, \dots, r-1, \\
 p_r(\lambda) &= \sum_{i=r-1}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!}, \\
 \frac{dp_j(\lambda)}{d\lambda} &= -e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!} + (j-1) \frac{\lambda^{j-2}}{(j-1)!} e^{-\lambda} = e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!} \left(\frac{j-1}{\lambda} - 1 \right) \\
 &= p_j(\lambda) \cdot \left(\frac{j-1}{\lambda} - 1 \right), \quad j = 1, \dots, r-1 \\
 \frac{dp_r(\lambda)}{d\lambda} &= \sum_{i \geq r-1} p_i(\lambda) \left(\frac{i-1}{\lambda} - 1 \right).
 \end{aligned}$$

Die Maximum-Likelihood-Gleichung lautet

$$0 = \sum_{j=1}^{r-1} Z_j \cdot \left(\frac{j-1}{\lambda} - 1 \right) + Z_r \frac{\sum_{i \geq r-1} p_i(\lambda) \left(\frac{i-1}{\lambda} - 1 \right)}{p_r(\lambda)}$$

Falls $r \rightarrow \infty$, so findet sich $r(n)$ für jedes n , für das $Z_{r(n)} = 0$. Deshalb gilt für $r > r(n)$:

$$\sum_{j=1}^{r-1} (j-1)Z_j - \lambda \underbrace{\sum_{j=1}^r Z_j}_{=n} = 0,$$

woraus der Maximum-Likelihood-Schätzer

$$\frac{1}{n} \sum_{j=1}^{r-1} (j-1)Z_j = \frac{1}{n} \sum_{j=1}^n X_j = \bar{X}_n$$

folgt. Der χ^2 -Pearson-Fisher-Test lautet: H_0 wird verworfen, falls

$$\hat{T}_n = \sum_{j=1}^r \frac{(Z_j - np_\lambda(\bar{X}_n))^2}{(np_j(\bar{X}_n))^2} > \chi_{r-2, 1-\alpha}^2.$$

1.4.3 Anpassungstest von Shapiro

Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen, $X_i \sim F$. Getestet werden soll die Hypothese

$$H_0 : F \in \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} \text{ vs. } H_1 : F \notin \{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

Die in den Abschnitten 1.4.1 - 1.4.2 vorgestellten χ^2 -Tests sind asymptotisch; deshalb können sie für relativ kleine Stichprobenumfänge nicht verwendet werden.

Der folgende Test wird diese Lücke füllen und eine Testentscheidung über H_0 selbst bei kleinen Stichproben ermöglichen.

Man bildet Ordnungsstatistiken $X_{(1)}, \dots, X_{(n)}$, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ und vergleicht ihre Korreliertheit mit den Mittelwerten der entsprechenden Ordnungsstatistiken der $N(0, 1)$ -Verteilung. Sei (Y_1, \dots, Y_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen, $Y_1 \sim N(0, 1)$. Es sei $a_i = \mathbb{E} Y_{(i)}$, $i = 1, \dots, n$. Falls der empirische Korrelationskoeffizient ρ_{aX} zwischen (a_1, \dots, a_n) und $(X_{(1)}, \dots, X_{(n)})$ bei 1 liegt, dann ist die Stichprobe normalverteilt. Formalisieren wir diese Heuristik:

Es sei b_i der Erwartungswert der i -ten Ordnungsstatistik in einer Stichprobe von $N(\mu, \sigma^2)$ -verteilten, unabhängigen Zufallsvariablen Z_i : $b_i = \mathbb{E} Z_{(i)}$, $i = 1, \dots, n$. Es gilt: $b_i = \mu + \sigma a_i$, $i = 1, \dots, n$. Betrachten wir den Korrelationskoeffizienten

$$\rho_{bX} = \frac{\sum_{i=1}^n (b_i - \bar{b}_n) (X_{(i)} - \bar{X}_n)}{\sqrt{\sum_{i=1}^n (b_i - \bar{b}_n)^2 \sum_{i=1}^n (X_{(i)} - \bar{X}_n)^2}}. \quad (1.4.5)$$

Da ρ invariant bezüglich Lineartransformationen ist und

$$\sum_{i=1}^n a_i = \sum_{i=1}^n \mathbb{E} Y_i = \mathbb{E} \left(\sum_{i=1}^n Y_i \right) = 0, \quad \text{gilt:}$$

$$\begin{aligned} \rho_{bX} &\stackrel{\text{(Stochastik I)}}{=} \rho_{aX} = \frac{\sum_{i=1}^n a_i (X_{(i)} - \bar{X}_n)}{\sqrt{\sum_{i=1}^n a_i^2 \sum_{i=1}^n (X_i - \bar{X}_n)^2}} = \frac{\sum_{i=1}^n a_i X_{(i)} - \bar{X}_n \overbrace{\sum_{i=1}^n a_i}^{=0}}{\sqrt{\sum_{i=1}^n a_i^2 \sum_{i=1}^n (X_i - \bar{X}_n)^2}} \\ &= \frac{\sum_{i=1}^n a_i X_{(i)}}{\sqrt{\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n (X_i - \bar{X}_n)^2}} \end{aligned}$$

Die Teststatistik lautet:

$$T_n = \frac{\sum_{i=1}^n a_i X_{(i)}}{\sqrt{\sum_{i=1}^n a_i^2 \sum_{i=1}^n (X_i - \bar{X}_n)^2}} \quad (\text{Shapiro-Francia-Test})$$

Die Werte a_i sind bekannt und können den Tabellen bzw. der Statistik-Software entnommen werden. Es gilt: $|T_n| \leq 1$.

H_0 wird abgelehnt, falls $T_n \leq q_{n,\alpha}$, wobei $q_{n,\alpha}$ das α -Quantil der Verteilung von T_n ist. Diese Quantile sind aus den Tabellen bekannt, bzw. können durch Monte-Carlo-Simulationen berechnet werden.

Bemerkung 1.4.3. Einen anderen, weit verbreiteten Test dieser Art bekommt man, wenn man die Lineartransformation $b_i = \mu + \sigma a_i$ durch eine andere Lineartransformation ersetzt:

$$(a'_1, \dots, a'_n)^\top = K^{-1} \cdot (a_1, \dots, a_n),$$

wobei $K = (k_{ij})_{j=1}^n$ die Kovarianzmatrix von $(Y_{(1)}, \dots, Y_{(n)})$ ist:

$$k_{ij} = \mathbb{E} (Y_{(i)} - a_i) (Y_{(j)} - a_j), \quad i, j = 1, \dots, n,$$

Der so konstruierte Test trägt den Namen *Shapiro-Wilk-Test*.

1.5 Weitere, nicht parametrische Tests

1.5.1 Binomialtest

Es sei (X_1, \dots, X_n) eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen, wobei $X_i \sim \text{Bernoulli}(p)$. Getestet werden soll:

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0$$

Die Teststatistik lautet

$$T_n = \sum_{i=1}^n X_i \underset{H_0}{\sim} \text{Bin}(n, p_0),$$

und die Entscheidungsregel ist: H_0 wird verworfen, falls

$$T_n \notin [\text{Bin}(n, p_0)_{\alpha/2}, \text{Bin}(n, p_0)_{1-\alpha/2}],$$

wobei $\text{Bin}(n, p)_\alpha$ das α -Quantil der $\text{Bin}(n, p)$ -Verteilung ist.

Für andere H_0 , wie zum Beispiel $p \leq p_0$ ($p \geq p_0$) muss der Ablehnungsbereich entsprechend angepasst werden.

Die Quantile $\text{Bin}(n, p)_\alpha$ erhält man aus Tabellen oder aus Monte-Carlo-Simulationen. Falls n groß ist, können diese Quantile durch die Normalapproximation berechnet werden:

Nach dem zentralen Grenzwertsatz von DeMoivre-Laplace gilt:

$$\mathbb{P}(T_n \leq x) = \mathbb{P}\left(\frac{T_n - np_0}{\sqrt{np_0(1-p_0)}} \leq \frac{x - np_0}{\sqrt{np_0(1-p_0)}}\right) \underset{n \rightarrow \infty}{\approx} \Phi\left(\frac{x - np_0}{\sqrt{np_0(1-p_0)}}\right).$$

Daraus folgt:

$$z_\alpha \approx \frac{\text{Bin}(n, p_0)_\alpha - np_0}{\sqrt{np_0(1-p_0)}}$$

$$\Rightarrow \text{Bin}(n, p_0)_\alpha \approx \sqrt{np_0(1-p_0)} \cdot z_\alpha + np_0$$

Nach der Poisson-Approximation (für $n \rightarrow \infty, np_0 \rightarrow \lambda_0$) gilt:

$$\text{Bin}(n, p_0)_{\alpha/2} \approx \text{Poisson}(\lambda_0)_{\alpha/2},$$

$$\text{Bin}(n, p_0)_{1-\alpha/2} \approx \text{Poisson}(\lambda_0)_{1-\alpha/2}, \quad \text{wobei } \lambda_0 = np_0.$$

Zielstellung: Wie kann mit Hilfe des oben beschriebenen Binomialtests die Symmetrieeigenschaft einer Verteilung getestet werden?

Es sei (Y_1, \dots, Y_n) eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen mit Verteilungsfunktion F . Getestet werden soll:

$$H_0 : F \text{ ist symmetrisch vs. } H_1 : F \text{ ist nicht symmetrisch.}$$

Eine symmetrische Verteilung besitzt den Median bei Null. Deswegen vergrößern wir die Hypothese H_0 und testen:

$$H'_0 : F^{-1}(0,5) = 0 \text{ vs. } H'_1 : F^{-1}(0,5) \neq 0.$$

Noch allgemeiner: Für ein $\beta \in [0, 1]$:

$$H''_0 : F^{-1}(\beta) = \gamma_\beta \text{ vs. } H''_1 : F^{-1}(\beta) \neq \gamma_\beta.$$

H''_0 vs. H''_1 wird mit Hilfe des Binomialtests wie folgt getestet: Sei $X_i = \mathbb{I}(Y_i \leq \gamma_\beta)$. Unter H''_0 gilt:

$$X_i \sim \text{Bernoulli}(F(\gamma_\beta)) = \text{Bernoulli}(\beta).$$

Seien $a_1 = -\infty, b_1 = \gamma_\alpha, a_2 = b_1, b_2 = +\infty$ zwei disjunkte Klassen $(a_1, b_1], (a_2, b_2]$ in der Sprache des χ^2 -Pearson-Tests. Die Testgröße ist:

$$T_n = \sum_{i=1}^n X_i = \# \{Y_i : Y_i \leq \gamma_\beta\} \sim \text{Bin}(n, \beta), \quad p = F(\gamma_\beta)$$

Die Hypothese $F^{-1}(\beta) = \gamma_\beta$ ist äquivalent zu $H'''_0 : p = \beta$. Die Entscheidungsregel lautet dann: H'''_0 wird verworfen, falls $T_n \notin [\text{Bin}(n, \beta)_{\alpha/2}, \text{Bin}(n, \beta)_{1-\alpha/2}]$. Dies ist ein Test zum Niveau α .

1.5.2 Iterationstests auf Zufälligkeit

In manchen Fragestellungen der Biologie untersucht man eine Folge von 0 oder 1 auf ihre „Zufälligkeit“ bzw. Vorhandensein von größeren Clustern von 0 oder 1. Diese Hypothesen kann man mit Hilfe der sogenannten *Iterationstests* statistisch überprüfen.

Sei eine Stichprobe $X_i, i = 1, \dots, n$ gegeben, $X_i \in \{0, 1\}$, $\sum_{i=1}^n X_i = n_1$ die Anzahl der Einsen, $n_2 = n - n_1$ die Anzahl der Nullen, n_1, n_2 vorgegeben. Eine Realisierung von (X_1, \dots, X_n) mit $n = 18, n_1 = 12$ wäre zum Beispiel

$$x = (0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1)$$

Es soll getestet werden, ob

H_0 : jede Folge x ist gleichwahrscheinlich vs.

H_1 : Es gibt bevorzugte Folgen (Clusterbildung)

stimmt.

Sei

$$\Omega = \left\{ x = (x_1, \dots, x_n) : x_i = 0 \text{ oder } 1, i = 1, \dots, n, \sum_{i=1}^n x_i = n_1 \right\}$$

der Stichprobenraum. Dann ist der Raum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathcal{F} = \mathcal{P}(\Omega)$,

$$\mathbb{P}(x) = \frac{1}{|\Omega|} = \frac{1}{\binom{n}{n_1}}$$

ein Laplacescher Wahrscheinlichkeitsraum.

Sei

$$\begin{aligned} T_n(X) &= \#\{\text{Iterationen in } X\} = \#\{\text{Teilfolgen der Nullen oder Einsen}\} \\ &= \#\{\text{Wechselstellen von 0 auf 1 oder von 1 auf 0}\} + 1. \end{aligned}$$

Zum Beispiel ist für $x = (0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0)$, $T_n(x) = 7 = 6 + 1$.

$T_n(X)$ wird folgendermaßen als Teststatistik für H_0 vs. H_1 benutzt. H_0 wird abgelehnt, falls $T(x)$ klein ist, das heißt, falls $T_n(x) < F_{T_n}^{-1}(\alpha)$. Dies ist ein Test zum Niveau α . Wie berechnen wir die Quantile $F_{T_n}^{-1}$?

Satz 1.5.1. Unter H_0 gelten folgende Aussagen:

1.

$$\mathbb{P}(T_n = k) = \begin{cases} \frac{2 \binom{n_1-1}{i-1} \binom{n_2-1}{i-1}}{\binom{n}{n_1}}, & \text{falls } k = 2i, \\ \frac{\binom{n_1-1}{i} \binom{n_2-1}{i-1} + \binom{n_1-1}{i-1} \binom{n_2-1}{i}}{\binom{n}{n_1}}, & \text{falls } k = 2i + 1. \end{cases}$$

2.

$$\mathbb{E} T_n = 1 + \frac{2n_1 n_2}{n}$$

3.

$$\text{Var}(T_n) = \frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n - 1)}$$

Beweis. 1. Wir nehmen an, daß $k = 2i$ (der ungerade Fall ist analog). Wie können i Klumpen von Einsen gewählt werden? Die Anzahl dieser Möglichkeiten = die Anzahl der Möglichkeiten, wie n_1 Teilchen auf i Klassen verteilt werden.

$$0|00|\dots|0|(n_1)$$

Dies ist gleich der Anzahl an Möglichkeiten, wie $i - 1$ Trennwände auf $n_1 - 1$ Positionen verteilt werden können = $\binom{n_1-1}{i-1}$. Das selbe gilt für die Nullen.

2. Sei $Y_j = \mathbb{I}\{X_{j-1} \neq X_j\}_{j=2,\dots,n}$.

$$\Rightarrow \mathbb{E} T_n(X) = 1 + \sum_{j=2}^n \mathbb{E} Y_j = 1 + \sum_{j=2}^n \mathbb{P}(X_{j-1} \neq X_j).$$

$$\begin{aligned} \mathbb{P}(X_{j-1} \neq X_j) &= \frac{2 \binom{n-2}{n_1-1}}{\binom{n}{n_1}} = 2 \cdot \frac{\frac{(n-2)!}{(n-2-(n_1-1))!(n_1-1)!}}{\frac{n!}{(n-n_1)!n_1!}} \\ &= \frac{2n_1(n-n_1)}{(n-1)n} \\ &= \frac{2n_1 n_2}{n(n-1)}. \end{aligned}$$

Daraus folgt

$$\mathbb{E} T_n = 1 + (n-1) \frac{2n_1 n_2}{n(n-1)} = 1 + 2 \frac{n_1 n_2}{n}.$$

3.

Übungsaufgabe 1.5.1. Beweisen Sie Punkt 3. □

Beispiel 1.5.1 (*Test von Wald-Wolfowitz*). Seien $Y = (Y_1, \dots, Y_n)$, $Z = (Z_1, \dots, Z_n)$ unabhängige Stichproben von unabhängigen und identisch verteilten Zufallsvariablen, $Y_i \sim F$, $Z_i \sim G$. Getestet werden soll:

$$H_0 : F = G \text{ vs. } H_1 : F \neq G.$$

Sei $(Y, Z) = (Y_1, \dots, Y_n, Z_1, \dots, Z_n)$ und seien X'_i Stichprobenvariablen von (Y, Z) , $i = 1, \dots, n$, $n = n_1 + n_2$. Wir bilden die Ordnungsstatistiken $X'_{(i)}$, $i = 1, \dots, n$ und setzen

$$X_i = \begin{cases} 1, & \text{falls } X'_{(i)} = Y_j \text{ für ein } j = 1, \dots, n_1, \\ 0, & \text{falls } X'_{(i)} = Z_j \text{ für ein } j = 1, \dots, n_2. \end{cases}$$

Unter H_0 sind die Stichprobenwerte in (Y, Z) gut gemischt, das heißt jede Kombination von 0 und 1 in (X_1, \dots, X_n) ist gleichwahrscheinlich. Darum können wir den Iterationstest auf Zufälligkeit anwenden, um H_0 vs. H_1 zu testen: H_0 wird verworfen, falls $T_n(x) \leq F^{-1}(\alpha)$, $x = (x_1, \dots, x_n)$.

Wie können die Quantile von F_{T_n} für große n berechnet werden? Falls

$$\frac{n_1}{n_1 + n_2} \xrightarrow{n \rightarrow \infty} p \in (0, 1),$$

dann ist T_n asymptotisch normalverteilt.

Satz 1.5.2. Unter der obigen Voraussetzung gilt:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mathbb{E} T_n}{n} &= 2p(1-p) \\ \lim_{n \rightarrow \infty} \frac{1}{n} \text{Var } T_n &= 4p^2(1-p)^2 \\ \frac{T_n - 2p(1-p)}{2\sqrt{np}(1-p)} &\xrightarrow[n \rightarrow \infty]{d} Y \sim N(0, 1), \quad \text{falls } \frac{n_1}{n_1 + n_2} \xrightarrow{n \rightarrow \infty} p \in (0, 1). \end{aligned}$$

So können Quantile von T_n näherungsweise für große n folgendermaßen berechnet werden:

$$\begin{aligned} \alpha &= \mathbb{P}(T_n \leq F_{T_n}^{-1}(\alpha)) = \mathbb{P}\left(\frac{T_n - 2np(1-p)}{2\sqrt{np}(1-p)} \leq \frac{x - 2np(1-p)}{2\sqrt{np}(1-p)}\right) \Bigg|_{x=F_{T_n}^{-1}(\alpha)} \\ &\approx \Phi\left(\frac{F_{T_n}^{-1}(\alpha) - 2np(1-p)}{2\sqrt{np}(1-p)}\right) \\ &\Rightarrow z_\alpha \approx \frac{F_{T_n}^{-1}(\alpha) - 2np(1-p)}{2\sqrt{np}(1-p)} \end{aligned}$$

Damit erhalten wir für die Quantile:

$$F_{T_n}^{-1}(\alpha) \approx 2np(1-p) + 2\sqrt{np}(1-p) \cdot z_\alpha$$

In der Praxis setzt man $\hat{p} = \frac{n_1}{n_1+n_2}$ für p ein.

2 Lineare Regression

In Stochastik I betrachteten wir die einfache lineare Regression der Form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

In Matrix-Form schreiben wir $Y = X\beta + \varepsilon$, wobei $Y = (Y_1, \dots, Y_n)^\top$ der Vektor der Zielzufallsvariablen ist,

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

eine $(n \times 2)$ -Matrix, die die Ausgangsvariablen $x_i, i = 1, \dots, n$ enthält und deshalb *Design-Matrix* genannt wird, $\beta = (\beta_0, \beta_1)^\top$ der Parametervektor und $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ der Vektor der Störgrößen. Bisher waren oft $\varepsilon_i \sim N(0, \sigma^2)$ für $i = 1, \dots, n$ und $\varepsilon \sim N(0, \mathcal{I} \cdot \sigma^2)$ multivariat normalverteilt.

Die multivariate (das bedeutet, nicht einfache) lineare Regression lässt eine beliebige $(n \times m)$ -Design-Matrix

$$X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$$

und einen m -dimensionalen Parametervektor $\beta = (\beta_1, \dots, \beta_m)^\top$ zu, für $m \geq 2$. Das heißt, es gilt

$$Y = X\beta + \varepsilon, \tag{2.0.1}$$

wobei $\varepsilon \sim N(0, K)$ ein multivariat normalverteilter Zufallsvektor der Störgrößen mit Kovarianzmatrix K ist, die im Allgemeinen nicht unabhängig voneinander sind:

$$K \neq \text{diag}(\sigma_1^2, \dots, \sigma_n^2).$$

Das Ziel dieses Kapitels ist es, Schätzer und Tests für β zu entwickeln. Zuvor müssen jedoch die Eigenschaften der multivariaten Normalverteilung untersucht werden.

2.1 Multivariate Normalverteilung

Im Vorlesungsskript Wahrscheinlichkeitsrechnung wurde die multivariate Normalverteilung in Beispiel 3.4.1 folgendermaßen eingeführt:

Definition 2.1.1. Es sei $X = (X_1, \dots, X_n)^\top$ ein n -dimensionaler Zufallsvektor, $\mu \in \mathbb{R}^n$, K eine symmetrische, positiv definite $(n \times n)$ -Matrix. X ist *multivariat normalverteilt* mit den Parametern μ und K ($X \sim N(\mu, K)$), falls X absolut stetig verteilt ist mit der Dichte

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(K)}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top K^{-1} (x - \mu) \right\}, \quad x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n.$$

Wir geben drei weitere Definitionen von $N(\mu, K)$ an und wollen die Zusammenhänge zwischen ihnen untersuchen:

Definition 2.1.2. Der Zufallsvektor $X = (X_1, \dots, X_n)^\top$ ist multivariat normalverteilt ($X \sim N(\mu, K)$) mit Parametern $\mu \in \mathbb{R}^n$ und K (eine symmetrische, nicht-negativ definite $(n \times n)$ -Matrix), falls die charakteristische Funktion $\varphi_X(t) = \mathbb{E} e^{i(t, X)}$, $t \in \mathbb{R}^n$, gegeben ist durch

$$\varphi_X(t) = \exp \left\{ it^\top \mu - \frac{1}{2} t^\top K t \right\}, \quad t \in \mathbb{R}^n.$$

Definition 2.1.3. Der Zufallsvektor $X = (X_1, \dots, X_n)^\top$ ist multivariat normalverteilt ($X \sim N(\mu, K)$) mit Parametern $\mu \in \mathbb{R}^n$ und einer symmetrischen, nicht negativ definiten $(n \times n)$ -Matrix K , falls

$$\forall a \in \mathbb{R}^n : \text{ die Zufallsvariable } (a, X) = a^\top X \sim N(a^\top \mu, a^\top K a)$$

eindimensional normalverteilt ist.

Definition 2.1.4. Es sei $\mu \in \mathbb{R}^n$, K eine nicht-negativ definite, symmetrische $(n \times n)$ -Matrix. Ein Zufallsvektor $X = (X_1, \dots, X_n)^\top$ ist multivariat normalverteilt mit Parametern μ und K ($X \sim N(\mu, K)$), falls

$$X \stackrel{d}{=} \mu + C \cdot Y,$$

wobei C eine $(n \times m)$ -Matrix mit $\text{rang}(C) = m$, $K = C \cdot C^\top$ und $Y \sim N(0, \mathcal{I}) \in \mathbb{R}^m$ ein m -dimensionaler Zufallsvektor mit unabhängigen und identisch verteilten Koordinaten $Y_j \sim N(0, 1)$ ist, $j = 1, \dots, m$.

Bemerkung: Dies ist das Analogon im eindimensionalen Fall: $Y \sim N(\mu, \sigma^2) \Leftrightarrow Y \stackrel{d}{=} \mu + \sigma X$ mit $X \sim N(0, 1)$.

Übungsaufgabe 2.1.1. Prüfen Sie, daß die in Definition 2.1.1 angegebene Dichte

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(K)}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top K^{-1} (x - \mu) \right\}, \quad x \in \mathbb{R}^n$$

tatsächlich eine Verteilungsdichte darstellt.

Lemma 2.1.1. Es seien X und Y n -dimensionale Zufallsvektoren mit charakteristischen Funktionen

$$\begin{aligned}\varphi_X(t) &= \mathbb{E} e^{i(t,X)} = \mathbb{E} e^{it^\top X} \\ \varphi_Y(t) &= \mathbb{E} e^{i(t,Y)} = \mathbb{E} e^{it^\top Y}\end{aligned}$$

für $t \in \mathbb{R}^n$. Es gelten folgende Eigenschaften:

1. *Eindeutigkeitssatz:*

$$X \stackrel{d}{=} Y \Leftrightarrow \varphi_X(t) = \varphi_Y(t), \quad t \in \mathbb{R}^n$$

2. Falls X und Y unabhängig sind, dann gilt:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t), \quad t \in \mathbb{R}^n.$$

ohne Beweis: vergleiche den Beweis des Satzes 5.1.1 (5), Folgerung 5.1.1, Vorlesungsskript WR.

Satz 2.1.1. 1. Die Definitionen 2.1.2 - 2.1.4 der multivariaten Normalverteilung sind äquivalent.

2. Die Definitionen 2.1.1 und 2.1.4 sind im Falle $n = m$ äquivalent.

Bemerkung 2.1.1. 1. Falls die Matrix K in Definition 2.1.4 den vollen Rang n besitzt, so besitzt sie die Dichte aus Definition 2.1.1. Sie wird in dem Fall *regulär* genannt.

2. Falls $\text{Rang}(K) = m < n$, dann ist die Verteilung $N(\mu, K)$ laut Definition 2.1.4 auf dem m -dimensionalen linearen Unterraum

$$\{y \in \mathbb{R}^n : y = \mu + Cx, x \in \mathbb{R}^m\}$$

konzentriert. $N(\mu, K)$ ist in diesem Fall offensichtlich nicht absolutstetig verteilt und wird daher *singulär* genannt.

Beweis. Wir beweisen: Definition 2.1.3 \Leftrightarrow 2.1.2 \Leftrightarrow 2.1.4.

1. a) Wir zeigen: Die Definitionen 2.1.2 und 2.1.3 sind äquivalent. Dazu ist zu zeigen: Für die Zufallsvariable X mit der charakteristischen Funktion

$$\varphi_X(t) = \exp\left\{it^\top \mu - \frac{1}{2}t^\top Kt\right\} \Leftrightarrow \forall a \in \mathbb{R}^n : a^\top X \sim N(a^\top \mu, a^\top K a).$$

Es gilt:

$$\varphi_{t^\top X}(1) = \mathbb{E} e^{it^\top X \cdot 1} \stackrel{\varphi_{N(\mu, \sigma^2)}}{=} \exp\left\{it^\top \mu - \frac{1}{2}t^\top Kt\right\} = \varphi_X(t) \quad \forall t \in \mathbb{R}.$$

(Dies nennt man das *Verfahren von Cramér-Wold*, vergleiche den multivariaten zentralen Grenzwertsatz).

- b) Wir zeigen: Die Definitionen 2.1.3 und 2.1.4 sind äquivalent. Dazu ist zu zeigen: $X = \mu + C \cdot Y$ (mit μ , C , und Y wie in Definition 2.1.4) $\Leftrightarrow \varphi_X(t) = \exp\{it^\top \mu - \frac{1}{2}t^\top Kt\}$, wobei $K = C \cdot C^\top$. Es gilt:

$$\begin{aligned} \varphi_{\mu+CY}(t) &= \mathbb{E} e^{i(t, \mu+CY)} = \mathbb{E} e^{it^\top \mu + it^\top CY} = e^{it^\top \mu} \cdot \mathbb{E} e^{i \overbrace{(C^\top t, Y)}^y} \\ &\stackrel{Y \sim N(0, \mathcal{I})}{=} e^{it^\top \mu} \cdot \exp\left(-\frac{1}{2}y^\top \cdot y\right) = \exp\left\{it^\top \mu - \frac{1}{2}t^\top C \cdot C^\top t\right\} \\ &= \exp\left\{it^\top \mu - \frac{1}{2}t^\top Kt\right\}, \quad t \in \mathbb{R}^n. \end{aligned}$$

2. Zu zeigen ist: Aus $X \sim N(\mu, K)$ im Sinne von Definition 2.1.4, $Y \sim N(\mu, K)$ im Sinne der Definition 2.1.1, $\text{Rang}(K) = n$ folgt, daß $\varphi_X = \varphi_Y$.

Aus der Definition 2.1.2 (die äquivalent zu Definition 2.1.4 ist) folgt, daß

$$\varphi_X(t) = \exp\left\{it^\top \mu - \frac{1}{2}t^\top Kt\right\}, \quad t \in \mathbb{R}^n,$$

$$\begin{aligned} \varphi_Y(t) &= \mathbb{E} e^{it^\top Y} = \int_{\mathbb{R}^n} e^{it^\top y} \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det K}} \cdot \exp\left\{-\frac{1}{2}\overbrace{(y-\mu)^\top}^x K^{-1} \overbrace{(y-\mu)}^x\right\} dy \\ &= e^{it^\top \mu} \cdot \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2} \sqrt{\det K}} \cdot \exp\left\{it^\top x - \frac{1}{2}x^\top K^{-1}x\right\} dx \end{aligned}$$

Wir diagonalisieren K : \exists orthogonale $(n \times n)$ -Matrix V : $V^\top = V^{-1}$ und $V^\top K V = \text{diag}(\lambda_1, \dots, \lambda_n)$, wobei $\lambda_i > 0$, $i = 1, \dots, n$. Mit der neuen Substitution: $x = Vz$, $t = Vs$ erhalten wir:

$$\begin{aligned} \varphi_Y(t) &= \frac{e^{it^\top \mu}}{(2\pi)^{n/2} \sqrt{\det K}} \cdot \int_{\mathbb{R}^n} \exp\left\{is^\top V^\top V z - \frac{1}{2}z^\top V^\top K^{-1}V z\right\} dz \\ &= \frac{e^{it^\top \mu}}{\sqrt{(2\pi)^n \lambda_1 \dots \lambda_n}} \cdot \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \exp\left\{is^\top z - \frac{1}{2} \sum_{i=1}^n \frac{z_i^2}{\lambda_i}\right\} dz_1 \dots dz_n \\ &= e^{it^\top \mu} \prod_{i=1}^n \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi \lambda_i}} e^{is_i z_i - \frac{z_i^2}{2\lambda_i}} dz_i = e^{it^\top \mu} \cdot \prod_{i=1}^n \varphi_{N(0, \lambda_i)}(s_i) = e^{it^\top \mu} \prod_{i=1}^n e^{-\frac{s_i^2 \lambda_i}{2}} \\ &= \exp\left\{it^\top \mu - \frac{1}{2}s^\top \text{diag}(\lambda_1, \dots, \lambda_n)s\right\} = \exp\left\{it^\top \mu - \frac{1}{2}(V^\top t)^\top V^\top K V V^\top t\right\} \\ &= \exp\left\{it^\top \mu - \frac{1}{2}t^\top \underbrace{V V^\top}_{\mathcal{I}} K \underbrace{V V^\top}_{\mathcal{I}} t\right\} = \exp\left\{it^\top \mu - \frac{1}{2}t^\top Kt\right\}, \quad t \in \mathbb{R}^n. \end{aligned}$$

□

2.1.1 Eigenschaften der multivariaten Normalverteilung

Satz 2.1.2. Es sei $X = (X_1, \dots, X_n) \sim N(\mu, K)$, $\mu \in \mathbb{R}^n$, K symmetrisch und nicht-negativ definit. Dann gelten folgende Eigenschaften:

1. μ ist der Erwartungswertvektor von X :

$$\mathbb{E} X = \mu, \quad \text{das heißt: } \mathbb{E} X_i = \mu_i, \quad i = 1, \dots, n.$$

K ist die Kovarianzmatrix von X :

$$K = (k_{ij}), \quad \text{mit } k_{ij} = \text{Cov}(X_i, X_j).$$

2. Jeder Teilvektor $X' = (X_{i_1}, \dots, X_{i_k})^\top$ ($1 \leq i_1 < \dots < i_k \leq n$) von X ist ebenso multivariat normalverteilt, $X' \sim N(\mu', K')$, wobei $\mu' = (\mu_{i_1}, \dots, \mu_{i_k})^\top$, $K' = (k'_{jl}) = (\text{Cov}(X_{i_j}, X_{i_l}))$, $j, l = 1, \dots, k$. Insbesondere sind $X_i \sim N(\mu_i, k_{ii})$, wobei $k_{ii} = \text{Var } X_i$, $i = 1, \dots, n$.
3. Zwei Teilvektoren von X sind unabhängig genau dann, wenn entsprechende Elemente k_{ij} von K , die ihre Kreuzkovarianzen darstellen, Null sind, das heißt: $X' = (X_1, \dots, X_k)^\top$, $X'' = (X_{k+1}, \dots, X_n)$ unabhängig (wobei die Reihenfolge nur wegen der Einfachheit so gewählt wurde, aber unerheblich ist) $\Leftrightarrow k_{ij} = 0$ für $1 \leq i \leq k$, $j > k$ oder $i > k$, $1 \leq j \leq k$.

$$K = \left(\begin{array}{c|c} K' & 0 \\ \hline 0 & K'' \end{array} \right)$$

K' und K'' sind Kovarianzmatrizen von X' bzw. X'' .

4. *Faltungsstabilität:* Falls X und Y unabhängige, n -dimensionale Zufallsvektoren mit $X \sim N(\mu_1, K_1)$ und $Y \sim N(\mu_2, K_2)$ sind, dann ist

$$X + Y \sim N(\mu_1 + \mu_2, K_1 + K_2).$$

Übungsaufgabe 2.1.2. Beweisen Sie Satz 2.1.2.

Satz 2.1.3 (*Lineare Transformation von $N(\mu, K)$*). Sei $X \sim N(\mu, K)$ ein n -dimensionaler Zufallsvektor, A eine $(m \times n)$ -Matrix mit $\text{Rang}(A) = m \leq n$, $b \in \mathbb{R}^m$. Dann ist der Zufallsvektor $Y = AX + b$ multivariat normalverteilt:

$$Y \sim N(A\mu + b, AK A^\top).$$

Beweis. Ohne Beschränkung der Allgemeinheit setzen wir $\mu = 0$ und $b = 0$, weil $\varphi_{Y-a}(t) = e^{-it^\top a} \cdot \varphi_Y(t)$, für $a = A\mu + b$. Es ist zu zeigen:

$$Y = AX, \quad X \sim N(0, K) \Rightarrow Y \sim N(0, AK A^\top)$$

Es ist

$$\begin{aligned}\varphi_Y(t) &= \varphi_{AX}(t) = \mathbb{E} e^{it^\top AX} = \mathbb{E} e^{i(X, \overbrace{A^\top t}^{:=s})} \\ &\stackrel{(\text{Def. 2.1.2})}{=} \exp \left\{ -\frac{1}{2} s^\top K s \right\} = \exp \left\{ -\frac{1}{2} t^\top A K A^\top t \right\}, t \in \mathbb{R}^n \\ &\Rightarrow Y \sim N(0, A K A^\top).\end{aligned}$$

□

2.1.2 Lineare und quadratische Formen von normalverteilten Zufallsvariablen

Definition 2.1.5. Seien $X = (X_1, \dots, X_n)^\top$ und $Y = (Y_1, \dots, Y_n)^\top$ Zufallsvektoren auf (Ω, \mathcal{F}, P) , A eine $(n \times n)$ -Matrix aus \mathbb{R}^{n^2} , die symmetrisch ist.

1. $Z = AX$ heißt *lineare Form* von X mit Matrix A .
2. $Z = Y^\top AX$ heißt *bilineare Form* von X und Y mit Matrix A ,

$$Z = \sum_{i=1}^n \sum_{j=1}^n a_{ij} X_j Y_i.$$

3. Die Zufallsvariable $Z = X^\top AX$ (die eine bilineare Form aus 2. mit $Y = X$ ist) heißt *quadratische Form* von X mit Matrix A .

Satz 2.1.4. Sei $Z = Y^\top AX$ eine bilineare Form von Zufallsvektoren $X, Y \in \mathbb{R}^n$ bzgl. der symmetrischen Matrix A . Falls $\mu_X = \mathbb{E} X$, $\mu_Y = \mathbb{E} Y$ und $K_{XY} = (\text{Cov}(X_i, Y_j))_{i,j=1,\dots,n}$ die Kreuzkovarianzmatrix von X und Y ist, dann gilt:

$$\mathbb{E} Z = \mu_Y^\top A \mu_X + \text{Spur}(A K_{XY}).$$

Beweis.

$$\begin{aligned}\mathbb{E} Z &= \mathbb{E} \text{Spur}(Z) = \mathbb{E} \text{Spur}(Y^\top AX) \quad (\text{wegen } \text{Spur}(AB) = \text{Spur}(BA)) \\ &= \mathbb{E} \text{Spur}(AXY^\top) = \text{Spur}(A \mathbb{E}(XY^\top)) \quad (\text{wobei } XY^\top = (X_i Y_j)_{i,j=1,\dots,n}) \\ &= \text{Spur} \left(A \mathbb{E} \left((X - \mu_X) \cdot (Y - \mu_Y)^\top + \mu_X Y^\top + X \mu_Y^\top - \mu_X \mu_Y^\top \right) \right) \\ &= \text{Spur} \left(A (K_{XY} + \mu_X \mu_Y^\top + \mu_X \mu_Y^\top - \mu_X \mu_Y^\top) \right) = \text{Spur} \left(A K_{XY} + A \mu_X \mu_Y^\top \right) \\ &= \text{Spur}(A K_{XY}) + \text{Spur} \left(A \mu_X \cdot \mu_Y^\top \right) \\ &= \text{Spur} \left(\mu_Y^\top A \mu_X \right) + \text{Spur}(A K_{XY}) = \mu_Y^\top A \mu_X + \text{Spur}(A K_{XY}).\end{aligned}$$

□

Folgerung 2.1.1. Für quadratische Formen gilt

$$\mathbb{E}(X^\top AX) = \mu_X^\top A \mu_X + \text{Spur}(A \cdot K),$$

wobei $\mu_X = \mathbb{E} X$ und K die Kovarianzmatrix von X ist.

Satz 2.1.5 (*Kovarianz quadratischer Formen*). Es sei $X \sim N(\mu, K)$ ein n -dimensionaler Zufallsvektor und $A, B \in \mathbb{R}^{n^2}$ zwei symmetrische $(n \times n)$ -Matrizen. Dann gilt Folgendes:

$$\text{Cov}\left(X^\top AX, X^\top BX\right) = 4\mu^\top AKB\mu + 2 \cdot \text{Spur}(AKBK).$$

Lemma 2.1.2 (*gemischte Momente*). Es sei $Y = (Y_1, \dots, Y_n)^\top \sim N(0, K)$ ein Zufallsvektor. Dann gilt Folgendes:

$$\begin{aligned} \mathbb{E}(Y_i Y_j Y_k) &= 0, \\ \mathbb{E}(Y_i Y_j Y_k Y_l) &= k_{ij} \cdot k_{kl} + k_{ik} \cdot k_{jl} + k_{jk} \cdot k_{il}, \quad 1 \leq i, j, k, l \leq n, \end{aligned}$$

wobei $K = (k_{ij})_{i,j=1,\dots,n}$ die Kovarianzmatrix von Y ist.

Übungsaufgabe 2.1.3. Beweisen Sie dieses Lemma.

folgt:

$$\begin{aligned} \text{Cov}\left(X^\top AX, X^\top BX\right) &= 2 \cdot \text{Spur}(AKBK) + \text{Spur}(AK) \cdot \text{Spur}(BK) + 4\mu^\top AKB\mu \\ &\quad - \text{Spur}(AK) \cdot \text{Spur}(BK) = 4\mu^\top AKB\mu + 2 \cdot \text{Spur}(AKBK). \end{aligned}$$

□

Folgerung 2.1.2.

$$\text{Var}\left(X^\top AX\right) = 4\mu^\top AKA\mu + 2 \cdot \text{Spur}\left((AK)^2\right)$$

Satz 2.1.6. Es seien $X \sim N(\mu, K)$ und $A, B \in \mathbb{R}^{n^2}$ zwei symmetrische Matrizen. Dann gilt:

$$\text{Cov}(BX, X^\top AX) = 2BKA\mu$$

Beweis.

$$\begin{aligned} \text{Cov}(BX, X^\top AX) &\stackrel{\text{(Folgerung 2.1.1)}}{=} \mathbb{E}\left[(BX - B\mu)(X^\top AX - \mu^\top A\mu - \text{Spur}(AK))\right] \\ &= \mathbb{E}\left[B(X - \mu)\left((X - \mu)^\top A(X - \mu) + 2\mu^\top AX - 2\mu^\top A\mu - \text{Spur}(AK)\right)\right], \end{aligned}$$

denn

$$(X - \mu)^\top A(X - \mu) = X^\top AX - \mu^\top AX - X^\top A\mu + \mu^\top A\mu$$

und mit der Substitution $Z = X - \mu$ (und damit $\mathbb{E}Z = 0$)

$$\begin{aligned} \text{Cov}(BX, X^\top AX) &= \mathbb{E}\left[BZ(Z^\top AZ + 2\mu^\top AZ - \text{Spur}(AK))\right] \\ &= \mathbb{E}(BZ \cdot Z^\top AZ) + 2\mathbb{E}(BZ \cdot \mu^\top AZ) - \text{Spur}(AK) \cdot \underbrace{\mathbb{E}(BZ)}_{=B\mathbb{E}Z=0} \\ &= 2\mathbb{E}(BZ \cdot Z^\top A\mu) + \mathbb{E}(BZZ^\top AZ) = 2B \underbrace{\mathbb{E}(ZZ^\top)}_{\text{Cov}X=K} A\mu \\ &\quad + B \cdot \underbrace{\mathbb{E}(ZZ^\top AZ)}_{=0} = 2BKA\mu, \end{aligned}$$

wegen $Z \sim N(0, K)$ und Lemma 2.1.2 und dem Beweis von Satz 2.1.5. □

Definition 2.1.6. Es seien $X_i \sim N(\mu_i, 1)$, $i = 1, \dots, n$ unabhängig. Dann besitzt die Zufallsvariable

$$Y = X_1^2 + \dots + X_n^2$$

die sogenannte *nicht-zentrale* $\chi_{n,\mu}^2$ -Verteilung mit n Freiheitsgraden und dem *Nichtzentralitätsparameter*

$$\mu = \sum_{i=1}^n \mu_i^2.$$

(in Stochastik I betrachteten wir den Spezialfall der zentralen χ_n^2 -Verteilung mit $\mu = 0$).

In Bemerkung 5.2.1, Vorlesungsskript WR, haben wir momenterzeugende Funktionen von Zufallsvariablen eingeführt. Jetzt benötigen wir für den Beweis des Satzes 2.1.7 folgenden Eindeutigkeitssatz:

Lemma 2.1.3 (*Eindeutigkeitssatz für momenterzeugende Funktionen*). Es seien X_1 und X_2 zwei absolutstetige Zufallsvariablen mit momenterzeugenden Funktionen

$$M_{X_i}(t) = \mathbb{E} e^{tX_i}, \quad i = 1, 2,$$

die auf einem Intervall (a, b) definiert sind. Falls f_1 und f_2 die Dichten der Verteilung von X_1 und X_2 sind, dann gilt

$$f_1(x) = f_2(x) \text{ für fast alle } x \in \mathbb{R} \Leftrightarrow M_{X_1}(t) = M_{X_2}(t), t \in (a, b).$$

Ohne Beweis.

Satz 2.1.7. Die Dichte einer $\chi_{n,\mu}^2$ -verteilten Zufallsvariable X (mit $n \in \mathbb{N}$ und $\mu > 0$) ist gegeben durch die Mischung der Dichten von χ_{n+2j}^2 -Verteilungen mit Mischungsvariable $J \sim \text{Poisson}(\mu/2)$:

$$f_X(x) = \begin{cases} \sum_{j=0}^{\infty} e^{-\mu/2} \frac{(\mu/2)^j}{j!} \cdot \frac{e^{-x/2} x^{\frac{n+2j}{2}-1}}{\Gamma(\frac{n+2j}{2}) \cdot 2^{\frac{n+2j}{2}}}, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (2.1.1)$$

Beweis. 1. Wir berechnen zuerst $M_X(t)$, $X \sim \chi_{n,\mu}^2$:

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) = \mathbb{E} \exp \left\{ t \sum_{i=1}^n X_i^2 \right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{tx_i^2} \cdot e^{-\frac{(x_i - \mu_i)^2}{2}} dx_i \quad \left(t < \frac{1}{2}, X_i \sim N(\mu_i, 1) \right) \end{aligned}$$

Es gilt:

$$\begin{aligned}
 tx_i^2 - \frac{(x_i - \mu_i)^2}{2} &= \frac{1}{2}(2tx_i^2 - x_i^2 + 2x_i\mu_i - \mu_i^2) \\
 &= -\frac{1}{2}\left(x_i^2(1-2t) - 2x_i\mu_i + \frac{\mu_i^2}{(1-2t)} - \frac{\mu_i^2}{(1-2t)} + \mu_i^2\right) \\
 &= -\frac{1}{2}\left(\left(x_i \cdot \sqrt{1-2t} - \frac{\mu_i}{\sqrt{1-2t}}\right)^2 + \mu_i^2\left(1 - \frac{1}{1-2t}\right)\right) \\
 &= -\frac{1}{2}\left(\frac{(x_i(1-2t) - \mu_i)^2}{1-2t} - \mu_i^2 \cdot \frac{2t}{1-2t}\right)
 \end{aligned}$$

Wir substituieren

$$y_i = \frac{(x_i \cdot (1-2t) - \mu_i)}{\sqrt{1-2t}}$$

und erhalten

$$\begin{aligned}
 M_X(t) &= (1-2t)^{-\frac{n}{2}} \prod_{i=1}^n \exp\left\{\mu_i^2 \cdot \left(\frac{t}{1-2t}\right)\right\} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y_i^2}{2}} dy_i}_{=1} \\
 &= (1-2t)^{-\frac{n}{2}} \cdot \exp\left\{\frac{t}{1-2t} \cdot \sum_{i=1}^n \mu_i^2\right\} = \frac{1}{(1-2t)^{n/2}} \cdot \exp\left\{\frac{\mu t}{1-2t}\right\}, \quad t < \frac{1}{2}.
 \end{aligned}$$

2. Es sei Y eine Zufallsvariable mit der Dichte (2.1.1). Wir berechnen $M_Y(t)$:

$$\begin{aligned}
 M_Y(t) &= \sum_{j=0}^{\infty} e^{-\frac{\mu}{2}} \frac{(\mu/2)^j}{j!} \cdot \underbrace{\int_0^{\infty} e^{xt} \cdot \frac{e^{-\frac{x}{2}} \cdot x^{\frac{n+2j}{2}-1}}{\Gamma\left(\frac{n+2j}{2}\right) \cdot \frac{n+2j}{2}} dx}_{=M_{\chi_{n+2j}^2}(t) = \frac{1}{(1-2t)^{(n+2j)/2}} \text{ (Stochastik I, Satz 3.2.1)}} \\
 &= \frac{e^{-\frac{\mu}{2}}}{(1-2t)^{\frac{n}{2}}} \cdot \sum_{j=1}^{\infty} \left(\frac{\mu}{2(1-2t)}\right)^j \cdot \frac{1}{j!} \\
 &= \frac{1}{(1-2t)^{\frac{n}{2}}} \cdot \exp\left\{-\frac{\mu}{2} + \frac{\mu}{2(1-2t)}\right\} = \frac{1}{(1-2t)^{\frac{n}{2}}} \cdot \exp\left\{\frac{\mu \cdot (1 - (1-2t))}{2 \cdot (1-2t)}\right\} \\
 &= (1-2t)^{-\frac{n}{2}} \cdot \exp\left\{\frac{\mu t}{1-2t}\right\} \\
 \implies M_X(t) &= M_Y(t), \quad t < \frac{1}{2}
 \end{aligned}$$

Nach Lemma 2.1.3 gilt dann, $f_X(x) = f_Y(x)$ für fast alle $x \in \mathbb{R}$.

□

Bemerkung 2.1.2. 1. Die Definition 2.1.6 kann in folgender Form umgeschrieben werden:

Falls $X \sim N(\vec{\mu}, \mathcal{I})$, $\vec{\mu} = (\mu_1, \dots, \mu_n)^\top$, dann gilt $|X|^2 = X^\top X \sim \chi_{n,\mu}^2$, wobei $\mu = |\vec{\mu}|^2$.

2. Die obige Eigenschaft kann auf $X \sim N(\vec{\mu}, K)$, mit einer symmetrischen, positiv definiten $(n \times n)$ -Matrix K verallgemeinert werden:

$$X^\top K^{-1} X \sim \chi_{n,\tilde{\mu}}^2, \quad \text{wobei } \tilde{\mu} = \vec{\mu}^\top K^{-1} \vec{\mu},$$

denn weil K positiv definit ist, gibt es ein $K^{\frac{1}{2}}$, sodaß $K = K^{\frac{1}{2}} K^{\frac{1}{2}\top}$. Dann gilt

$$Y = K^{-\frac{1}{2}} X \sim N(K^{-\frac{1}{2}} \mu, \mathcal{I}), \quad \text{weil } K^{-\frac{1}{2}} K K^{-\frac{1}{2}\top} = K^{-\frac{1}{2}} \cdot K^{\frac{1}{2}} \cdot K^{\frac{1}{2}\top} \cdot K^{-\frac{1}{2}\top} = \mathcal{I}$$

und daher

$$Y^\top Y \stackrel{\text{Punkt 1}}{\sim} \chi_{n,\tilde{\mu}}^2, \quad \text{mit } \tilde{\mu} = \left(K^{-\frac{1}{2}} \vec{\mu}\right)^\top K^{-\frac{1}{2}} \vec{\mu} = \vec{\mu}^\top K^{-\frac{1}{2}\top} K^{-\frac{1}{2}} \vec{\mu} = \vec{\mu}^\top K^{-1} \vec{\mu}.$$

Satz 2.1.8. Es sei $X \sim N(\mu, K)$, wobei K eine symmetrische, positiv definite $(n \times n)$ -Matrix ist, und sei A eine weitere symmetrische $(n \times n)$ -Matrix mit der Eigenschaft $AK = (AK)^2$ (Idempotenz) und $\text{Rang}(A) = r \leq n$. Dann gilt:

$$X^\top AX \sim \chi_{r,\tilde{\mu}}^2, \quad \text{wobei } \tilde{\mu} = \mu^\top A \mu.$$

Beweis. Wir zeigen, daß A nicht negativ definit ist.

$$\begin{aligned} AK &= (AK)^2 = AK \cdot AK \quad | \cdot K^{-1} \\ \implies A &= AK A \implies \forall x \in \mathbb{R}^n : x^\top Ax = x^\top AK Ax \\ &= \underbrace{(Ax)^\top}_{=y} K \underbrace{(Ax)}_{=y} \geq 0 \quad \text{wegen der positiven Definitheit von } K. \\ \implies A &\text{ ist nicht negativ definit.} \\ \implies \exists H &: \text{ eine } (n \times r)\text{-Matrix mit } \text{Rang}(H) = r : A = HH^\top \end{aligned}$$

Somit gilt

$$X^\top AX = X^\top H \cdot H^\top X = \underbrace{(H^\top X)^\top}_{=Y} \cdot H^\top X = Y^\top Y$$

Es gilt: $Y \sim N(H^\top \mu, \mathcal{I}_r)$, denn nach Satz 2.1.3 ist $Y \sim N(H^\top \mu, H^\top K H)$ und $\text{Rang}(H) = r$. Das heißt, $H^\top H$ ist eine invertierbare $(r \times r)$ -Matrix, und

$$\begin{aligned} H^\top K H &= (H^\top H)^{-1} \underbrace{(H^\top H \cdot H^\top K H \cdot (H^\top H))}_{=AKA=A} (H^\top H)^{-1} \\ &= (H^\top H)^{-1} H^\top \cdot \underbrace{A}_{=HH^\top} \cdot H (H^\top H)^{-1} \\ &= \mathcal{I}_r \end{aligned}$$

Dann ist

$$X^\top AX = |Y|^2 \sim \chi_{r, \tilde{\mu}}^2 \text{ mit } \tilde{\mu} = (H^\top \mu)^2 = \mu^\top H \cdot H^\top \mu = \mu^\top A \mu.$$

□

Satz 2.1.9 (Unabhängigkeit). Es sei $X \sim N(\mu, K)$ und K eine symmetrische, nicht-negativ definite $(n \times n)$ -Matrix.

1. Es seien A, B ($r_1 \times n$) bzw. ($r_2 \times n$)-Matrizen, $r_1, r_2 \leq n$ mit $AKB^\top = 0$. Dann sind die Vektoren AX und BX unabhängig.
2. Sei ferner C eine symmetrische, nicht-negativ definite $(n \times n)$ -Matrix mit der Eigenschaft $AKC = 0$. Dann sind AX und $X^\top CX$ unabhängig.

Beweis. 1. Nach Satz 2.1.2, 3) gilt: AX und BX sind unabhängig $\iff \varphi_{(AX, BX)}(t) = \varphi_{AX}(t) \cdot \varphi_{BX}(t)$, $t = (t_1, t_2)^\top \in \mathbb{R}^{r_1+r_2}$, $t_1 \in \mathbb{R}^{r_1}$, $t_2 \in \mathbb{R}^{r_2}$. Es ist zu zeigen:

$$\varphi_{(AX, BX)}(t) = \mathbb{E} e^{i(t_1^\top A + t_2^\top B) \cdot X} \stackrel{!}{=} \mathbb{E} e^{i t_1^\top AX} \cdot \mathbb{E} e^{i t_2^\top BX}.$$

Es gilt

$$\varphi_{(AX, BX)}(t) = \mathbb{E} e^{i(t_1^\top A + t_2^\top B) \cdot X} \stackrel{(Def. 2.1.2)}{=} e^{i(t_1^\top A + t_2^\top B) \cdot \mu - \frac{1}{2} \cdot (t_1^\top A + t_2^\top B) \cdot K \cdot (t_1^\top A + t_2^\top B)^\top},$$

und mit

$$\begin{aligned} & (t_1^\top A + t_2^\top B) \cdot K \cdot (t_1^\top A + t_2^\top B)^\top \\ &= (t_1^\top A) K (t_1^\top A)^\top + (t_1^\top A)^\top K (t_2^\top B) + (t_2^\top B) K (t_1^\top A)^\top + (t_2^\top B) K (t_2^\top B)^\top \\ &= t_1^\top A K A^\top t_1 + t_1^\top \cdot \underbrace{AKB^\top}_{=0} \cdot t_2 + t_2^\top \cdot \underbrace{BK A^\top}_{=(AKB^\top)^\top=0} \cdot t_1 + t_2^\top B K B^\top t_2 \end{aligned}$$

ist

$$\begin{aligned} \varphi_{(AX, BX)}(t) &= e^{i t_1^\top A - \frac{1}{2} t_1^\top A K A^\top t_1} \cdot e^{i t_2^\top B - \frac{1}{2} t_2^\top B K B^\top t_2} \\ &= \varphi_{AX}(t_1) \cdot \varphi_{BX}(t_2), \quad t_1 \in \mathbb{R}^{r_1}, t_2 \in \mathbb{R}^{r_2} \end{aligned}$$

2. C ist symmetrisch, nicht-negativ definit \implies Es gibt eine $(n \times r)$ -Matrix H mit $\text{Rang}(H) = r \leq n$ und $C = HH^\top$, $\implies H^\top H$ hat Rang r und ist somit invertierbar. Dann gilt:

$$X^\top CX = X^\top HH^\top X = (H^\top X)^\top \cdot H^\top X = |H^\top X|^2.$$

Falls AX und $H^\top X$ unabhängig sind, dann sind auch AX und $X^\top CX = |H^\top X|^2$ unabhängig, nach dem Transformationssatz für Zufallsvektoren. Nach 1) sind AX und $H^\top X$ unabhängig, falls $AK(H^\top)^\top = AKH = 0$. Da nach Voraussetzung

$$AKC = AKH \cdot H^\top = 0 \implies AKH \cdot H^\top H = 0,$$

da aber $\exists(H^\top H)^{-1}$, folgt, daß

$$\begin{aligned} 0 &= AKH \cdot H^\top H \cdot (H^\top H)^{-1} = AKH \implies AKH = 0 \\ &\implies AX \text{ und } H^\top X \text{ sind unabhängig} \\ &\implies AX \text{ und } X^\top CX \text{ sind unabhängig.} \end{aligned}$$

□

2.2 Multivariate lineare Regressionsmodelle mit vollem Rang

Die *multivariate lineare Regression* hat die Form

$$Y = X\beta + \varepsilon,$$

wobei $Y = (Y_1, \dots, Y_n)^\top$ der Zufallsvektor der Zielvariablen ist,

$$X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$$

ist eine deterministische *Design-Matrix* mit vollem Rang, $\text{Rang}(X) = r = m \leq n$, $\beta = (\beta_1, \dots, \beta_m)^\top$ ist der *Parametervektor* und $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ ist der Zufallsvektor der *Störgrößen*, mit $\mathbb{E}\varepsilon_i = 0$, $\text{Var}\varepsilon_i = \sigma^2 > 0$. Das Ziel dieses Abschnittes wird sein, β und σ^2 geeignet zu schätzen.

2.2.1 Methode der kleinsten Quadrate

Sei $X = (X_1, \dots, X_m)$, wobei die deterministischen Vektoren $X_j = (x_{1j}, x_{2j}, \dots, x_{nj})^\top$, $j = 1, \dots, m$ einen m -dimensionalen linearen Unterraum $L_X = \langle X_1, \dots, X_m \rangle$ aufspannen. Sei

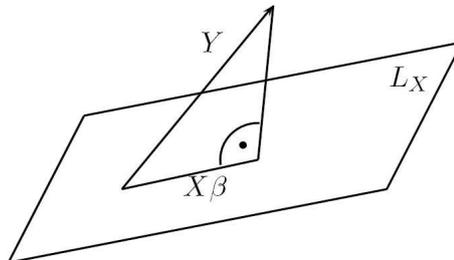
$$e(\beta) = \frac{1}{n} \|Y - X\beta\|^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - x_{i1}\beta_1 - \dots - x_{im}\beta_m)^2$$

die mittlere quadratische Abweichung zwischen Y und $X\beta$.

Der *MKQ-Schätzer* $\hat{\beta}$ für β ist definiert durch

$$\hat{\beta} = \text{argmin}(e(\beta)). \quad (2.2.1)$$

Warum existiert eine Lösung $\beta \in \mathbb{R}^m$ des quadratischen Optimierungsproblems (2.2.1)? Geometrisch kann $X\hat{\beta}$ als die orthogonale Projektion des Datenvektors Y auf den linearen Unterraum L_X interpretiert werden. Formal zeigen wir die Existenz der Lösung mit folgendem Satz.

Abbildung 2.1: Projektion auf den linearen Unterraum L_X 

Satz 2.2.1. Unter den obigen Voraussetzungen existiert der eindeutig bestimmte MKQ-Schätzer $\hat{\beta}$, der die Lösung der sogenannten *Normalgleichung* ist:

$$X^T X \beta = X^T Y. \quad (2.2.2)$$

Daher gilt:

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Beweis. Die notwendige Bedingung für die Existenz des Minimums ist $e'(\beta) = 0$, das heißt

$$e'(\beta) = \left(\frac{\partial e(\beta)}{\partial \beta_1}, \dots, \frac{\partial e(\beta)}{\partial \beta_m} \right)^T = 0.$$

Es gilt:

$$e'(\beta) = \frac{2}{n} (X^T X \beta - X^T Y)$$

$\implies \hat{\beta}$ ist eine Lösung der Normalgleichung $X^T X \beta = X^T Y$. Wir zeigen die hinreichende Bedingung des Minimums:

$$e''(\beta) = \left(\frac{\partial^2 e(\beta)}{\partial \beta_i \partial \beta_j} \right)_{i,j=1,\dots,m} = \frac{2}{n} X^T X.$$

$X^T X$ ist symmetrisch und positiv definit, weil X einen vollen Rang hat:

$$\forall y \neq 0, y \in \mathbb{R}^m : y^T X^T X y = (Xy)^T Xy = |Xy|^2 > 0$$

und aus $y \neq 0 \implies Xy \neq 0$, folgt, daß $e''(\beta)$ positiv definit ist. Also ist $X^\top X$ invertierbar. Das heißt, $\hat{\beta}$ ist der Minimumpunkt von $e(\beta)$. Den Schätzer $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ bekommt man, indem man die Normalengleichung $X^\top X\beta = X^\top Y$ von links mit $(X^\top X)^{-1}$ multipliziert. \square

Beispiel 2.2.1. 1. *Einfache lineare Regression*

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad m = 2, \beta = (\beta_1, \beta_2)^\top, Y = X\beta + \varepsilon$$

$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$ ergibt den MKQ-Schätzer aus der Stochastik I

$$\hat{\beta}_2 = \frac{S_{XY}^2}{S_{XX}^2}, \quad \hat{\beta}_1 = \bar{Y}_n - \bar{X}_n \hat{\beta}_2,$$

wobei

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i, & \bar{Y}_n &= \frac{1}{n} \sum_{i=1}^n Y_i \\ S_{XY}^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n) \\ S_{XX}^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \end{aligned}$$

Übungsaufgabe 2.2.1. Beweisen Sie dies!

2. *Multiple lineare Regression*

$Y = X\beta + \varepsilon$ mit Designmatrix

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix} \quad \text{für } \beta = (\beta_0, \beta_1, \dots, \beta_m)^\top.$$

Der MKQ-Schätzer $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ ist offensichtlich ein linearer Schätzer bezüglich Y .

Wir werden jetzt zeigen, daß $\hat{\beta}$ der *beste lineare, erwartungstreue Schätzer* von β (im Englischen *BLUE = best linear unbiased estimator*) in der Klasse

$$\mathcal{L} = \left\{ \tilde{\beta} = AY + b : \mathbb{E} \tilde{\beta} = \beta \right\}$$

aller linearen erwartungstreuen Schätzer ist.

Satz 2.2.2 (*Güteeigenschaften des MKQ-Schätzers $\hat{\beta}$*). Es sei $Y = X\beta + \varepsilon$ ein multivariates lineares Regressionsmodell mit vollem Rang m und Störgrößen $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$, die folgende Voraussetzungen erfüllen:

$$\mathbb{E}\varepsilon = 0, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{ij}, \quad i, j = 1, \dots, n \text{ für ein } \sigma^2 \in (0, \infty).$$

Dann gilt Folgendes:

1. Der MKQ-Schätzer $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ ist erwartungstreu: $\mathbb{E}\hat{\beta} = \beta$.
2. $\text{Cov}(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}$
3. $\hat{\beta}$ besitzt die minimale Varianz:

$$\forall \tilde{\beta} \in \mathcal{L}: \quad \text{Var} \tilde{\beta}_j \geq \text{Var} \hat{\beta}_j, \quad j = 1, \dots, m.$$

Beweis. 1. Es gilt:

$$\begin{aligned} \mathbb{E}\hat{\beta} &= \mathbb{E} \left[(X^\top X)^{-1} X^\top (X\beta + \varepsilon) \right] \\ &= (X^\top X)^{-1} \cdot X^\top X \cdot \beta + (X^\top X)^{-1} X^\top \cdot \underbrace{\mathbb{E}\varepsilon}_{=0} \\ &= \beta \quad \forall \beta \in \mathbb{R}^m. \end{aligned}$$

2. Für alle $\tilde{\beta} = AY + b \in \mathcal{L}$ gilt:

$$\begin{aligned} \beta &= \mathbb{E}\tilde{\beta} = A\mathbb{E}Y + b = AX\beta + b \quad \forall \beta \in \mathbb{R}^m. \\ &\implies b = 0, \quad AX = \mathcal{I}. \\ &\implies \tilde{\beta} = AY = A(X\beta + \varepsilon) = AX\beta + A\varepsilon \\ &= \beta + A\varepsilon. \end{aligned}$$

Für

$$\hat{\beta} = \underbrace{(X^\top X)^{-1} X^\top}_=A Y$$

gilt:

$$\begin{aligned} \text{Cov}\hat{\beta} &= \left(\mathbb{E} \left(\left(\hat{\beta}_i - \beta_i \right) \left(\hat{\beta}_j - \beta_j \right) \right) \right)_{i,j=1,\dots,m} \\ &= \mathbb{E} \left(A\varepsilon \cdot (A\varepsilon)^\top \right) = \mathbb{E} \left(A\varepsilon\varepsilon^\top A^\top \right) = A\mathbb{E} \left(\varepsilon\varepsilon^\top \right) \cdot A^\top \\ &= A \cdot \sigma^2 \mathcal{I} A^\top = \sigma^2 A A^\top = \sigma^2 (X^\top X)^{-1} X^\top \left((X^\top X)^{-1} X^\top \right)^\top \\ &= \sigma^2 (X^\top X)^{-1} X^\top X (X^\top X)^{-1} = \sigma^2 (X^\top X)^{-1}. \end{aligned}$$

3. Sei $\tilde{\beta} \in \mathcal{L}$, $\tilde{\beta} = \beta + A\varepsilon$. Zu zeigen ist, daß

$$\left(\text{Cov}(\tilde{\beta})\right)_{ii} = \sigma^2(AA^\top)_{ii} \geq \left(\text{Cov}(\hat{\beta})\right)_{ii} = \sigma^2(X^\top X)_{ii}^{-1}, \quad i = 1, \dots, m.$$

Sei $D = A - (X^\top X)^{-1}X^\top$, dann folgt: $A = D + (X^\top X)^{-1}X^\top$,

$$\begin{aligned} AA^\top &= \left(D + (X^\top X)^{-1}X^\top\right) \left(D^\top + X(X^\top X)^{-1\top}\right) \\ &= DD^\top + (X^\top X)^{-1}, \text{ weil} \end{aligned}$$

$$\begin{aligned} DX(X^\top X)^{-1} &= \underbrace{(AX)}_{=I} - \underbrace{(X^\top X)^{-1}X^\top X}_{=I} (X^\top X)^{-1} = 0 \\ (X^\top X)^{-1}X^\top D^\top &= (X^\top X)^{-1}X^\top \left(A^\top - X(X^\top X)^{-1\top}\right) \\ &= (X^\top X)^{-1} \left(\underbrace{(AX)^\top}_{=I} - \underbrace{X^\top X(X^\top X)^{-1}}_{=I}\right) = 0. \end{aligned}$$

$$\begin{aligned} \implies (AA^\top)_{ii} &= \underbrace{(DD^\top)_{ii}}_{\geq 0} + (X^\top X)_{ii}^{-1} \geq (X^\top X)_{ii}^{-1} \\ \implies \text{Var } \hat{\beta}_i &\leq \text{Var } \tilde{\beta}_i, \quad i = 1, \dots, m. \end{aligned}$$

□

Satz 2.2.3. Es sei $\hat{\beta}_n$ der MKQ-Schätzer im oben eingeführten multivariaten linearen Regressionsmodell. Sei $\{a_n\}_{n \in \mathbb{N}}$ eine Zahlenfolge mit $a_n \neq 0$, $n \in \mathbb{N}$, $a_n \rightarrow 0$ ($n \rightarrow \infty$). Es wird vorausgesetzt, daß eine invertierbare $(m \times m)$ -Matrix Q existiert mit

$$Q = \lim_{n \rightarrow \infty} a_n (X_n^\top X_n).$$

Dann ist $\hat{\beta}_n$ schwach konsistent:

$$\hat{\beta}_n \xrightarrow[n \rightarrow \infty]{p} \beta.$$

Beweis.

$$\hat{\beta}_n \xrightarrow[n \rightarrow \infty]{p} \beta \iff \mathbb{P}\left(\left|\hat{\beta}_n - \beta\right| > \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 0 \quad \forall \varepsilon > 0.$$

$$\begin{aligned}
\mathbb{P}\left(\left|\hat{\beta}_n - \beta\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\hat{\beta}_n - \beta\right|^2 > \varepsilon^2\right) \\
&= \mathbb{P}\left(\sum_{i=1}^m \left|\hat{\beta}_{in} - \beta_i\right|^2 > \varepsilon^2\right) \leq \mathbb{P}\left(\bigcup_{i=1}^m \left\{\left|\hat{\beta}_{in} - \beta_i\right|^2 > \frac{\varepsilon^2}{m}\right\}\right) \\
&\leq \sum_{i=1}^m \mathbb{P}\left(\left|\hat{\beta}_{in} - \beta_i\right| > \frac{\varepsilon}{\sqrt{m}}\right) \\
&\leq m \sum_{i=1}^m \frac{\text{Var} \hat{\beta}_{in}}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0, \quad (\text{aus der Ungleichung von Tschebyschew}) \\
&\text{falls } \text{Var} \hat{\beta}_{in} \xrightarrow{n \rightarrow \infty} 0, \quad i = 1, \dots, m.
\end{aligned}$$

$\text{Var} \hat{\beta}_{in}$ ist ein Diagonaleintrag von der Matrix

$$\text{Cov} \hat{\beta}_n \stackrel{(\text{Satz 2.2.2})}{=} \sigma^2 \left(X_n^\top X_n\right)^{-1}.$$

Wenn wir zeigen, daß $\text{Cov} \hat{\beta}_n \xrightarrow{n \rightarrow \infty} 0$, ist der Satz bewiesen.

Es existiert

$$Q^{-1} = \lim_{n \rightarrow \infty} \frac{1}{a_n} \left(X_n^\top X_n\right)^{-1}$$

und damit gilt:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \text{Cov} \hat{\beta}_n &= \sigma^2 \lim_{n \rightarrow \infty} \left(X_n^\top X_n\right)^{-1} = \sigma^2 \lim_{n \rightarrow \infty} a_n \cdot \frac{1}{a_n} \left(X_n^\top X_n\right)^{-1} \\
&= 0 \cdot Q^{-1} \cdot \sigma^2 = 0.
\end{aligned}$$

□

2.2.2 Schätzer der Varianz σ^2

Wir führen den Schätzer $\hat{\sigma}^2$ für die Varianz σ^2 der Störgrößen ε_i folgendermaßen ein:

$$\hat{\sigma}^2 = \frac{1}{n-m} \left|Y - X\hat{\beta}\right|^2. \quad (2.2.3)$$

Dies ist eine verallgemeinerte Version des Varianzschätzers aus der einfachen linearen Regression, die wir bereits in Stochastik I kennenlernten. Dabei ist $\hat{Y} = Y - X\hat{\beta}$ der Vektor der Residuen.

Satz 2.2.4 (*Erwartungstreue*). Der Varianzschätzer

$$\hat{\sigma}^2 = \frac{1}{n-m} \left|Y - X\hat{\beta}\right|^2$$

ist erwartungstreu. Das heißt,

$$\mathbb{E} \hat{\sigma}^2 = \sigma^2.$$

Beweis.

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-m} (Y - X\hat{\beta})^\top (Y - X\hat{\beta}) \\ &= \frac{1}{n-m} (Y - X(X^\top X)^{-1}X^\top Y)^\top (Y - X(X^\top X)^{-1}X^\top Y) \\ &= \frac{1}{n-m} (DY)^\top DY\end{aligned}$$

wobei $D = \mathcal{I} - X(X^\top X)^{-1}X^\top$ eine $(n \times n)$ -Matrix ist. Dann ist

$$\hat{\sigma}^2 = \frac{1}{n-m} Y^\top D^\top DY = \frac{1}{n-m} Y^\top D^2 Y = \frac{1}{n-m} Y^\top DY, \text{ falls}$$

$D^\top = D$ und $D^2 = D$ (das heißt, daß D symmetrisch und idempotent ist). Tatsächlich gilt:

$$\begin{aligned}D^\top &= \mathcal{I} - (X^\top)^\top (X^\top X)^{\top -1} X^\top = \mathcal{I} - X (X^\top X)^{-1} X^\top = D. \\ D^2 &= (\mathcal{I} - X(X^\top X)^{-1}X^\top) (\mathcal{I} - X(X^\top X)^{-1}X^\top) \\ &= \mathcal{I} - 2X(X^\top X)^{-1}X^\top + X(X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top \\ &= \mathcal{I} - X(X^\top X)^{-1}X^\top = D.\end{aligned}$$

Weiterhin gilt:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-m} \cdot \text{Spur} (Y^\top DY) = \frac{1}{n-m} \cdot \text{Spur} (DYY^\top) \\ \implies \mathbb{E} \hat{\sigma}^2 &= \frac{1}{n-m} \cdot \text{Spur} (D \mathbb{E} (YY^\top)) = \frac{\sigma^2}{n-m} \cdot \text{Spur} (D),\end{aligned}$$

denn

$$\begin{aligned}\text{Spur} (D \cdot \mathbb{E} (YY^\top)) &= \text{Spur} (D(X\beta)(X\beta)^\top + \underbrace{DX\beta \mathbb{E} \varepsilon^\top}_{=0} + \underbrace{D \mathbb{E} \varepsilon (X\beta)^\top}_{=0} + \underbrace{D \cdot \mathbb{E} \varepsilon \varepsilon^\top}_{= \text{Cov} \varepsilon = \sigma^2 \cdot \mathcal{I}})\end{aligned}$$

und

$$\begin{aligned}DX &= (\mathcal{I} - X(X^\top X)^{-1}X^\top) X \\ &= X - X(X^\top X)^{-1}X^\top X = X - X = 0.\end{aligned}$$

Es bleibt zu zeigen, daß $\text{Spur}(D) = n - m$:

$$\begin{aligned} \text{Spur}(D) &= \text{Spur} \left(\mathcal{I} - X \left(X^\top X \right)^{-1} X^\top \right) = \text{Spur}(\mathcal{I}) - \text{Spur} \left(X \left(X^\top X \right)^{-1} X^\top \right) \\ &= n - \text{Spur} \left(\underbrace{X^\top X \cdot \left(X^\top X \right)^{-1}}_{\text{eine } (m \times m)\text{-Matrix}} \right) = n - m. \end{aligned}$$

□

2.2.3 Maximum-Likelihood-Schätzer für β und σ^2

Um Maximum-Likelihood-Schätzer für β und σ^2 bzw. Verteilungseigenschaften der MKQ-Schätzer $\hat{\beta}$ und $\hat{\sigma}^2$ herleiten zu können, muß die Verteilung von ε bzw. Y präzisiert werden. Wir werden ab sofort normalverteilte Störgrößen betrachten, die unabhängig und identisch verteilt sind:

$$\varepsilon \sim N(0, \sigma^2 \mathcal{I}), \quad \sigma^2 > 0.$$

Daraus folgt:

$$Y \sim N(X\beta, \sigma^2 \mathcal{I}).$$

Wie sieht die Verteilung der MKQ-Schätzer $\hat{\beta}$ und $\hat{\sigma}^2$ aus? Da $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ linear von Y abhängt, erwartungstreu ist und die $\text{Cov}\hat{\beta} = \hat{\sigma}^2 (X^\top X)^{-1}$ besitzt, gilt:

$$\hat{\beta} \sim N \left(\beta, \sigma^2 \left(X^\top X \right)^{-1} \right)$$

Berechnen wir nun Maximum-Likelihood-Schätzer für β und σ^2 , und zwar $\tilde{\beta}$ und $\tilde{\sigma}^2$. Dann zeigen wir, daß sie im Wesentlichen mit den MKQ-Schätzern übereinstimmen.

$$\begin{aligned} \tilde{\beta} &= \hat{\beta}, \\ \tilde{\sigma}^2 &= \frac{n-m}{n} \hat{\sigma}^2. \end{aligned}$$

Betrachten wir zunächst die Likelihood-Funktion von Y :

$$L(y, \beta, \sigma^2) = f_Y(y) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \cdot \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)^\top (y - X\beta) \right\}$$

und die Log-Likelihood-Funktion

$$\log L(y, \beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \underbrace{\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} |y - X\beta|^2}_{:=g}.$$

Die Maximum-Likelihood-Schätzer sind dann

$$\left(\tilde{\beta}, \tilde{\sigma}^2 \right) = \underset{\beta \in \mathbb{R}^m, \sigma^2 > 0}{\text{argmax}} \log L(y, \beta, \sigma^2),$$

sofern sie existieren.

Satz 2.2.5 (*Maximum-Likelihood-Schätzung von $\tilde{\beta}$ und $\tilde{\sigma}^2$*). Es existieren eindeutig bestimmte Maximum-Likelihood-Schätzer für β und σ^2 , die folgendermaßen aussehen:

$$\begin{aligned}\tilde{\beta} &= \hat{\beta} = (X^\top X)^{-1} X^\top Y \\ \tilde{\sigma}^2 &= \frac{n-m}{n} \hat{\sigma}^2 = \frac{1}{n} |Y - X\tilde{\beta}|^2.\end{aligned}$$

Beweis. Wir fixieren $\sigma^2 > 0$ und suchen

$$\tilde{\beta} = \operatorname{argmax}_{\beta \in \mathbb{R}^m} \log L(Y, \beta, \sigma^2) = \operatorname{argmin}_{\beta \in \mathbb{R}^m} |Y - X\beta|^2,$$

woraus folgt, daß $\tilde{\beta}$ mit dem bekannten MKQ-Schätzer $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ identisch ist, der nicht von σ^2 abhängt. Berechnen wir jetzt

$$\tilde{\sigma}^2 = \operatorname{argmax}_{\sigma^2 > 0} \log L(Y, \tilde{\beta}, \sigma^2) = \operatorname{argmax}_{\sigma^2 > 0} g(\sigma^2).$$

Es gilt

$$g(\sigma^2) \xrightarrow{\sigma^2 \rightarrow +\infty} -\infty, \quad g(\sigma^2) \xrightarrow{\sigma^2 \rightarrow 0} -\infty,$$

weil $|Y - X\tilde{\beta}|^2 \neq 0$, dadurch, daß $Y \sim N(X\beta, \sigma^2 I) \in \{Xy : y \in \mathbb{R}^m\}$ mit Wahrscheinlichkeit Null. Da

$$g'(\sigma^2) = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{|Y - X\tilde{\beta}|^2}{2(\sigma^2)^2} = 0, \quad \text{ist } \tilde{\sigma}^2 = \frac{1}{n} |Y - X\tilde{\beta}|^2$$

ein Maximumpunkt von $g(\sigma^2)$, das heißt, $\tilde{\sigma}^2$ ist ein Maximum-Likelihood-Schätzer für σ^2 . \square

Satz 2.2.6. Unter den obigen Voraussetzungen gilt:

1. $\mathbb{E} \tilde{\sigma}^2 = \frac{n-m}{n} \sigma^2$, das heißt, $\tilde{\sigma}^2$ ist nicht erwartungstreu; allerdings ist er asymptotisch unverzerrt.
2. $\frac{n}{\tilde{\sigma}^2} \tilde{\sigma}^2 \sim \chi_{n-m}^2$, $\frac{n-m}{\tilde{\sigma}^2} \tilde{\sigma}^2 \sim \chi_{n-m}^2$.

Beweis. 1. Trivial (vergleiche den Beweis von Satz 2.2.4)

2. Wir zeigen den Satz nur für $\hat{\sigma}^2$.

$$\begin{aligned}\frac{n-m}{\sigma^2} \hat{\sigma}^2 &= \frac{1}{\sigma^2} |Y - X\hat{\beta}|^2 \\ &= \frac{1}{\sigma^2} Y^\top \underbrace{D}_{=D^2} Y \quad (\text{nach dem Beweis von Satz 2.2.4}) \\ &= \frac{1}{\sigma^2} (DY)^\top DY = \frac{1}{\sigma^2} (D(X\beta + \varepsilon))^\top \cdot D(X\beta + \varepsilon) \\ &= \frac{1}{\sigma^2} (D\varepsilon)^\top D\varepsilon = \left(\frac{\varepsilon^\top}{\sigma}\right) D \left(\frac{\varepsilon}{\sigma}\right),\end{aligned}$$

wobei

$$\begin{pmatrix} \varepsilon \\ \sigma \end{pmatrix} \sim N(0, \mathcal{I}).$$

Nach Satz 2.1.8 gilt

$$\frac{\varepsilon^\top}{\sigma} D \frac{\varepsilon}{\sigma} \sim \chi_r^2,$$

wobei $r = \text{Rang}(D)$, weil $D\mathcal{I} = D$ idempotent ist. Falls $r = n - m$, dann ist $\frac{n-m}{\sigma^2} \sim \chi_{n-m}^2$. Zeigen wir, daß $\text{Rang}(D) = r = n - m$. Aus der linearen Algebra ist bekannt, daß $\text{Rang}(D) = n - \dim(\text{Kern}(D))$. Wir zeigen, daß $\text{Kern}(D) = \{Xx : x \in \mathbb{R}^m\}$ und damit $\dim(\text{Kern}(D)) = m$, weil $\text{Rang}(X) = m$. Es ist $\{Xx : x \in \mathbb{R}^m\} \subseteq \text{Kern}(D)$, da

$$DX = (\mathcal{I} - X(X^\top X)^{-1}X^\top)X = X - (X^\top X)^{-1}X^\top X = 0.$$

und $\text{Kern}(D) \subseteq \{Xx : x \in \mathbb{R}^m\}$, weil

$$\begin{aligned} \forall y \in \text{Kern}(D) : \quad Dy = 0 &\iff (\mathcal{I} - X(X^\top X)^{-1}X^\top)y = 0 \\ &\iff y = X \cdot \underbrace{(X^\top X)^{-1}X^\top Y}_x = Xx \in \{Xx : x \in \mathbb{R}^m\}. \end{aligned}$$

□

Satz 2.2.7. Sei $Y = X\beta + \varepsilon$ ein multivariates lineares Regressionsmodell mit $Y = (Y_1, \dots, Y_n)^\top$, Designmatrix X mit $\text{Rang}(X) = m$, $\beta = (\beta_1, \dots, \beta_m)^\top$, $\varepsilon \sim N(0, \sigma^2 \mathcal{I})$. Dann sind die Schätzer $\hat{\beta} = (X^\top X)^{-1}X^\top Y$ für β bzw. $\hat{\sigma}^2 = \frac{1}{n-m} |Y - X\hat{\beta}|^2$ für σ^2 unabhängig voneinander.

Beweis. In diesem Beweis verwenden wir den Satz 2.1.9, für dessen Anwendung wir $\hat{\beta}$ als lineare und $\hat{\sigma}^2$ als quadratische Form von ε darstellen. Es ist in den Beweisen der Sätze 2.2.2 und 2.2.6 gezeigt worden, daß

$$\begin{aligned} \hat{\beta} &= \beta + \underbrace{(X^\top X)^{-1}X^\top}_{=A} \varepsilon, \\ \hat{\sigma}^2 &= \frac{1}{n-m} \varepsilon^\top D \varepsilon, \quad \text{wobei } D = \mathcal{I} - X(X^\top X)^{-1}X^\top. \end{aligned}$$

Zusätzlich gilt $AD = 0$, weil nach dem Beweis des Satzes 2.2.4

$$(AD)^\top = D^\top A^\top = \underbrace{D \cdot X}_{=0} ((X^\top X)^{-1})^\top = 0.$$

Da $\varepsilon \sim N(0, \sigma^2 \mathcal{I})$, folgt daraus

$$A\sigma^2 \mathcal{I} D = 0.$$

Deshalb sind die Voraussetzungen des Satzes 2.1.9 erfüllt, und $\hat{\beta}$ und $\hat{\sigma}^2$ sind unabhängig. □

2.2.4 Tests für Regressionsparameter

In diesem Abschnitt wird zunächst die Hypothese

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0$$

für ein $\beta_0 \in \mathbb{R}^m$ getestet. Dafür definieren wir die Testgröße

$$T = \frac{(\hat{\beta} - \beta_0)^\top X^\top X (\hat{\beta} - \beta_0)}{m\hat{\sigma}^2}.$$

Man kann zeigen (vergleiche Satz 2.2.8), daß unter H_0 gilt:

$$T \sim F_{m,n-m}.$$

Daraus folgt, daß H_0 abgelehnt werden soll, falls $T > F_{m,n-m,1-\alpha}$, wobei $F_{m,n-m,1-\alpha}$ das $(1-\alpha)$ -Quantil der $F_{m,n-m}$ -Verteilung darstellt. Dies ist ein Test zum Niveau $\alpha \in (0, 1)$.

Spezialfall: Der Fall $\beta_0 = 0$ beschreibt einen *Test auf Zusammenhang*; das heißt, man testet, ob die Parameter β_1, \dots, β_m für die Beschreibung der Daten Y relevant sind.

Bemerkung 2.2.1. 1. Wie kann man verstehen, daß die Testgröße T tatsächlich H_0 von H_1 unterscheiden soll? Führen wir die Bezeichnung

$$\tilde{Y} = Y - \underbrace{X\hat{\beta}}_{:=\hat{Y}}$$

ein; dabei gilt:

$$\hat{\sigma}^2 = \frac{1}{n-m} |\tilde{Y}|^2$$

und \tilde{Y} ist der Vektor der *Residuen*.

Ohne Beschränkung der Allgemeinheit setzen wir $\beta_0 = 0$. Falls H_0 nicht gelten soll, dann ist $\beta \neq 0$, und somit

$$|X\beta|^2 = (X\beta)^\top X\beta = \beta^\top X^\top X\beta > 0,$$

weil X den vollen Rang hat. Daraus folgt, daß H_0 abgelehnt werden soll, falls

$$|\hat{Y}|^2 = |X\hat{\beta}|^2 = \hat{\beta}^\top X^\top X\hat{\beta} \gg 0.$$

In der Testgröße $|X\hat{\beta}|^2$ sind allerdings die Schwankungen der Schätzung von β nicht berücksichtigt. Deswegen teilt man $|X\hat{\beta}|^2$ durch $\hat{\sigma}^2$:

$$T = \frac{\hat{\beta}^\top X^\top X\hat{\beta}}{m \cdot \hat{\sigma}^2} = \frac{|\hat{Y}|^2}{\frac{m}{n-m} |Y - \hat{Y}|^2}.$$

Der Satz von Pythagoras liefert

$$|Y|^2 = |\tilde{Y}|^2 + |\hat{Y}|^2,$$

wobei unter H_0

$$\mathbb{E} |\hat{Y}|^2 = \mathbb{E} |Y|^2 - \mathbb{E} |Y - \hat{Y}|^2 = n\sigma^2 - \mathbb{E} |\tilde{Y}|^2 \quad \text{gilt, und somit}$$

$$\frac{\mathbb{E} |\hat{Y}|^2}{\mathbb{E} \left(\frac{m}{n-m} |\tilde{Y}|^2 \right)} \stackrel{(H_0)}{=} \frac{n\sigma^2 - \mathbb{E} |\tilde{Y}|^2}{\frac{m}{n-m} \mathbb{E} |\tilde{Y}|^2} = \frac{n-m}{m} \left(\frac{n\sigma^2}{\mathbb{E} |\tilde{Y}|^2} - 1 \right),$$

$$\text{weil } \mathbb{E} |Y|^2 = \mathbb{E} (Y^\top Y) = \sigma^2 \cdot n, \quad \text{wegen } Y \sim N(0, \sigma^2 \mathcal{I}).$$

\implies Die Testgröße T ist sensibel gegenüber Abweichungen von H_0 .

2. Die Größe

$$|\tilde{Y}|^2 = |Y - \hat{Y}|^2$$

wird *Reststreuung* genannt. Mit deren Hilfe kann der Begriff des *Bestimmtheitsmaßes* R^2 aus der Stochastik I wie folgt verallgemeinert werden:

$$R^2 = 1 - \frac{|\tilde{Y}|^2}{|Y - \bar{Y}_n \cdot e|^2},$$

wobei $e = (1, \dots, 1)^\top$, $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$.

Satz 2.2.8. Unter $H_0 : \beta = \beta_0$ gilt

$$T = \frac{(\hat{\beta} - \beta_0)^\top X^\top X (\hat{\beta} - \beta_0)}{m\hat{\sigma}^2} \sim F_{m, n-m}.$$

Beweis. Es gilt

$$\hat{\beta} \sim N \left(\beta_0, \sigma^2 (X^\top X)^{-1} \right)$$

$$\implies \hat{\beta} - \beta_0 \sim N \left(0, \underbrace{\sigma^2 (X^\top X)^{-1}}_{:=K} \right).$$

Falls $A = \frac{X^\top X}{\sigma^2}$, dann ist $AK = \mathcal{I}$ idempotent. Dann gilt nach Satz 2.1.8

$$(\hat{\beta} - \beta_0)^\top A (\hat{\beta} - \beta_0) \stackrel{H_0}{\sim} \chi_m^2$$

(Zur Information: Unter H_1 wäre $(\hat{\beta} - \beta_0)^\top A (\hat{\beta} - \beta_0)$ nicht-zentral χ^2 -verteilt).

Es gilt zusätzlich:

$$\frac{n-m}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-m}^2.$$

Aus Satz 2.2.7 folgt die Unabhängigkeit von $(\hat{\beta} - \beta_0)^\top A(\hat{\beta} - \beta_0)$ und $\frac{n-m}{\sigma^2} \hat{\sigma}^2$.

$$\implies T = \frac{(\hat{\beta} - \beta_0)^\top (X^\top X)(\hat{\beta} - \beta_0)/m}{(n-m)\hat{\sigma}^2/(n-m)} \sim F_{m,n-m}$$

nach der Definition der F -Verteilung. □

Jetzt wird die Relevanz der einzelnen Parameter β_j getestet:

$$H_0 : \beta_j = \beta_{0j} \text{ vs. } H_1 : \beta_j \neq \beta_{0j}.$$

Satz 2.2.9. Unter $H_0 : \beta_j = \beta_{0j}$ gilt:

$$T_j = \frac{\hat{\beta}_j - \beta_{0j}}{\hat{\sigma} \sqrt{x^{jj}}} \sim t_{n-m}, \text{ wobei}$$

$$(X^\top X)^{-1} = (x^{ij})_{i,j=1,\dots,m}.$$

Beweis. Aus $\hat{\beta} \stackrel{H_0}{\sim} N(\beta_0, \sigma^2(X^\top X)^{-1})$ folgt $\hat{\beta}_j \stackrel{H_0}{\sim} N(\beta_{0j}, \sigma^2 x^{jj})$ und somit $\hat{\beta}_j - \beta_{0j} \sim N(0, \sigma^2 x^{jj})$. Dann ist $\frac{\hat{\beta}_j - \beta_{0j}}{\sigma \sqrt{x^{jj}}} \sim N(0, 1)$. Zusätzlich gilt: $\frac{(n-m)\hat{\sigma}^2}{\sigma^2} \stackrel{H_0}{\sim} \chi_{n-m}^2$, und nach Satz 2.2.7 sind beide Größen unabhängig. Daraus folgt:

$$T_j = \frac{\frac{\hat{\beta}_j - \beta_{0j}}{\sigma \sqrt{x^{jj}}}}{\sqrt{\frac{(n-m)\hat{\sigma}^2}{(n-m)\sigma^2}}} \sim t_{n-m}.$$

□

Somit wird $H_0 : \beta_j = \beta_{0j}$ abgelehnt, falls $|T| > t_{n-m, 1-\alpha/2}$. Dies ist ein Test von H_0 vs. H_1 zum Niveau α .

Sei nun

$$H_0 : \beta_{j_1} = \beta_{0j_1}, \dots, \beta_{j_l} = \beta_{0j_l} \text{ vs. } H_1 : \exists i \in \{1, \dots, l\} : \beta_{j_i} \neq \beta_{0j_i}$$

die zu testende Hypothese.

Übungsaufgabe 2.2.2. Zeigen Sie, daß unter H_0 folgende Verteilungsaussage gilt:

$$T = \frac{(\hat{\beta}' - \beta_0')^\top K'(\hat{\beta}' - \beta_0')}{l\hat{\sigma}^2} \sim F_{l,n-m},$$

wobei

$$\begin{aligned}\hat{\beta}' &= (\hat{\beta}_{j_1}, \dots, \hat{\beta}_{j_i}), \\ \beta'_0 &= (\beta_{0j_1}, \dots, \beta_{0j_i}), \\ K' &= \begin{pmatrix} x^{j_1j_1} & \dots & x^{j_1j_i} \\ \vdots & \ddots & \vdots \\ x^{j_ij_1} & \dots & x^{j_ij_i} \end{pmatrix}^{-1}.\end{aligned}$$

Konstruieren Sie den dazugehörigen F -Test!

Test auf Linearkombination von Parametern

Sei nun

$$H_0 : H\beta = c \text{ vs. } H_1 : H\beta \neq c,$$

wobei H eine $(r \times m)$ -Matrix und $c \in \mathbb{R}^r$ sind.

Satz 2.2.10. Unter H_0 gilt

$$T = \frac{(H\hat{\beta} - c)^\top (H(X^\top X)^{-1}H^\top)^{-1}(H\hat{\beta} - c)}{r\hat{\sigma}^2} \sim F_{r,n-m}.$$

Deshalb wird $H_0 : H\beta = c$ abgelehnt, falls $T > F_{r,n-m,1-\alpha}$.

Übungsaufgabe 2.2.3. Beweisen Sie Satz 2.2.10!

2.2.5 Konfidenzbereiche

1. Konfidenzintervall für β_j

Im Satz 2.2.9 haben wir gezeigt, daß

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \cdot \sqrt{x^{jj}}} \sim t_{n-m},$$

wobei $(X^\top X)^{-1} = (x^{ij})_{i,j=1,\dots,m}$. Daraus kann mit den üblichen Überlegungen folgendes Konfidenzintervall für β_j zum Niveau $1 - \alpha$ abgeleitet werden:

$$\mathbb{P}\left(\hat{\beta}_j - t_{n-m,1-\alpha/2} \cdot \hat{\sigma} \sqrt{x^{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{n-m,1-\alpha/2} \cdot \hat{\sigma} \sqrt{x^{jj}}\right) = 1 - \alpha.$$

2. Simultaner Konfidenzbereich für $\beta = (\beta_1, \dots, \beta_m)^\top$

Falls A_j wie unten definiert ist, dann erhält man mit Hilfe folgender *Bonferroni-Ungleichung*

$$\mathbb{P}\left(\bigcap_{j=1}^m A_j\right) \geq \sum_{j=1}^m \mathbb{P}(A_j) - (m - 1),$$

daß

$$\mathbb{P}\left(\underbrace{\hat{\beta}_j - t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma} \sqrt{x^{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma} \sqrt{x^{jj}}}_{:=A_j}, \quad j = 1, \dots, m\right)$$

$$\stackrel{\text{(Bonferroni)}}{\geq} \sum_{j=1}^m \mathbb{P}(A_j) - (m-1) = m \cdot \left(1 - \frac{\alpha}{m}\right) - m + 1 = 1 - \alpha.$$

Daraus folgt, daß

$$\left\{ \beta = (\beta_1, \dots, \beta_m)^\top : \beta_j \in \left[\hat{\beta}_j - t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma} \sqrt{x^{jj}}, \hat{\beta}_j + t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma} \sqrt{x^{jj}} \right] \right\}$$

ein simultaner Konfidenzbereich für β zum Niveau $1 - \alpha$ ist.

3. Konfidenzellipsoid für β .

In Satz 2.2.8 haben wir bewiesen, daß

$$T = \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{m\hat{\sigma}^2} \sim F_{m,n-m}.$$

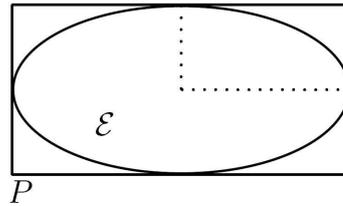
Daraus folgt, daß

$$\mathbb{P}(T \leq F_{m,n-m,1-\alpha}) = 1 - \alpha \quad \text{und}$$

$$\mathcal{E} = \left\{ \beta \in \mathbb{R}^m : \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{m\hat{\sigma}^2} \leq F_{m,n-m,1-\alpha} \right\}$$

ein Konfidenzellipsoid zum Niveau $1 - \alpha$ ist, siehe Abbildung 2.2.

Abbildung 2.2: Konfidenzellipsoid



Da ein Ellipsoid in das minimale Parallelepiped P eingebettet werden kann, sodaß die Seitenlängen von P gleich $2 \times$ der Halbachsenlängen von \mathcal{E} sind, ergibt sich folgender simultaner Konfidenzbereich für $\beta = (\beta_1, \dots, \beta_m)^\top$:

$$P = \left\{ \beta : \hat{\beta}_j - \hat{\sigma} \sqrt{m x^{jj} F_{m,n-m,1-\alpha}} \leq \beta_j \leq \hat{\beta}_j + \hat{\sigma} \sqrt{m x^{jj} F_{m,n-m,1-\alpha}} \right\}$$

$j = 1, \dots, m.$

4. Konfidenzintervall für den erwarteten Zielwert $x_{01}\beta_1 + \dots + x_{0m}\beta_m$.

Sei $Y_0 = x_{01}\beta_1 + \dots + x_{0m}\beta_m + \varepsilon_0$ eine neue Zielvariable mit $\mathbb{E}\varepsilon_0 = 0$. Dann ist

$$\mathbb{E}Y_0 = \sum_{i=1}^n x_{0i}\beta_i.$$

Wir konstruieren ein Konfidenzintervall für $\mathbb{E}Y_0$. Dazu verwenden wir die Beweisidee des Satzes 2.2.9 kombiniert mit Satz 2.2.10 mit $H = (x_{01}, \dots, x_{0m}) = x_0^\top$, $r = 1$. Dann ist

$$T = \frac{\sum_{i=1}^m \hat{\beta}_i x_{0i} - \sum_{i=1}^m \beta_i x_{0i}}{\hat{\sigma} \sqrt{x_0^\top (X^\top X)^{-1} x_0}} \sim t_{n-m}.$$

Darum ist

$$\left\{ \beta = (\beta_1, \dots, \beta_m)^\top : \sum_{i=1}^m x_{0i} \hat{\beta}_i - \hat{\sigma} \sqrt{x_0^\top (X^\top X)^{-1} x_0} \cdot t_{n-m, 1-\alpha/2} \leq \sum_{i=1}^m x_{0i} \beta_i \leq \sum_{i=1}^m x_{0i} \hat{\beta}_i + \hat{\sigma} \sqrt{x_0^\top (X^\top X)^{-1} x_0} \cdot t_{n-m, 1-\alpha/2} \right\}$$

ein Konfidenzintervall für $\sum_{i=1}^m x_{0i} \beta_i$ zum Niveau $1 - \alpha$.

5. Prognoseintervall für die Zielvariable Y_0 .

Für $Y_0 = \sum_{i=1}^m x_{0i} \beta_i + \varepsilon_0$ mit $\varepsilon_0 \sim N(0, \sigma^2)$, ε_0 unabhängig von $\varepsilon_1, \dots, \varepsilon_n$, gilt:

$$\begin{aligned} x_0^\top \hat{\beta} - Y_0 &\sim N(0, \sigma^2(1 + x_0^\top (X^\top X)^{-1} x_0)) \\ \implies \frac{x_0^\top \hat{\beta} - Y_0}{\sigma \sqrt{1 + x_0^\top (X^\top X)^{-1} x_0}} &\sim N(0, 1) \\ \implies \frac{x_0^\top \hat{\beta} - Y_0}{\hat{\sigma} \sqrt{1 + x_0^\top (X^\top X)^{-1} x_0}} &\sim t_{n-m} \end{aligned}$$

Also ist

$$\left(x_0^\top \hat{\beta} + c, x_0^\top \hat{\beta} - c \right)$$

$$\text{mit } c = \hat{\sigma} \sqrt{1 + x_0^\top (X^\top X)^{-1} \cdot x_0} \cdot t_{n-m, 1-\alpha/2}$$

ein Prognoseintervall für die Zielvariable Y_0 zum Niveau $1 - \alpha$.

6. *Konfidenzband für die Regressionsebene $y = \beta_1 + \sum_{i=2}^m x_i \beta_i$ im multiplen Regressionsmodell.*

Es sei $Y = X\beta + \varepsilon$, wobei

$$X = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1m} \\ 1 & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \cdots & x_{nm} \end{pmatrix} \quad \text{und } \varepsilon \sim N(0, \sigma^2 \cdot \mathcal{I}).$$

Wir wollen ein zufälliges Konfidenzband $B(x)$ für y angeben. Es gilt

$$\mathbb{P} \left(y = \beta_1 + \sum_{i=2}^m \beta_i x_i \in B(x) \right) = 1 - \alpha \quad \forall x \in \mathbb{R}_1^{m-1}, \quad \text{wobei}$$

$$\mathbb{R}_1^{m-1} = \left\{ (1, x_2, \dots, x_m)^\top \in \mathbb{R}^m \right\}.$$

Satz 2.2.11. Es gilt:

$$\mathbb{P} \left(\max_{x \in \mathbb{R}_1^{m-1}} \frac{\left(x^T \hat{\beta} - \overbrace{\left(\beta_1 + \sum_{i=2}^m \beta_i x_i \right)}^{=y} \right)^2}{\hat{\sigma}^2 x^\top (X^\top X)^{-1} x} \leq m \cdot F_{m, n-m, 1-\alpha} \right) = 1 - \alpha.$$

Ohne Beweis.

2.3 Multivariate lineare Regression mit $\text{Rang}(X) < m$

Es sei $Y = X\beta + \varepsilon$, $Y \in \mathbb{R}^n$, wobei X eine $(n \times m)$ -Matrix mit $\text{Rang}(X) = r < m$ ist, $\beta = (\beta_1, \dots, \beta_m)^\top$, $\varepsilon \in \mathbb{R}^n$, $\mathbb{E} \varepsilon = 0$, $\mathbb{E}(\varepsilon_i \varepsilon_j) = \delta_{ij} \sigma^2$, $i, j = 1, \dots, n$, $\sigma^2 > 0$.

Der MKQ-Schätzer $\hat{\beta}$ ist nach wie vor eine Lösung der Normalgleichung

$$(X^\top X) \beta = X^\top Y.$$

$X^\top X$ ist aber nicht mehr invertierbar, weil

$$\text{Rang}(X^\top X) \leq \min \{ \text{Rang}(X), \text{Rang}(X^\top) \} = r < m.$$

Um $\hat{\beta}$ aus der Normalgleichung zu gewinnen, sollen beide Seiten der Gleichung mit der sogenannten *verallgemeinerten Inversen* von $X^\top X$ multipliziert werden.

2.3.1 Verallgemeinerte Inverse

Definition 2.3.1. Sei A eine $(n \times m)$ -Matrix. Eine $(m \times n)$ -Matrix A^- heißt *verallgemeinerte Inverse* von A , falls

$$AA^-A = A \quad \text{gilt.}$$

Die Matrix A^- ist nicht eindeutig bestimmt, was die folgenden Hilfssätze zeigen.

Lemma 2.3.1. Sei A eine $(n \times m)$ -Matrix, $m \leq n$ mit $\text{Rang}(A) = r \leq m$. Es existieren invertierbare Matrizen P ($n \times n$) und Q ($m \times m$), sodaß

$$PAQ = \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{wobei } \mathcal{I}_r = \text{diag}(\underbrace{1, \dots, 1}_{r \text{ Mal}}). \quad (2.3.1)$$

Folgerung 2.3.1. Für eine beliebige $(n \times m)$ -Matrix A mit $n \geq m$, $\text{Rang}(A) = r \leq m$ gilt

$$A^- = Q \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} P, \quad (2.3.2)$$

wobei P und Q Matrizen aus der Darstellung (2.3.1) sind, $\mathcal{I}_r = \text{diag}(\underbrace{1, \dots, 1}_{r \text{ Mal}})$, und A_1, A_2, A_3 beliebige $((m-r) \times r)$, $(r \times (n-r))$ bzw. $((m-r) \times (n-r))$ -Matrizen sind.

Insbesondere kann

$$\begin{aligned} A_1 &= 0, \\ A_2 &= 0, \\ A_3 &= \text{diag}(\underbrace{1, \dots, 1}_{s-r \text{ Mal}}, 0, \dots, 0), \\ s &\in \{r, \dots, m\} \end{aligned}$$

gewählt werden, das heißt, $\text{Rang}(A^-) = s \in \{r, \dots, m\}$ für

$$A^- = Q \begin{pmatrix} \mathcal{I}_s & 0 \\ 0 & 0 \end{pmatrix} P.$$

Beweis. Zeigen wir, daß für A^- wie in (2.3.2) gegeben, $AA^-A = A$ gilt. Aus Lemma 2.3.1 folgt, daß

$$\begin{aligned} A &= P^{-1} \cdot \text{diag}(1, \dots, 1, 0, \dots, 0) \cdot Q^{-1} \quad \text{und somit} \\ AA^-A &= P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} Q \cdot \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} P P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} \\ &= P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} \\ &= A. \end{aligned}$$

□

Lemma 2.3.2. Sei A eine beliebige $(n \times m)$ -Matrix mit $\text{Rang}(A) = r \leq m, m \leq n$.

1. Falls $(A^\top A)^-$ eine verallgemeinerte Inverse von $A^\top A$ ist, dann ist $((A^\top A)^-)^\top$ ebenfalls eine verallgemeinerte Inverse von $A^\top A$.
2. Es gilt die Darstellung

$$\begin{aligned} (A^\top A)(A^\top A)^- A^\top &= A^\top \quad \text{bzw.} \\ A(A^\top A)^-(A^\top A) &= A. \end{aligned}$$

Beweis. 1. $A^\top A$ ist symmetrisch, also

$$\underbrace{\left(A^\top A (A^\top A)^- A^\top A \right)^\top}_{=A^\top A ((A^\top A)^-)^\top A^\top A} = \left(A^\top A \right)^\top = A^\top A.$$

Also ist $((A^\top A)^-)^\top$ eine verallgemeinerte Inverse von $A^\top A$.

2. Es sei $B = (A^\top A)(A^\top A)^- A^\top - A^\top$. Wir zeigen, daß $B = 0$, indem wir zeigen, daß $BB^\top = 0$.

$$\begin{aligned} BB^\top &= \left((A^\top A)(A^\top A)^- A^\top - A^\top \right) \left(A \left((A^\top A)^- \right)^\top A^\top A - A \right) \\ &= A^\top A (A^\top A)^- A^\top A \left((A^\top A)^- \right)^\top A^\top A - \underbrace{A^\top A (A^\top A)^- A^\top A}_{=A^\top A} \\ &\quad - \underbrace{A^\top A \left((A^\top A)^- \right)^\top \cdot A^\top A}_{=A^\top A} + A^\top A = A^\top A - 2A^\top A + A^\top A = 0. \end{aligned}$$

Die Aussage $A(A^\top A)^- A^\top A = A$ erhält man, indem man die Matrizen an beiden Seiten der Gleichung $A^\top A (A^\top A)^- A^\top = A^\top$ transponiert. □

2.3.2 MKQ-Schätzer für β

Satz 2.3.1. Es sei X eine $(n \times m)$ -Designmatrix mit $\text{Rang}(X) = r \leq m$ in der linearen Regression $Y = X\beta + \varepsilon$. Die allgemeine Lösung der Normalgleichung

$$(X^\top X) \beta = X^\top Y$$

sieht folgendermaßen aus:

$$\beta = (X^\top X)^- X^\top Y + \left(\mathcal{I}_m - (X^\top X)^- X^\top X \right) z, \quad z \in \mathbb{R}^m. \quad (2.3.3)$$

Beweis. 1. Zeigen wir, daß β wie in (2.3.3) angegeben, eine Lösung der Normalgleichung darstellt.

$$\begin{aligned} X^\top X \beta &= \underbrace{(X^\top X)(X^\top X)^{-1} X^\top Y}_{=X^\top \text{ (Lemma 2.3.2, 2.)}} + \left(X^\top X - \underbrace{X^\top X (X^\top X)^{-1} X^\top X}_{=X^\top X} \right) z \\ &= X^\top Y \end{aligned}$$

2. Zeigen wir, daß eine beliebige Lösung β' der Normalgleichung die Form (2.3.3) besitzt. Sei β die Lösung (2.3.3). Wir bilden die Differenz der Gleichungen

$$\begin{array}{rcl} (X^\top X) \beta' & = & X^\top Y \\ - (X^\top X) \beta & = & X^\top Y \\ \hline (X^\top X) (\beta' - \beta) & = & 0 \end{array}$$

$$\begin{aligned} \beta' &= (\beta' - \beta) + \beta \\ &= \beta' - \beta + (X^\top X)^{-1} X^\top Y + \left(\mathcal{I}_m - (X^\top X)^{-1} X^\top X \right) z \\ &= (X^\top X)^{-1} X^\top Y + \left(\mathcal{I}_m - (X^\top X)^{-1} X^\top X \right) z + (\beta' - \beta) - \underbrace{(X^\top X)^{-1} X^\top X (\beta' - \beta)}_{=0} \\ &= (X^\top X)^{-1} X^\top Y + \left(\mathcal{I}_m - (X^\top X)^{-1} X^\top X \right) \underbrace{\left(z + \beta' - \beta \right)}_{=z_0} \\ &\implies \beta' \text{ besitzt die Darstellung (2.3.3).} \end{aligned}$$

□

Bemerkung 2.3.1. Der Satz 2.3.1 liefert die Menge aller Extremalpunkte der MKQ-Minimierungsaufgabe

$$e(\beta) = \frac{1}{n} |Y - X\beta|^2 \longrightarrow \min_{\beta}.$$

Deshalb soll die Menge aller MKQ-Schätzer von β in (2.3.3) zusätzliche Anforderungen erfüllen.

Satz 2.3.2. 1. Alle MKQ-Schätzer von β haben die Form

$$\bar{\beta} = \left(X^\top X \right)^{-} X^\top Y, \quad \text{wobei}$$

$(X^\top X)^{-}$ eine beliebige verallgemeinerte Inverse von $X^\top X$ ist.

2. $\bar{\beta}$ ist nicht erwartungstreu, denn

$$\mathbb{E} \bar{\beta} = \left(X^\top X \right)^{-} X^\top X \beta.$$

3. Es gilt:

$$\text{Cov}\bar{\beta} = \sigma^2 \left(X^\top X \right)^- \left(X^\top X \right) \left(\left(X^\top X \right)^- \right)^\top.$$

Beweis. 1. Zeigen wir, daß $e(\beta) \geq e(\bar{\beta}) \quad \forall \beta \in \mathbb{R}^m$.

$$\begin{aligned} n \cdot e(\beta) &= |Y - X\beta|^2 = (Y - X\bar{\beta} + X(\bar{\beta} - \beta))^\top (Y - X\bar{\beta} + X(\bar{\beta} - \beta)) \\ &= (Y - X\bar{\beta})^\top (Y - X\bar{\beta}) + (X(\bar{\beta} - \beta))^\top (X(\bar{\beta} - \beta)) \\ &\quad + 2(\bar{\beta} - \beta)^\top X^\top (Y - X\bar{\beta}) \\ &= n \cdot e(\bar{\beta}) + \underbrace{2 \cdot (\bar{\beta} - \beta)^\top (X^\top Y - (X^\top X)\bar{\beta})}_{=0} + |X(\bar{\beta} - \beta)|^2 \\ &\geq n \cdot e(\bar{\beta}) + 0 = n \cdot e(\bar{\beta}), \quad \text{denn} \end{aligned}$$

$\bar{\beta}$ hat die Form (2.3.3) mit $z = 0$ und ist somit eine Lösung der Normalengleichung.

2. Es gilt:

$$\begin{aligned} \mathbb{E}\bar{\beta} &= \mathbb{E} \left((X^\top X)^- X^\top Y \right) = \left(X^\top X \right)^- X^\top \mathbb{E} Y \\ &= \left(X^\top X \right)^- X^\top X \beta, \quad \text{weil aus} \\ Y &= X\beta + \varepsilon, \quad \mathbb{E}\varepsilon = 0 \quad \text{die Relation } \mathbb{E} Y = X\beta \text{ folgt.} \end{aligned}$$

Warum ist $\bar{\beta}$ nicht erwartungstreu? Also warum ist $(X^\top X)^- X^\top X \beta \neq \beta$, $\beta \in \mathbb{R}^m$? Da $\text{Rang}(X) = r < m$, ist $\text{Rang}(X^\top X) < m$ und damit $\text{Rang}((X^\top X)^- X^\top X) < m$. Darum existiert ein $\beta \neq 0$, für das gilt:

$$\left(X^\top X \right)^- X^\top X \beta = 0 \neq \beta,$$

also ist $\bar{\beta}$ nicht erwartungstreu. Es gilt sogar, daß alle Lösungen von (2.3.3) keine erwartungstreuen Schätzer sind. Wenn wir den Erwartungswert an (2.3.3) anwenden, so erhielten wir im Falle der Erwartungstreue:

$$\begin{aligned} \forall \beta \in \mathbb{R}^m : \quad \beta &= (X^\top X)^- X^\top X \beta + \left(\mathcal{I}_m - (X^\top X)^- (X^\top X) \right) z, \quad z \in \mathbb{R}^m. \\ \implies \left(\mathcal{I}_m - (X^\top X)^- (X^\top X) \right) (z - \beta) &= 0 \quad \forall z, \beta \in \mathbb{R}^m \\ \implies (X^\top X)^- (X^\top X) (\beta - z) &= \beta - z, \quad \forall z, \beta \in \mathbb{R}^m. \end{aligned}$$

Da diese Gleichung nicht für alle $\beta \in \mathbb{R}^m$ gelten kann (siehe oben), führt die Annahme der Erwartungstreue zum Widerspruch.

3. Es gilt:

$$\begin{aligned}
\text{Cov}(\bar{\beta}_i, \bar{\beta}_j) &= \text{Cov}\left(\underbrace{\left((X^\top X)^{-1} X^\top Y\right)}_{:=A=(a_{kl})}, \left((X^\top X)^{-1} X^\top Y\right)_j\right) \\
&= \text{Cov}\left(\sum_{k=1}^n a_{ik} Y_k, \sum_{l=1}^n a_{jl} Y_l\right) \\
&= \sum_{k,l=1}^n a_{ik} a_{jl} \underbrace{\text{Cov}(Y_k, Y_l)}_{=\sigma^2 \cdot \delta_{kl}} = \sigma^2 \sum_{k=1}^n a_{ik} a_{jk} = \left(\sigma^2 A A^\top\right)_{i,j} \\
&= \left(\sigma^2 (X^\top X)^{-1} X^\top X \left((X^\top X)^{-1}\right)^\top\right)_{i,j}.
\end{aligned}$$

□

2.3.3 Erwartungstreu schätzbare Funktionen

Definition 2.3.2. Eine Linearkombination $a^\top \beta$ von β_1, \dots, β_m , $a \in \mathbb{R}^m$ heißt (*erwartungstreu*) *schätzbar*, falls

$$\exists c \in \mathbb{R}^n : \quad \mathbb{E}(c^\top Y) = a^\top \beta,$$

das heißt, falls es einen linearen, erwartungstreuen Schätzer $c^\top Y$ für $a^\top \beta$ gibt.

Satz 2.3.3. Die Funktion $a^\top \beta$, $a \in \mathbb{R}^m$ ist genau dann erwartungstreu schätzbar, wenn eine der folgenden Bedingungen erfüllt ist:

1. $\exists c \in \mathbb{R}^n : \quad a^\top = c^\top X$.
2. a erfüllt die Gleichung

$$a^\top (X^\top X)^{-1} X^\top X = a^\top. \quad (2.3.4)$$

Beweis. 1. „ \implies “: Falls $a^\top \beta$ schätzbar, dann existiert ein $d \in \mathbb{R}^n$ mit $\mathbb{E}(d^\top Y) = a^\top \beta \quad \forall \beta \in \mathbb{R}^m$. Also

$$\begin{aligned}
a^\top \beta &= d^\top \mathbb{E} Y = d^\top X \beta \Rightarrow (a^\top - d^\top X) \beta = 0, \quad \forall \beta \in \mathbb{R}^m \\
&\implies a^\top = d^\top X,
\end{aligned}$$

setze $c = d$, damit ist die erste Richtung bewiesen.

„ \impliedby “: $\mathbb{E}(c^\top Y) = c^\top \mathbb{E} Y = c^\top X \beta = a^\top \beta$, also ist $a^\top \beta$ erwartungstreu schätzbar.

2. „ \implies “: Falls $a^\top \beta$ erwartungstreu schätzbar ist, dann gilt:

$$a^\top (X^\top X)^- X^\top X \stackrel{\text{Punkt 1}}{=} c^\top \underbrace{X \cdot (X^\top X)^- X^\top X}_{=X \text{ (Lemma 2.3.2)}} = c^\top X \stackrel{\text{(Punkt 1)}}{=} a^\top.$$

Also ist (2.3.4) erfüllt.

„ \impliedby “: Falls $a^\top (X^\top X)^- X^\top X = a^\top$, dann gilt mit $c = (a^\top (X^\top X)^- X^\top X)^\top$ nach Punkt 1, daß $a^\top \beta$ schätzbar ist. □

Bemerkung 2.3.2. Im Falle der Regression mit $\text{Rang}(X) = m$ ist die Gleichung (2.3.4) immer erfüllt, denn $(X^\top X)^- = (X^\top X)^{-1}$ und damit ist $a^\top \beta$ schätzbar für alle $a \in \mathbb{R}^m$.

Satz 2.3.4 (*Beispiele schätzbarer Funktionen*). Falls $\text{Rang}(X) = r < m$, dann sind folgende Linearkombinationen von β schätzbar:

1. Die Koordinaten $\sum_{j=1}^m x_{ij} \beta_j$, $i = 1, \dots, n$ des Erwartungswertvektors $\mathbb{E}Y = X\beta$.
2. Beliebige Linearkombinationen schätzbarer Funktionen.

Beweis. 1. Führe die Bezeichnung $\tilde{x}_i = (x_{i1}, \dots, x_{im})$, $i = 1, \dots, n$ ein. Dann ist

$$\sum_{j=1}^m x_{ij} \beta_j = \tilde{x}_i^\top \beta \quad \forall i = 1, \dots, n,$$

$$X\beta = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)^\top \beta.$$

$\tilde{x}_i \beta$ ist schätzbar, falls \tilde{x}_i die Gleichung (2.3.4) erfüllt, die für alle $i = 1, \dots, n$ folgendermaßen in Matrixform dargestellt werden kann:

$$X \left(X^\top X \right)^- X^\top X = X,$$

was nach Lemma 2.3.2 Gültigkeit besitzt.

2. Für $a_1, \dots, a_k \in \mathbb{R}^m$ seien $a_1^\top \beta, \dots, a_k^\top \beta$ schätzbare Funktionen. Für alle $\lambda = (\lambda_1, \dots, \lambda_k)^\top \in \mathbb{R}^k$ zeigen wir, daß $\sum_{i=1}^k \lambda_i \cdot a_i^\top \beta = \lambda^\top A \beta$ schätzbar ist, wobei $A = (a_1, \dots, a_k)^\top$. Zu zeigen bleibt: $b = (\lambda^\top A)^\top$ erfüllt (2.3.4), also

$$\lambda^\top A \left(X^\top X \right)^- X^\top X = \lambda^\top A.$$

Diese Gleichung stimmt, weil $a_i^\top (X^\top X)^- X^\top X = a_i^\top$, $i = 1, \dots, k$. Nach Satz 2.3.3, 2.) ist $\lambda^\top A \beta$ schätzbar. □

Satz 2.3.5 (Gauß-Markov). Es sei $a^\top \beta$ eine schätzbare Funktion, $a \in \mathbb{R}^m$ im linearen Regressionsmodell $Y = X\beta + \varepsilon$ mit $\text{Rang}(X) \leq m$.

1. Der beste lineare erwartungstreu Schätzer (engl. BLUE - best linear unbiased estimator) von $a^\top \beta$ ist durch $a^\top \bar{\beta}$ gegeben, wobei

$$\bar{\beta} = (X^\top X)^{-1} X^\top Y$$

ein MKQ-Schätzer für β ist.

2. $\text{Var}(a^\top \bar{\beta}) = \sigma^2 a^\top (X^\top X)^{-1} a$.

Beweis. Die Linearität von $a^\top \bar{\beta} = a^\top (X^\top X)^{-1} X^\top Y$ als Funktion von Y ist klar. Zeigen wir die Erwartungstreue:

$$\begin{aligned} \mathbb{E}(a^\top \bar{\beta}) &= a^\top \mathbb{E} \bar{\beta} = a^\top (X^\top X)^{-1} X^\top X \beta \\ &= c^\top \underbrace{X(X^\top X)^{-1} X^\top X}_{=X \text{ (Lemma 2.3.2)}} \beta = \underbrace{c^\top X}_{=a^\top} \beta = a^\top \beta \quad \forall \beta \in \mathbb{R}^m. \end{aligned}$$

Berechnen wir $\text{Var}(a^\top \bar{\beta})$ (also beweisen wir Punkt 2), und zeigen, daß sie minimal ist.

$$\begin{aligned} \text{Var}(a^\top \bar{\beta}) &= \text{Var} \left(\sum_{i=1}^m a_i \bar{\beta}_i \right) = \sum_{i,j=1}^m a_i a_j \cdot \text{Cov}(\bar{\beta}_i, \bar{\beta}_j) \\ &= a^\top \text{Cov}(\bar{\beta}) a \stackrel{\text{(Satz 2.3.2)}}{=} a^\top \sigma^2 \left((X^\top X)^{-1} X^\top X (X^\top X)^{-1} \right)^\top a \\ &= \sigma^2 \cdot a^\top \underbrace{\left((X^\top X)^{-1} \right)^\top}_{=(X^\top X)^{-1}} X^\top X \underbrace{\left((X^\top X)^{-1} \right)^\top}_{=(X^\top X)^{-1}} a \\ &\stackrel{\text{Lemma 2.3.2, 1.)}}{=} \sigma^2 a^\top (X^\top X)^{-1} X^\top X (X^\top X)^{-1} a \\ &\stackrel{\text{Satz 2.3.3, 1.)}}{=} \sigma^2 \cdot c^\top \underbrace{X \cdot (X^\top X)^{-1} X^\top X}_{=X} (X^\top X)^{-1} X^\top c \\ &= \sigma^2 \underbrace{c^\top X}_{=a^\top} (X^\top X)^{-1} \underbrace{X^\top c}_{=a} = \sigma^2 a^\top (X^\top X)^{-1} a. \end{aligned}$$

Jetzt zeigen wir, daß für einen beliebigen linearen, erwartungstreuen Schätzer $b^\top Y$ von $a^\top \beta$ gilt: $\text{Var}(b^\top Y) \geq \text{Var}(a^\top \bar{\beta})$. Weil $b^\top Y$ erwartungstreu ist, gilt: $\mathbb{E}(b^\top Y) = a^\top \beta$. Nach Satz 2.3.3 gilt: $a^\top = b^\top X$. Betrachten wir die Varianz von

$$\begin{aligned} 0 &\leq \text{Var}(b^\top Y - a^\top \bar{\beta}) = \text{Var}(b^\top Y) - 2\text{Cov}(b^\top Y, a^\top \bar{\beta}) + \text{Var}(a^\top \bar{\beta}) \\ &= \text{Var}(b^\top Y) - 2\sigma^2 a^\top (X^\top X)^{-1} a + \sigma^2 a^\top (X^\top X)^{-1} a = \text{Var}(b^\top Y) - \text{Var}(a^\top \bar{\beta}) \end{aligned}$$

mit

$$\begin{aligned} \text{Cov}(b^\top Y, a^\top \bar{\beta}) &= \text{Cov}(b^\top Y, a^\top (X^\top X)^- X^\top Y) = \sigma^2 a^\top (X^\top X)^- \underbrace{X^\top b}_{=a} \\ &= \sigma^2 a^\top (X^\top X)^- a. \end{aligned}$$

Damit ist $\text{Var}(b^\top Y) \geq \text{Var}(a^\top \bar{\beta})$ und $a^\top \bar{\beta}$ ist ein bester, linearer, erwartungstreuer Schätzer für $a^\top \beta$. \square

Bemerkung 2.3.3. 1. Falls $\text{Rang}(X) = m$, dann ist $a^\top \hat{\beta}$ der beste lineare, erwartungstreue Schätzer für $a^\top \beta$, $a \in \mathbb{R}^m$.

2. Wie im folgenden Satz gezeigt wird, hängt der Schätzer $a^\top \bar{\beta} = a^\top (X^\top X)^- X^\top Y$ nicht von der Wahl der verallgemeinerten Inversen ab.

Satz 2.3.6. Der beste lineare, erwartungstreue Schätzer $a^\top \bar{\beta}$ für $a^\top \beta$ ist eindeutig bestimmt.

Beweis.

$$a^\top \bar{\beta} = a^\top (X^\top X)^- X^\top Y \stackrel{\text{Satz 2.3.3, 1.})}{=} c^\top X (X^\top X)^- X^\top Y.$$

Wir zeigen, daß $X(X^\top X)^- X^\top$ nicht von der Wahl von $(X^\top X)^-$ abhängt. Zeigen wir, daß für beliebige verallgemeinerte Inverse A_1 und A_2 von $(X^\top X)$ gilt: $XA_1 X^\top = XA_2 X^\top$. Nach Lemma 2.3.2, 2.) gilt:

$$XA_1 X^\top X = X = XA_2 X^\top X.$$

Multiplizieren wir alle Teile der Gleichung mit $A_1 X^\top$ von rechts:

$$XA_1 \underbrace{X^\top X A_1 X^\top}_{=X^\top} = XA_1 X^\top = XA_2 \underbrace{X^\top X A_1 X^\top}_{=X^\top}$$

Also ist $XA_1 X^\top = XA_2 X^\top$. \square

2.3.4 Normalverteilte Störgrößen

Sei $Y = X\beta + \varepsilon$ ein lineares Regressionsmodell mit $\text{Rang}(X) = r < m$ und $\varepsilon \sim N(0, \sigma^2 \mathcal{I})$. Genauso wie in Abschnitt 2.2.3 können Maximum-Likelihood-Schätzer $\tilde{\beta}$ und $\tilde{\sigma}^2$ für β und σ^2 hergeleitet werden. Und genauso wie im Satz 2.2.5 kann gezeigt werden, daß

$$\begin{aligned} \tilde{\beta} &= \bar{\beta} = (X^\top X)^- X^\top Y \quad \text{und} \\ \tilde{\sigma}^2 &= \frac{1}{n} |Y - X\bar{\beta}|^2. \end{aligned}$$

Jetzt werden die Verteilungseigenschaften von $\bar{\beta}$ und $\tilde{\sigma}^2$ untersucht. Wir beginnen mit der Erwartungstreue von $\tilde{\sigma}^2$. Wir zeigen, daß $\tilde{\sigma}^2$ nicht erwartungstreu ist, dafür ist aber der korrigierte Schätzer

$$\bar{\sigma}^2 = \frac{1}{n-r} |Y - X\beta|^2 = \frac{n}{n-r} \tilde{\sigma}^2$$

erwartungstreu.

Satz 2.3.7. Der Schätzer $\bar{\sigma}^2$ ist erwartungstreu für σ^2 .

Der Beweis des Satzes 2.3.7 folgt dem Beweis des Satzes 2.2.4, in dem $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ und $\hat{\sigma}^2 = \frac{1}{n-m} |Y - X\hat{\beta}|^2$ im Fall $\text{Rang}(X) = m$ betrachtet wurden. Somit ist die Aussage des Satzes 2.2.4 ein Spezialfall des Satzes 2.3.7. Führen wir die Matrix $D = \mathcal{I} - X(X^\top X)^{-1} X^\top$ ein.

Lemma 2.3.3. Für D gelten folgende Eigenschaften:

1. $D^\top = D$ (Symmetrie),
2. $D^2 = D$ (Idempotenz),
3. $DX = 0$,
4. $\text{Spur}(D) = n - r$.

Beweis. 1. Es gilt:

$$\begin{aligned} D^\top &= \left(\mathcal{I} - X(X^\top X)^- X^\top \right)^\top = \mathcal{I} - X \left((X^\top X)^- \right)^\top X^\top \\ &= \mathcal{I} - X(X^\top X)^- X^\top = D, \end{aligned}$$

weil $\left((X^\top X)^- \right)^\top$ auch eine verallgemeinerte Inverse von $X^\top X$ ist (vergleiche Lemma 2.3.2, 1.)).

2. Es gilt:

$$\begin{aligned} D^2 &= \left(\mathcal{I} - X(X^\top X)^- X^\top \right)^2 = \mathcal{I} - 2X(X^\top X)^- X^\top + \underbrace{X(X^\top X)^- X^\top X(X^\top X)^- X^\top}_{=X(\text{Lemma 2.3.2, 2.})} \\ &= \mathcal{I} - X(X^\top X)^- X^\top = D. \end{aligned}$$

$$3. \quad DX = X - \underbrace{X(X^\top X)^- X^\top X}_{=X(\text{Lemma 2.3.2, 2.})} = X - X = 0.$$

4. Es gilt:

$$\text{Spur}(D) = \text{Spur}(I) - \text{Spur} \left(X(X^\top X)^- X^\top \right) = n - \text{Spur} \left(X(X^\top X)^- X^\top \right).$$

Verwenden wir die Eigenschaft der symmetrischen idempotenten Matrizen A aus der linearen Algebra, daß $\text{Spur}(A) = \text{Rang}(A)$. Da $X(X^\top X)^- X^\top$ symmetrisch und idempotent ist, genügt es zu zeigen, daß $\text{Rang}(X(X^\top X)^- X^\top) = r$. Nach Lemma 2.3.2 2.) gilt:

$$\begin{aligned} \text{Rang}(X) &= r = \text{Rang}(X(X^\top X)^- X^\top X) \\ &\leq \min \left\{ \text{Rang}(X(X^\top X)^- X^\top), \underbrace{\text{Rang}(X)}_{=r} \right\} \\ &\leq \text{Rang} \left(X(X^\top X)^- X^\top \right) \leq \text{Rang}(X) = r \\ &\implies \text{Rang} \left(X(X^\top X)^- X^\top \right) = r \\ &\implies \text{Spur} \left(X(X^\top X)^- X^\top \right) = r. \end{aligned}$$

□

Beweis des Satzes 2.3.7. Mit Hilfe des Lemmas 2.3.3 bekommt man

$$\begin{aligned}\bar{\sigma}^2 &= \frac{1}{n-r} |Y - X\bar{\beta}|^2 = \frac{1}{n-r} \left| Y - X(X^\top X)^{-1} X^\top Y \right|^2 = \frac{1}{n-r} |DY|^2 \\ &= \frac{1}{n-r} \left| \underbrace{DX}_{=0} \beta + D\varepsilon \right|^2 = \frac{1}{n-r} |D\varepsilon|^2 = \frac{1}{n-r} \varepsilon^\top \underbrace{D^\top D}_{=D^2=D} \varepsilon = \frac{1}{n-r} \varepsilon^\top D\varepsilon.\end{aligned}$$

Deshalb gilt:

$$\begin{aligned}\mathbb{E} \bar{\sigma}^2 &= \frac{1}{n-r} \mathbb{E} \left(\varepsilon^\top D\varepsilon \right) = \frac{1}{n-r} \mathbb{E} \text{Spur} \left(\varepsilon^\top D\varepsilon \right) = \frac{1}{n-r} \text{Spur} \left(D \cdot \mathbb{E} \left(\underbrace{\varepsilon \varepsilon^\top}_{\sigma^2 \mathcal{I}} \right) \right) \\ &= \frac{\sigma^2}{n-r} \cdot \text{Spur}(D) = \sigma^2 \text{ nach Lemma 2.3.3, 4.)}, \text{ weil } \mathbb{E} \varepsilon \varepsilon^\top = \sigma^2 \mathcal{I} \\ &\text{wegen } \varepsilon \sim N(0, \sigma^2 \mathcal{I}).\end{aligned}$$

□

Satz 2.3.8. Es gelten folgende Verteilungseigenschaften:

1. $\bar{\beta} \sim N \left((X^\top X)^{-1} X^\top X\beta, \sigma^2 (X^\top X)^{-1} (X^\top X)^{-1} \right)^\top$,
2. $\frac{(n-r)\bar{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2$,
3. $\bar{\beta}$ und $\bar{\sigma}^2$ sind unabhängig.

Beweis. 1. Es gilt:

$$\bar{\beta} = (X^\top X)^{-1} X^\top Y = (X^\top X)^{-1} X^\top (X\beta + \varepsilon) = \underbrace{(X^\top X)^{-1} X^\top X\beta}_{=\mu} + \underbrace{(X^\top X)^{-1} X^\top}_{=A} \varepsilon$$

und mit der Definition von $N(\cdot, \cdot)$ bekommt man

$$\begin{aligned}\bar{\beta} &\sim N \left(\mu, \sigma^2 AA^\top \right) = N \left((X^\top X)^{-1} X^\top X\beta, \sigma^2 (X^\top X)^{-1} X^\top X (X^\top X)^{-1} \right)^\top \\ &\text{mit } AA^\top = (X^\top X)^{-1} X^\top X (X^\top X)^{-1}\end{aligned}$$

2. Es gilt $\bar{\sigma}^2 = \frac{1}{n-r} \varepsilon^\top D\varepsilon$ aus dem Beweis des Satzes 2.3.7. Deshalb

$$\frac{(n-r)\bar{\sigma}^2}{\sigma^2} = \underbrace{\left(\frac{\varepsilon}{\sigma} \right)^\top}_{\sim N(0, \mathcal{I})} D \left(\frac{\varepsilon}{\sigma} \right) \stackrel{(\text{Satz 2.1.8})}{\sim} \chi_{n-r}^2.$$

3. Betrachten wir $A\varepsilon$ und $\varepsilon^\top D\varepsilon$. Es genügt zu zeigen, daß sie unabhängig sind, um die Unabhängigkeit von $\bar{\beta}$ und $\bar{\sigma}^2$ zu beweisen, weil $\bar{\beta} = \mu + A\varepsilon$, $\bar{\sigma}^2 = \frac{1}{n-r} \varepsilon^\top D\varepsilon$. Es gilt: $A \cdot \sigma^2 \mathcal{I} \cdot D = 0$. Nach Satz 2.1.9 sind dann $A\varepsilon$ und $\varepsilon^\top D\varepsilon$ unabhängig.

□

2.3.5 Hypothesentests

Betrachten wir die Hypothesen $H_0 : H\beta = d$ vs. $H_1 : H\beta \neq d$, wobei H eine $(s \times m)$ -Matrix ($s \leq m$) mit $\text{Rang}(H) = s$ ist, und $d \in \mathbb{R}^s$.

Im Satz 2.2.10 haben wir im Fall $\text{Rang}(X) = r = m$ folgende Testgröße dafür betrachtet:

$$T = \frac{(H\hat{\beta} - d)^\top (H(X^\top X)^{-1}H^\top)^{-1}(H\hat{\beta} - d)}{s\hat{\sigma}^2} \stackrel{(H_0)}{\sim} F_{s,n-m}.$$

Im allgemeinen Fall betrachten wir

$$T = \frac{(H\bar{\beta} - d)^\top (H(X^\top X)^-H^\top)^{-1}(H\bar{\beta} - d)}{s\bar{\sigma}^2}. \quad (2.3.5)$$

Wir wollen zeigen, daß $T \stackrel{(H_0)}{\sim} F_{s,n-r}$. Dann wird H_0 verworfen, falls $T > F_{s,n-r,1-\alpha}$. Dies ist ein Test zum Niveau $\alpha \in (0, 1)$.

Definition 2.3.3. Die Hypothese $H_0 : H\beta = d$ heißt *testbar*, falls alle Koordinaten des Vektors $H\beta$ schätzbare Funktionen sind.

Satz 2.3.3 gibt Bedingungen an H an, unter denen $H_0 : H\beta = d$ testbar ist. Diese werden im folgendem Lemma formuliert:

Lemma 2.3.4. Die Hypothese $H_0 : H\beta = d$ ist testbar genau dann, wenn

1. $\exists (s \times n)$ -Matrix $C : H = CX$, oder
2. $H(X^\top X)^-X^\top X = H$.

Wir zeigen, daß die Testgröße T in (2.3.5) wohldefiniert ist, das heißt, die $(s \times s)$ -Matrix $H(X^\top X)^-H^\top$ positiv definit und damit invertierbar ist. Aus Folgerung 2.3.1 haben wir $X^\top X = P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} P^{-1}$ für eine $(m \times m)$ -Matrix P , die invertierbar und symmetrisch ist. Deshalb gilt

$$(X^\top X)^- = P \cdot \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & \mathcal{I}_{m-r} \end{pmatrix} P = P \cdot P,$$

das heißt, daß es eine eindeutige verallgemeinerte Inverse von $X^\top X$ mit dieser Darstellung gibt. Daraus folgt, daß die $(s \times s)$ -Matrix $HPPH^\top = (PH^\top)^\top \cdot PH^\top$ positiv definit ist, weil $\text{Rang}(PH^\top) = s$. Sei nun $(X^\top X)^-$ eine beliebige verallgemeinerte Inverse von $X^\top X$. Dann ist mit Lemma 2.3.4

$$H(X^\top X)^-H^\top = CX(X^\top X)^-X^\top C^\top = CXPPX^\top C^\top = HPPH^\top,$$

denn $X(X^\top X)^-X^\top$ ist invariant bezüglich der Wahl von $(X^\top X)^-$, laut Beweis des Satzes 2.3.6. Also ist $H(X^\top X)^-H^\top$ positiv definit für eine beliebige verallgemeinerte Inverse $(X^\top X)^-$ und die Testgröße T somit wohldefiniert.

Satz 2.3.9. Falls $H_0 : H\beta = d$ testbar ist, dann gilt $T \stackrel{(H_0)}{\sim} F_{s,n-r}$.

Beweis. Ähnlich, wie in Satz 2.2.10 gilt

$$H\bar{\beta} - d = H(X^\top X)^{-1} X^\top (X\beta + \varepsilon) - d = \underbrace{H(X^\top X)^{-1} X^\top X\beta - d}_{=\mu} + \underbrace{H(X^\top X)^{-1} X^\top \varepsilon}_{=B}$$

Zeigen wir, daß $\mu \stackrel{(H_0)}{=} 0$.

$$\mu \stackrel{(\text{Lemma 2.3.4})}{=} C \cdot \underbrace{X(X^\top X)^{-1} X^\top X}_{=X \text{ (Lemma 2.3.2, 2.)}} \cdot \beta - d = CX\beta - d = H\beta - d \stackrel{(H_0)}{=} 0.$$

Nach Satz 2.3.8 sind $(H\bar{\beta} - d)^\top (H(X^\top X)^{-1} H^\top)^{-1} (H\bar{\beta} - d)$ und $s \cdot \bar{\sigma}^2$ unabhängig, $\frac{(n-r)\bar{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2$. Also bleibt nur noch zu zeigen, daß

$$\left(\underbrace{H\bar{\beta} - d}_{=\varepsilon^\top B^\top} \right)^\top \left(H(X^\top X)^{-1} H^\top \right)^{-1} \left(\underbrace{H\bar{\beta} - d}_{=B\varepsilon} \right) \stackrel{(H_0)}{\sim} \chi_s^2.$$

Es gilt

$$\begin{aligned} & \varepsilon^\top B^\top \left(H(X^\top X)^{-1} H^\top \right)^{-1} B\varepsilon \\ &= \varepsilon^\top \underbrace{X \left((X^\top X)^{-1} \right)^\top H^\top \left(H(X^\top X)^{-1} H^\top \right)^{-1} H(X^\top X)^{-1} X^\top}_{A} \varepsilon \end{aligned}$$

Man kann leicht zeigen, daß A symmetrisch, idempotent und $\text{Rang}(A) = s$ ist. Zeigen wir zum Beispiel die Idempotenz:

$$\begin{aligned} A^2 &= X \left((X^\top X)^{-1} \right)^\top H^\top \left(H(X^\top X)^{-1} H^\top \right)^{-1} \underbrace{H(X^\top X)^{-1} X^\top X}_{H \text{ (Lemma 2.3.4, 2.)}} \left((X^\top X)^{-1} \right)^\top H^\top \\ &\quad \cdot \left(H(X^\top X)^{-1} H^\top \right)^{-1} H(X^\top X)^{-1} X^\top \\ &= X \left((X^\top X)^{-1} \right)^\top H^\top \left(H(X^\top X)^{-1} H^\top \right)^{-1} H(X^\top X)^{-1} X^\top = A, \end{aligned}$$

weil $\left((X^\top X)^{-1} \right)^\top$ auch eine verallgemeinerte Inverse von $X^\top X$ ist (nach Lemma 2.3.2). Somit hängt auch $H(X^\top X)^{-1} H^\top = CX(X^\top X)^{-1} X^\top C^\top$ nicht von der Wahl von $(X^\top X)^{-1}$ ab, vgl. den Beweis des Satzes 2.3.6. Nach Satz 2.1.8 ist $\frac{\varepsilon^\top}{\sigma} A \frac{\varepsilon}{\sigma} \sim \chi_s^2$, wegen $\varepsilon \sim N(0, \sigma^2 \mathcal{I})$ und somit $T \stackrel{H_0}{\sim} F_{s,n-r}$. \square

2.3.6 Konfidenzbereiche

Ähnlich wie in Abschnitt 2.2.5 werden wir Konfidenzbereiche für unterschiedliche Funktionen vom Parametervektor β angeben. Aus dem Satz 2.3.9 ergibt sich unmittelbar folgender Konfidenzbereich zum Niveau $1 - \alpha \in (0, 1)$:

Folgerung 2.3.1. Sei $Y = X\beta + \varepsilon$ ein multivariates Regressionsmodell mit $\text{Rang}(X) = r < m$, H eine $(s \times m)$ -Matrix mit $\text{Rang}(H) = s$, $s \in \{1, \dots, m\}$ und $H_0 : H\beta = d$ testbar $\forall d \in \mathbb{R}^s$. Dann ist

$$\left\{ d \in \mathbb{R}^s : \frac{(H\bar{\beta} - d)^\top (H(X^\top X)^- H^\top)^{-1} (H\bar{\beta} - d)}{s \cdot \bar{\sigma}^2} \leq F_{s, n-r, 1-\alpha} \right\}$$

ein Konfidenzbereich für $H\beta$ zum Niveau $1 - \alpha$.

Folgerung 2.3.2. Sei $h^\top \beta$ eine schätzbare lineare Funktion von β , $h \in \mathbb{R}^m$. Dann ist

$$\left(h^\top \bar{\beta} - t_{n-r, 1-\alpha/2} \cdot \bar{\sigma} \sqrt{h^\top (X^\top X)^- h}, h^\top \bar{\beta} + t_{n-r, 1-\alpha/2} \cdot \bar{\sigma} \sqrt{h^\top (X^\top X)^- h} \right)$$

ein Konfidenzintervall für $h^\top \beta$ zum Niveau $1 - \alpha$.

Beweis. Setzen wir $s = 1$ und $H = h^\top$. Aus Satz 2.3.9 folgt

$$\begin{aligned} T &= \frac{(h^\top \bar{\beta} - d)^\top (h^\top (X^\top X)^- h)^{-1} (h^\top \bar{\beta} - d)}{\bar{\sigma}^2} = \frac{(h^\top \bar{\beta} - d) (h^\top \bar{\beta} - d)}{\bar{\sigma}^2 (h^\top (X^\top X)^- h)} \\ &= \frac{(h^\top \bar{\beta} - d)^2}{\bar{\sigma}^2 (h^\top (X^\top X)^- h)} \sim F_{1, n-r} \end{aligned}$$

unter der Voraussetzung $h^\top \beta = d$, weil $h^\top (X^\top X)^- h$ eindimensional (eine Zahl) ist. Deshalb gilt

$$\sqrt{T} = \frac{h^\top \beta - h^\top \bar{\beta}}{\bar{\sigma} \sqrt{h^\top (X^\top X)^- h}} \sim t_{n-r}$$

und somit

$$\mathbb{P} \left(-t_{n-r, 1-\alpha/2} \leq \sqrt{T} \leq t_{n-r, 1-\alpha/2} \right) = 1 - \alpha.$$

Daraus folgt das obige Konfidenzintervall. \square

Man kann sogar eine stärkere Version von 2.3.2 beweisen, die für alle h aus einem linearen Unterraum gilt:

Satz 2.3.10 (Konfidenzband von Scheffé). Sei $H = (h_1, \dots, h_s)^\top$, $h_1, \dots, h_s \in \mathbb{R}^m$, $1 \leq s \leq m$ und $H_0 : H\beta = d$ testbar $\forall d \in \mathbb{R}^s$. Sei $\text{Rang}(H) = s$ und $\mathcal{L} = \langle h_1, \dots, h_s \rangle$ der lineare Unterraum, der von den Vektoren h_1, \dots, h_s aufgespannt wird. Dann gilt:

$$\mathbb{P} \left(\max_{h \in \mathcal{L}} \left\{ \frac{(h^\top \beta - h^\top \bar{\beta})^2}{\bar{\sigma}^2 h^\top (X^\top X)^{-1} h} \right\} \leq s F_{s, n-r, 1-\alpha} \right) = 1 - \alpha$$

Somit ist

$$\left[h^\top \bar{\beta} - \sqrt{s F_{s, n-r, 1-\alpha}} \cdot \bar{\sigma} \sqrt{h^\top (X^\top X)^{-1} h}, h^\top \bar{\beta} + \sqrt{s F_{s, n-r, 1-\alpha}} \cdot \bar{\sigma} \sqrt{h^\top (X^\top X)^{-1} h} \right]$$

ein (gleichmäßiges bzgl. $h \in \mathcal{L}$) Konfidenzintervall für $h^\top \beta$.

Beweis. Aus dem Satz 2.3.9 folgt $\forall \alpha \in (0, 1)$:

$$\mathbb{P} \left(\underbrace{(H\bar{\beta} - H\beta)^\top (H(X^\top X)^{-1} H^\top)^{-1} (H\bar{\beta} - H\beta)}_{T_1} \leq s \cdot \bar{\sigma}^2 F_{s, n-r, 1-\alpha} \right) = 1 - \alpha.$$

Falls wir zeigen können, daß

$$T_1 = \max_{x \in \mathbb{R}^s, x \neq 0} \left\{ \frac{(x^\top (H\bar{\beta} - H\beta))^2}{x^\top (H(X^\top X)^{-1} H^\top) x} \right\}, \quad (2.3.6)$$

dann ist der Satz bewiesen, denn

$$\begin{aligned} 1 - \alpha &= \mathbb{P} \left(T_1 \leq \underbrace{s \bar{\sigma}^2 F_{s, n-r, 1-\alpha}}_t \right) = \mathbb{P} \left(\max_{x \in \mathbb{R}^s, x \neq 0} \left\{ \frac{(x^\top (H\bar{\beta} - H\beta))^2}{x^\top (H(X^\top X)^{-1} H^\top) x} \right\} \leq t \right) \\ &= \mathbb{P} \left(\max_{x \in \mathbb{R}^s, x \neq 0} \left\{ \frac{((H^\top x)^\top \bar{\beta} - (H^\top x)^\top \beta)^2}{(H^\top x)^\top (X^\top X)^{-1} (H^\top x)} \right\} \leq t \right) \quad \text{und weil } H^\top x = h \in \mathcal{L} \\ &= \mathbb{P} \left(\max_{h \in \mathcal{L}} \left\{ \frac{(h^\top \bar{\beta} - h^\top \beta)^2}{h^\top (X^\top X)^{-1} h} \right\} \leq s \bar{\sigma}^2 F_{s, n-r, 1-\alpha} \right). \end{aligned}$$

Also, zeigen wir die Gültigkeit von (2.3.6). Es genügt zu zeigen, daß T_1 die obere Schranke von

$$\frac{(x^\top (H\bar{\beta} - H\beta))^2}{x^\top (H(X^\top X)^{-1} H^\top) x}$$

darstellt, die auch angenommen wird. Da $H(X^\top X)^{-1} H^\top$ positiv definit ist und invertierbar, existiert eine invertierbare $(s \times s)$ -Matrix B mit der Eigenschaft $BB^\top =$

$H(X^\top X)^{-1}H^\top$. Dann gilt

$$\begin{aligned} \left(x^\top(H\bar{\beta} - H\beta)\right)^2 &= \left(\underbrace{x^\top B}_{(B^\top x)^\top} \cdot B^{-1}(H\bar{\beta} - H\beta)\right)^2 \\ &\leq |B^\top x|^2 \cdot |B^{-1}(H\bar{\beta} - H\beta)|^2 \quad (\text{wegen der Ungleichung von Cauchy-Schwarz}) \\ &= x^\top B B^\top x (H\bar{\beta} - H\beta)^\top \cdot \underbrace{(B^{-1})^\top B^{-1}}_{=(B^\top)^{-1}B^{-1}=(BB^\top)^{-1}} (H\bar{\beta} - H\beta) \\ &= x^\top H(X^\top X)^{-1}H^\top x \cdot (H\bar{\beta} - H\beta)^\top \left(H(X^\top X)^{-1}H^\top\right)^{-1} (H\bar{\beta} - H\beta). \end{aligned}$$

Somit gilt

$$\frac{(x^\top(H\bar{\beta} - H\beta))^2}{x^\top(H(X^\top X)^{-1}H^\top)x} \leq (H\bar{\beta} - H\beta)^\top \left(H(X^\top X)^{-1}H^\top\right)^{-1} (H\bar{\beta} - H\beta) = T_1.$$

Man kann leicht prüfen, daß diese Schranke für $x = (H(X^\top X)^{-1}H^\top)^{-1} (H\bar{\beta} - H\beta)$ angenommen wird. \square

2.3.7 Einführung in die Varianzanalyse

In diesem Abschnitt geben wir ein Beispiel für die Verwendung linearer Modelle mit Design-Matrix, die keinen vollen Rang besitzt. Dabei handelt es sich um die Aussage der *Variabilität der Erwartungswerte* in der Stichprobe $Y = (Y_1, \dots, Y_n)^\top$, die auf englisch *analysis of variance*, kurz *ANOVA*, heißt. Später werden wir auch denselben Begriff *Varianzanalyse* dafür verwenden.

Betrachten wir zunächst die *einfaktorielle Varianzanalyse*, bei der man davon ausgeht, daß die Stichprobe (Y_1, \dots, Y_n) in k homogene Teilklassen $(Y_{ij}, j = 1, \dots, n_i), i = 1, \dots, k$ zerlegbar ist, mit den Eigenschaften:

$$1. \mathbb{E}(Y_{ij}) = \mu_i = \mu + \alpha_i, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k.$$

$$2. n_i > 1, \quad i = 1, \dots, k, \quad \sum_{i=1}^k n_i = n, \quad \sum_{i=1}^k n_i \alpha_i = 0.$$

Dabei ist μ ein Faktor, der allen Klassen gemeinsam ist, und α_i verkörpert die *klassenspezifischen Differenzen* zwischen den Erwartungswerten μ_1, \dots, μ_k . Die Nummer $i = 1, \dots, k$ der Klassen wird als *Stufe eines Einflussfaktors* (zum Beispiel die Dosis eines Medikaments in einer klinischen Studie) und $\alpha_i, i = 1, \dots, k$ als *Effekt* der i -ten Stufe gedeutet. Die Nebenbedingung $\sum_{i=1}^k n_i \alpha_i = 0$ bewirkt, daß die Umrechnung

$(\mu_1, \dots, \mu_k) \longleftrightarrow (\mu, \alpha_1, \dots, \alpha_k)$ eindeutig wird und daß $\mu = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbb{E} Y_{ij}$. Es wird vorausgesetzt, daß μ_i mit unkorrelierten Meßfehlern ε_{ij} gemessen werden kann, das heißt

$$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (2.3.7)$$

$$\mathbb{E} \varepsilon_{ij} = 0, \quad \text{Var} \varepsilon_{ij} = \sigma^2, \quad \varepsilon_{ij} \text{ unkorreliert}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i. \quad (2.3.8)$$

Es soll die *klassische ANOVA-Hypothese* getestet werden, daß *keine* Variabilität in den Erwartungswerten μ_i auffindbar ist:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k,$$

was bedeutet, daß

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k.$$

Aus der Nebenbedingung

$$\sum_{i=1}^k n_i \alpha_i = 0.$$

folgt: $\alpha_i = 0$

Die Problemstellung (2.3.7) kann in der Form der multivariaten linearen Regression folgendermaßen umgeschrieben werden:

$$Y = X\beta + \varepsilon, \text{ wobei } Y = (Y_{11}, \dots, Y_{1n_1}, Y_{21}, \dots, Y_{2n_2}, \dots, Y_{k1}, \dots, Y_{kn_k})^\top,$$

$$\beta = (\mu, \alpha_1, \dots, \alpha_k)^\top,$$

$$\varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \dots, \varepsilon_{k1}, \dots, \varepsilon_{kn_k})^\top,$$

$$X = \begin{pmatrix} 1 & 1 & 0 & \dots & \dots & 0 \\ 1 & 1 & 0 & \dots & \dots & 0 \\ \vdots & & & & & \\ 1 & 1 & 0 & \dots & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 1 & 0 & \dots & \dots & 0 & 1 \\ \vdots & & & & & \\ 1 & 0 & \dots & \dots & 0 & 1 \end{pmatrix} \begin{matrix} \left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} n_1 \\ \left. \begin{matrix} \\ \\ \end{matrix} \right\} n_2 \\ \vdots \\ \left. \begin{matrix} \\ \\ \end{matrix} \right\} n_k \end{matrix}$$

Die $(n \times (k + 1))$ -Matrix X hat den Rang $k < m = k + 1$, somit ist die Theorie von Abschnitt 2.3 auf dieses Modell komplett anwendbar.

Übungsaufgabe 2.3.1. Zeigen Sie, dass die ANOVA-Hypothese

$$H_0: \alpha_i = 0, \quad \forall i = 1, \dots, k$$

nicht testbar ist!

Um eine äquivalente testbare Hypothese aufzustellen, benutzt man

$$H_0: \alpha_1 - \alpha_2 = 0, \dots, \alpha_1 - \alpha_k = 0 \quad \text{bzw.} \quad H_0: H\beta = 0$$

für die $(k-1) \times (k+1)$ -Matrix

$$H = \begin{pmatrix} 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 \\ \vdots & & & & & \\ 0 & 1 & 0 & \dots & -1 & 0 \\ 0 & 1 & 0 & \dots & 0 & -1 \end{pmatrix}.$$

(Zeigen Sie es!)

Bei der *zweifaktoriellen Varianzanalyse* wird die Stichprobe (Y_1, \dots, Y_n) in Abhängigkeit von 2 Faktoren in $k_1 \cdot k_2$ homogene Gruppen aufgeteilt:

$$Y_{i_1 i_2 j}, \quad j = 1, \dots, n_{i_1 i_2}$$

für $i_1 = 1, \dots, k_1, i_2 = 1, \dots, k_2$, sodaß

$$\sum_{i_1=1}^{k_1} \sum_{i_2=1}^{k_2} n_{i_1 i_2} = n.$$

Hier wird angenommen, daß

$$\mathbb{E} Y_{i_1 i_2 j} = \mu_{i_1 i_2} = \mu + \alpha_{i_1} + \beta_{i_2} + \gamma_{i_1 i_2}, \quad i_1 = 1, \dots, k_1, i_2 = 1, \dots, k_2,$$

somit stellt man folgendes lineares Modell auf:

$$Y_{i_1 i_2 j} = \mu_{i_1 i_2} + \varepsilon_{i_1 i_2 j} = \mu + \alpha_{i_1} + \beta_{i_2} + \gamma_{i_1 i_2} + \varepsilon_{i_1 i_2 j}, \\ j = 1, \dots, n_{i_1 i_2}, i_1 = 1, \dots, k_1, i_2 = 1, \dots, k_2.$$

Übungsaufgabe 2.3.2. Schreiben Sie die Design-Matrix X für diesen Fall explizit auf! Zeigen Sie, daß sie wieder keinen vollen Rang besitzt.

*

3 Verallgemeinerte lineare Modelle

Eine andere Klasse von Regressionsmodellen erlaubt einerseits einen beliebigen funktionalen Zusammenhang g zwischen dem Mittelwert der Zielvariablen $\mathbb{E} Y_i$ und dem linearen Teil $X\beta$, der aus linearen Kombinationen der Einträge der Designmatrix $X = (x_{ij})$ und des Parametervektors $\beta = (\beta_1, \dots, \beta_m)^\top$ besteht; andererseits lässt sie andere Verteilungen von Y_i zu, die nicht notwendigerweise auf der Normalverteilung (und Funktionen davon) basieren. So ist es möglich, Daten Y_i zu betrachten, die eine endliche Anzahl von Ausprägungen haben (z.B. „Ja“ und „Nein“ in ökonomischen Meinungsumfragen). Die Klasse aller möglichen Verteilungen wird durch die sog. *Exponentialfamilie* begrenzt, die wir in Kürze einführen werden.

Sei Y_1, \dots, Y_n eine Zufallsstichprobe der Zielvariablen des Modells und sei

$$X = (x_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$$

die Designmatrix der Ausgangsvariablen, die hier nicht zufällig sind.

Definition 3.0.4. Das *verallgemeinerte lineare Modell* ist gegeben durch

$$(g(\mathbb{E} Y_1), \dots, g(\mathbb{E} Y_n))^\top = X\beta \quad \text{mit } \beta = (\beta_1, \dots, \beta_m)^\top, \quad (3.0.1)$$

wobei $g : G \subset \mathbb{R} \rightarrow \mathbb{R}$ die sog. *Linkfunktion* mit dem Definitionsbereich G ist. Der Rang $(X) = m$.

Unter der Annahme, dass g explizit bekannt ist, soll hier der Parametervektor β aus (Y_1, \dots, Y_n) geschätzt werden. Wir setzen voraus, dass $Y_i, i = 1, \dots, n$, unabhängig, aber nicht unbedingt identisch verteilt sind. Ihre Verteilung gehört jedoch zur folgenden Klasse von Verteilungen:

3.1 Exponentialfamilie von Verteilungen

Definition 3.1.1. Die Verteilung einer Zufallsvariable Y gehört zur *Exponentialfamilie*, falls es Funktionen $a : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ und $b : \Theta \rightarrow \mathbb{R}$ gibt, für die

- im *absolutstetigen Fall* die Dichte von Y gegeben ist durch

$$f_\theta(y) = \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\}, \quad y \in \mathbb{R} \quad (3.1.1)$$

- im *diskreten Fall* die Zähldichte von Y gegeben ist durch

$$P_\theta(Y = y) = \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\}, y \in C, \quad (3.1.2)$$

wobei C der (höchstens) abzählbare *Wertebereich* von Y , τ^2 der sog. *Störparameter*, $\theta \in \Theta \subset \mathbb{R}$ ein Parameter und

$$\Theta = \left\{ \theta \in \mathbb{R} : \int_{\mathbb{R}} \exp \left\{ \frac{y\theta + a(y, \tau)}{\tau^2} \right\} dy < \infty \right\}$$

bzw. im diskreten Fall:

$$\Theta = \left\{ \theta \in \mathbb{R} : \sum_{y \in C} \exp \left\{ \frac{y\theta + a(y, \tau)}{\tau^2} \right\} < \infty \right\}$$

der natürliche Parameterraum ist, der mindestens zwei verschiedene Elemente enthält.

Lemma 3.1.1. Θ ist ein Intervall.

Beweis. Zeigen wir, dass $\Theta \subset \mathbb{R}$ konvex ist. Dann ist es notwendigerweise ein (möglicherweise unendliches) Intervall. Für beliebige $\theta_1, \theta_2 \in \Theta$ (mindestens ein solches Paar gibt es nach Definition 3.1.1) zeigen wir, dass $\alpha\theta_1 + (1 - \alpha)\theta_2 \in \Theta$ für alle $\alpha \in (0, 1)$. Nehmen wir an, dass die Verteilung von Y absolut stetig ist. Da $\theta_i \in \Theta$, es gilt

$$\int_{\mathbb{R}} \exp \left\{ \frac{1}{\tau^2} (y\theta_i + a(y, \tau)) \right\} dy < \infty, \quad i = 1, 2.$$

Durch die offensichtliche Ungleichung

$$\alpha x_1 + (1 - \alpha)x_2 \leq \max\{x_1, x_2\}, \quad x_1, x_2 \in \mathbb{R} \quad \alpha \in (0, 1)$$

erhalten wir

$$\begin{aligned} & \exp \left\{ \frac{1}{\tau^2} (y(\alpha\theta_1 + (1 - \alpha)\theta_2) + a(y, \tau)) \right\} \\ &= \exp \left\{ \alpha \frac{1}{\tau^2} (y\theta_1 + a(y, \tau)) + (1 - \alpha) \frac{1}{\tau^2} (y\theta_2 + a(y, \tau)) \right\} \\ &\leq \max_{i=1,2} \exp \left\{ \frac{1}{\tau^2} (y\theta_i + a(y, \tau)) \right\} \leq \exp \left\{ \frac{1}{\tau^2} (y\theta_1 + a(y, \tau)) \right\} + \exp \left\{ \frac{1}{\tau^2} (y\theta_2 + a(y, \tau)) \right\}, \end{aligned}$$

so dass

$$\int_{\mathbb{R}} \exp \left\{ \frac{1}{\tau^2} (y(\alpha\theta_1 + (1 - \alpha)\theta_2) + a(y, \tau)) \right\} dy \leq \sum_{i=1}^2 \int_{\mathbb{R}} \exp \left\{ \frac{1}{\tau^2} (y\theta_i + a(y, \tau)) \right\} dy < \infty$$

nach Voraussetzungen des Lemmas.

$$\Rightarrow \alpha\theta_1 + (1 - \alpha)\theta_2 \in \Theta,$$

und Θ ist ein Intervall. □

Beispiel 3.1.1. Welche Verteilungen gehören zur Exponentialfamilie?

1. **Normalverteilung:** Falls $Y \sim \mathcal{N}(\mu, \sigma^2)$, dann ist der Erwartungswert μ der uns interessierende Parameter, σ^2 ist dagegen der Störparameter. Es gilt:

$$\begin{aligned} f_\mu(y) &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \\ &= \exp \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \left(\frac{y^2}{\sigma^2} - \frac{2y\mu}{\sigma^2} + \frac{\mu^2}{\sigma^2} \right) \right\} \\ &= \exp \left\{ \frac{1}{\sigma^2} \left(y\mu - \frac{y^2}{2} - \left(\frac{\mu^2}{2} + \frac{\sigma^2}{2} \log(2\pi\sigma^2) \right) \right) \right\}, \end{aligned}$$

so dass

$$\theta = \mu, \quad \tau = \sigma, \quad a(y, \tau) = -\frac{y^2}{2} - \frac{\sigma^2}{2} \log(2\pi\sigma^2) \quad \text{und} \quad b(\mu) = b(\theta) = \frac{\mu^2}{2}.$$

2. **Bernoulli-Verteilung:** $Y \sim \text{Bernoulli}(p)$, $p \in [0; 1]$.

Sie wird etwa im Falle von Meinungsumfragen in der Marktforschung verwendet, in denen

$$Y = \begin{cases} 1, & \text{falls die Antwort „ja“} \\ 0, & \text{falls die Antwort „nein“} \end{cases} \text{ auf eine Frage der Enquete gegeben wurde.}$$

Dabei ist die Wahrscheinlichkeit $P(Y = 1) = p$, $P(Y = 0) = 1 - p$. Dann gilt für $y \in \{0, 1\}$:

$$\begin{aligned} P_\theta(Y = y) &= p^y(1 - p)^{1-y} = e^{y \log p + (1-y) \log(1-p)} \\ &= e^{y \log \frac{p}{1-p} - (-\log(1-p))}. \end{aligned}$$

Somit gehört die Bernoulli-Verteilung zur Exponentialfamilie mit

$$\theta = \log \frac{p}{1-p}, \quad \tau = 1, \quad a(y, \tau) = 0, \quad b(\theta) = -\log(1-p) = \log(1 + e^\theta).$$

3. **Poisson-Verteilung:** Falls $Y \sim \text{Poisson}(\lambda)$, $\lambda > 0$, dann gilt für $y \in \mathbb{N}_0$

$$P_\theta(Y = y) = e^{-\lambda} \cdot \frac{\lambda^y}{y!} = e^{y \log \lambda - \log(y!) - \lambda} \quad .$$

Somit gehört die Poisson-Verteilung zur Exponentialfamilie mit

$$\theta = \log \lambda, \quad \tau = 1, \quad a(y, \tau) = -\log(y!), \quad b(\theta) = \lambda = e^\theta .$$

Lemma 3.1.2. Falls die Verteilung von Y zur Exponentialfamilie gehört, $\mathbb{E}Y^2 < \infty$ und $b : \Theta \rightarrow \mathbb{R}$ zweimal stetig differenzierbar ist mit $b''(\theta) > 0$ für alle $\theta \in \Theta$, dann gilt

$$\mathbb{E}Y = b'(\theta), \quad \text{Var} Y = \tau^2 b''(\theta) .$$

Beweis. 1. Führen wir den Beweis für den Fall der absolut stetigen Verteilung von Y . Der diskrete Fall läßt sich analog behandeln, wenn man das \int -Zeichen durch \sum ersetzt. Es gilt

$$\begin{aligned} \mathbb{E}Y &= \int_{\mathbb{R}} y f_\theta(y) dy = \int_{\mathbb{R}} y \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\} dy \\ &= e^{-\frac{b(\theta)}{\tau^2}} \cdot \tau^2 \int_{\mathbb{R}} \frac{\partial}{\partial \theta} \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau)) \right\} dy \\ &= e^{-\frac{b(\theta)}{\tau^2}} \cdot \tau^2 \frac{\partial}{\partial \theta} \int_{\mathbb{R}} \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau)) \right\} dy \\ &= e^{-\frac{b(\theta)}{\tau^2}} \cdot \tau^2 \frac{\partial}{\partial \theta} \left(\underbrace{e^{\frac{b(\theta)}{\tau^2}} \int_{\mathbb{R}} \exp \left\{ \frac{1}{\tau^2} (y\theta + a(y, \tau) - b(\theta)) \right\} dy}_{\int_{\mathbb{R}} f_\theta(y) dy = 1} \right) \\ &= e^{-\frac{b(\theta)}{\tau^2}} \tau^2 \frac{\partial}{\partial \theta} \left(e^{\frac{b(\theta)}{\tau^2}} \right) = e^{-\frac{b(\theta)}{\tau^2}} \cdot \tau^2 \frac{b'(\theta)}{\tau^2} e^{\frac{b(\theta)}{\tau^2}} = b'(\theta). \end{aligned}$$

2. Es bleibt noch zu zeigen:

Übungsaufgabe 3.1.1. Beweisen Sie die Formel

$$\text{Var} Y = \tau^2 b''(\theta) \quad (\text{analog zu 1}).$$

□

3.2 Linkfunktion

Die Zielgrößen Y_i , $i = 1, \dots, n$ seien also unabhängig verteilt mit einer Verteilung, die zur Exponentialfamilie gehört und einer (Zähl)Dichte wie in (3.1.1) bzw. (3.1.2). Setzen wir voraus, dass $b : \Theta \rightarrow \mathbb{R}$ zweimal stetig differenzierbar ist mit $b''(\theta) > 0$ für alle $\theta \in \Theta$. Sei ein verallgemeinertes lineares Modell (3.0.1) gegeben.

Definition 3.2.1. (Natürliche Linkfunktion)

Die Linkfunktion $g : G \rightarrow \mathbb{R}$ heißt *natürlich*, falls $g = (b')^{-1}$, $G = \{b'(\theta) : \theta \in \Theta\}$ und g zweimal stetig differenzierbar ist mit $g'(x) \neq 0$ für alle $x \in G$.

Die Frage, warum die natürliche Linkfunktion so heißt, beantwortet folgendes Lemma:

Lemma 3.2.1. Falls das verallgemeinerte lineare Modell (3.0.1) die natürliche Linkfunktion besitzt, dann gilt $(\theta_1, \dots, \theta_n)^\top = X\beta$.

Beweis. Wegen $b''(\theta) > 0$ ist $b'(\theta)$ monoton steigend, also invertierbar. Führen wir folgende Bezeichnungen ein:

$$\mu_i = \mathbb{E}Y_i, \quad \eta_i = x_i^\top \beta, \quad x_i = (x_{i1}, \dots, x_{im})^\top, \quad i = 1, \dots, n$$

Da g invertierbar ist, gilt

$$\mu_i = g^{-1}(x_i^\top \beta) = g^{-1}(\eta_i), \quad i = 1, \dots, n$$

Andererseits folgt $\mu_i = b'(\theta_i)$ aus Lemma 3.1.2, so dass

$$b'(\theta_i) = g^{-1}(\eta_i) \stackrel{\text{Definition 3.2.1}}{=} b'(\eta_i), \quad i = 1, \dots, n \quad .$$

Wegen der Monotonie von b' folgt die Behauptung $\theta_i = \eta_i$, $i = 1, \dots, n$. □

Beispiel 3.2.1. Berechnen wir die natürlichen Linkfunktionen für die Verteilungen von Beispiel 3.1.1.

1. **Normalverteilung:** da $b(\mu) = \frac{\mu^2}{2}$, gilt

$$b'(x) = \frac{2x}{2} = x \text{ und somit } g(x) = (b')^{-1}(x) = x \text{ .}$$

Die natürliche Linkfunktion ist $g(x) = x$, somit gilt hier

$$(\mu_1, \dots, \mu_n)^\top = (\mathbb{E}Y_1, \dots, \mathbb{E}Y_n)^\top = X\beta \text{ .}$$

Das ist genau der Fall der linearen Regression.

2. **Bernoulli-Verteilung:** da $b(\theta) = \log(1 + e^\theta)$, gilt

$$\begin{aligned} b'(x) &= \frac{1}{1 + e^x} \cdot e^x = y \\ \Leftrightarrow \frac{1}{e^{-x} + 1} &= y \\ \Leftrightarrow \frac{1}{y} - 1 &= e^{-x} \\ \Leftrightarrow x &= -\log \frac{1-y}{y} = \log \frac{y}{1-y} \\ \Rightarrow g(x) &= (b')^{-1}(x) = \log \frac{x}{1-x} . \end{aligned}$$

Das verallgemeinerte lineare Regressionsmodell im Falle der Bernoulli-Verteilung wird *binäre (kategoriale) Regression* genannt. Falls sie mit der natürlichen Linkfunktion verwendet wird, nennt man sie *logistische Regression*. In diesem Fall gilt

$$\begin{aligned} (p_1, \dots, p_n)^\top &= (\mathbb{E} Y_1, \dots, \mathbb{E} Y_n)^\top \\ \theta_i &= \log \frac{p_i}{1-p_i} = x_i^\top \beta, \quad i = 1, \dots, n \\ \Leftrightarrow e^{\theta_i} &= \frac{p_i}{1-p_i} \\ \Leftrightarrow p_i &= \frac{e^{\theta_i}}{1 + e^{\theta_i}} \\ \Leftrightarrow p_i &= \frac{e^{x_i^\top \beta}}{1 + e^{x_i^\top \beta}}, \quad i = 1, \dots, n . \end{aligned}$$

Das Verhältnis

$$\frac{p_i}{1-p_i} = \frac{P(Y_i = 1)}{P(Y_i = 0)}, \quad i = 1, \dots, n$$

wird in der englischsprachigen Literatur *Odds* genannt. Der Logarithmus des Odds heißt *Logit*:

$$\log \frac{p_i}{1-p_i}, \quad i = 1, \dots, n .$$

Logits sind also hier „neue Zielvariablen“, die durch Linearkombinationen $x_i^\top \beta$ geschätzt werden.

Eine alternative Linkfunktion, die oft benutzt wird, ist $g(x) = \Phi^{-1}(x)$, die *Quantilfunktion der Normalverteilung*. Sie ist keine natürliche Linkfunktion. Mit ihrer Hilfe bekommt man das sog. *Probit-Modell*:

$$p_i = \Phi(x_i^\top \beta), \quad i = 1, \dots, n .$$

3. **Poisson-Verteilung:** da $b(\theta) = e^\theta$, ist in diesem Fall

$$g(x) = (b')^{-1}(x) = \log x, \quad x > 0$$

die natürliche Linkfunktion. Somit hat das verallgemeinerte lineare Modell mit der natürlichen Linkfunktion folgende Darstellung

$$(\log \lambda_1, \dots, \log \lambda_n)^\top = X\beta \quad \text{oder} \quad \lambda_i = e^{x_i^\top \beta}, \quad i = 1, \dots, n.$$

3.3 Maximum-Likelihood-Schätzung von β

Da die (Zähl)Dichte von Y_i die Gestalt

$$\exp \left\{ \frac{1}{\tau^2} (y\theta_i + a(y, \tau) - b(\theta_i)) \right\}$$

hat und Y_i unabhängig sind, kann man die Log-Likelihood-Funktion der Stichprobe $Y = (Y_1, \dots, Y_n)$ in folgender Form aufschreiben:

$$\log L(Y, \theta) = \log \prod_{i=1}^n f_{\theta_i}(Y_i) = \frac{1}{\tau^2} \sum_{i=1}^n \left(Y_i \theta_i + a(Y_i, \tau) - b(\theta_i) \right). \quad (3.3.1)$$

Aus dem Beweis des Lemmas 3.2.1 folgt, dass

$$\theta_i = (b')^{-1}(g^{-1}(x_i^\top \beta)), \quad i = 1, \dots, n, \quad (3.3.2)$$

was bedeutet, dass die Funktion $\log L(Y, \theta)$ eine Funktion von Parameter β ist. In der Zukunft schreiben wir $\log L(Y, \beta)$, um diese Tatsache zu unterstreichen.

Unser Ziel ist es, den Maximum-Likelihood-Schätzer $\hat{\beta}$ für β zu berechnen:

$$\hat{\beta} = \operatorname{argmax}_{\beta} \log L(Y, \beta).$$

Dafür wird die notwendige Bedingung des Extremums

$$\frac{\partial \log L(Y, \beta)}{\partial \beta_i} = 0, \quad i = 1, \dots, m,$$

untersucht. Verwenden wir folgende Bezeichnungen:

$$U_i(\beta) = \frac{\partial \log L(Y, \beta)}{\partial \beta_i}, \quad i = 1, \dots, m,$$

$$U(\beta) = (U_1(\beta), \dots, U_m(\beta))^\top,$$

$$I_{ij}(\beta) = \mathbb{E} [U_i(\beta) U_j(\beta)], \quad i, j = 1, \dots, m.$$

Definition 3.3.1. 1. Die Matrix $I(\beta) = (I_{ij}(\beta))_{i,j=1}^m$ heißt *Fisher-Informationsmatrix*.

2. Führen wir die sog. *Hesse-Matrix* $W(\beta)$ als zufällige Matrix

$$W(\beta) = (W_{ij}(\beta))_{i,j=1}^m \quad \text{mit} \quad W_{ij}(\beta) = \frac{\partial^2}{\partial \beta_i \partial \beta_j} \log L(Y, \beta)$$

ein. Diese $(m \times m)$ -Matrix enthält die partiellen Ableitungen 2. Ordnung der Log-Likelihood-Funktion, die für die numerische Lösung der Maximierungsaufgabe

$$\log L(Y, \beta) \rightarrow \max_{\beta}$$

von Bedeutung sein werden.

Satz 3.3.1. Man kann zeigen, dass $U(\beta)$ und $I(\beta)$ folgende explizite Form haben:

1. Es gilt

$$U_j(\beta) = \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)}, \quad j = 1, \dots, m,$$

2. Es gilt

$$I_{jk}(\beta) = \sum_{i=1}^n x_{ij} x_{ik} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 \frac{1}{\sigma_i^2(\beta)}, \quad j, k = 1, \dots, m,$$

wobei $\eta_i = x_i^\top \beta$, $\mu_i(\beta) = g^{-1}(x_i^\top \beta)$ der Erwartungswert von Y_i und

$$\sigma_i^2(\beta) \stackrel{\text{Lemma 3.1.2}}{=} \tau^2 b''(\theta_i) \stackrel{(3.3.2)}{=} \tau^2 b''((b')^{-1}(g^{-1}(x_i^\top \beta))), \quad i = 1, \dots, n$$

die Varianz von Y_i ist.

Beweis. 1. Führen wir die Bezeichnung

$$l_i(\beta) = \frac{1}{\tau^2} (Y_i \theta_i + a(Y_i, \tau) - b(\theta_i)), \quad i = 1, \dots, n \text{ ein.}$$

Somit gilt

$$U_j(\beta) = \sum_{i=1}^n \frac{\partial l_i(\beta)}{\partial \beta_j}, \quad j = 1, \dots, m.$$

Durch die mehrfache Anwendung der Kettenregel ergibt sich

$$\frac{\partial l_i(\beta)}{\partial \beta_j} = \frac{\partial l_i(\beta)}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

Da

$$\frac{\partial l_i(\beta)}{\partial \theta_i} = \frac{1}{\tau^2} (Y_i - b'(\theta_i)) \stackrel{\text{Lemma 3.1.2}}{=} \frac{1}{\tau^2} (Y_i - \mu_i(\beta)),$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} = \left((b'(\theta_i))' \right)^{-1} = (b''(\theta_i))^{-1} \stackrel{\text{Lemma 3.1.2}}{=} \left(\frac{\sigma_i^2(\beta)}{\tau^2} \right)^{-1} = \frac{\tau^2}{\sigma_i^2(\beta)},$$

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i}$$

wegen $\mu_i = \mathbb{E}Y_i = g^{-1}(\eta_i)$,

$$\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial (x_i^\top \beta)}{\partial \beta_j} = x_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m,$$

bekommen wir

$$\begin{aligned} U_j(\beta) &= \frac{1}{\tau^2} \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \cdot \frac{\tau^2}{\sigma_i^2(\beta)} \cdot \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \\ &= \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot \frac{1}{\sigma_i^2(\beta)}, \quad j = 1, \dots, m. \end{aligned}$$

2. Für alle $i, j = 1, \dots, m$ gilt:

$$\begin{aligned} I_{ij}(\beta) &= \mathbb{E}(U_i(\beta)U_j(\beta)) = \sum_{k,l=1}^n x_{ki}x_{lj} \underbrace{\text{Cov}(Y_k, Y_l)}_{\delta_{kl}\sigma_k^2(\beta)} \cdot \frac{\partial g^{-1}(\eta_k)}{\partial \eta_k} \frac{\partial g^{-1}(\eta_l)}{\partial \eta_l} \frac{1}{\sigma_k^2(\beta)\sigma_l^2(\beta)} \\ &= \sum_{k=1}^n x_{ki}x_{kj} \left(\frac{\partial g^{-1}(\eta_k)}{\partial \eta_k} \right)^2 \frac{1}{\sigma_k^2(\beta)}. \end{aligned}$$

□

Bemerkung 3.3.1. Im Falle der natürlichen Linkfunktion vereinfachen sich die obigen Gleichungen. So sieht die Log-Likelihood-Funktion folgendermaßen aus:

$$\log L(Y, \beta) = \frac{1}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top \beta + a(Y_i, \tau) - b(x_i^\top \beta) \right).$$

Da in diesem Fall $g^{-1}(\eta_i) = b'(\eta_i)$, $\eta_i = x_i^\top \beta = \theta_i$ gilt

$$\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} = b''(\theta_i) \stackrel{\text{Lemma 3.1.2}}{=} \frac{1}{\tau^2} \sigma_i^2(\beta)$$

und somit

$$U_j(\beta) = \frac{1}{\tau^2} \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)), \quad j = 1, \dots, m,$$

$$I_{jk}(\beta) = \frac{1}{\tau^4} \sum_{i=1}^n x_{ij} x_{ik} \sigma_i^2(\beta), \quad j, k = 1, \dots, m.$$

Satz 3.3.2. Es gilt

$$W_{jk}(\beta) = \sum_{i=1}^n x_{ij} x_{ik} \left((Y_i - \mu_i(\beta)) \nu_i - u_i^2 \frac{1}{\sigma_i^2(\beta)} \right), \quad j, k = 1, \dots, m,$$

wobei

$$u_i = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \quad \text{und} \quad \nu_i = \frac{1}{\tau^2} \cdot \frac{\partial^2 ((b')^{-1} \circ g^{-1})(\eta_i)}{\partial \eta_i^2}, \quad i = 1, \dots, n,$$

$$\mu_i(\beta) = \mathbb{E}Y_i, \quad \sigma_i^2(\beta) = \text{Var}Y_i, \quad \eta_i = x_i^\top \beta.$$

Beweis. Für beliebige $j, k = 1, \dots, m$ gilt

$$\begin{aligned} W_{jk}(\beta) &= \frac{\partial}{\partial \beta_k} U_j(\beta) \stackrel{\text{Satz 3.3.1}}{=} \frac{\partial}{\partial \beta_k} \sum_{i=1}^n x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \\ &= \sum_{i=1}^n x_{ij} \left((Y_i - \mu_i(\beta)) \frac{\partial}{\partial \beta_k} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \right) - \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \frac{1}{\sigma_i^2(\beta)} \frac{\partial \mu_i(\beta)}{\partial \beta_k} \right) \\ &= \sum_{i=1}^n \left(x_{ij} (Y_i - \mu_i(\beta)) \frac{\partial}{\partial \beta_k} \left(\frac{\tau^2 b''((b')^{-1}(g^{-1}(\eta_i))) ((b')^{-1} \circ g^{-1})'(\eta_i)}{\tau^2 b''((b')^{-1}(g^{-1}(\eta_i)))} \right) \right. \\ &\quad \left. - \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right)^2 \frac{1}{\sigma_i^2(\beta)} x_{ik} \right) \\ &= \sum_{i=1}^n x_{ij} x_{ik} \left((Y_i - \mu_i(\beta)) \nu_i - u_i^2 \frac{1}{\sigma_i^2(\beta)} \right), \end{aligned}$$

wobei

$$\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot \frac{1}{\sigma_i^2(\beta)} \stackrel{\text{Lemma 3.1.2}}{=} \frac{\partial b'(\theta_i)}{\partial \eta_i} \cdot \frac{1}{\tau^2} \cdot \frac{1}{b''(\theta_i)} = \frac{\partial b'(\theta_i)}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \eta_i} \frac{1}{\tau^2} \frac{1}{b''(\theta_i)} = \frac{1}{\tau^2} \frac{\partial \theta_i}{\partial \eta_i}$$

und

$$\frac{\partial}{\partial \beta_k} \left(\frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot \frac{1}{\sigma_i^2(\beta)} \right) = \frac{1}{\tau^2} \frac{\partial^2 \theta_i}{\partial \eta_i^2} \cdot \frac{\partial \eta_i}{\partial \beta_k} \stackrel{\eta_i = x_i^\top \beta}{=} \frac{1}{\tau^2} \frac{\partial^2 \theta_i}{\partial \eta_i^2} \cdot x_{ik},$$

dabei ist

$$\frac{\overbrace{\frac{\partial g^{-1}(\eta_i)}{\partial \beta_k}}^{\mu_i(\beta)}}{\partial \beta_k} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_k} = \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \cdot x_{ik}$$

und $\theta_i = (b')^{-1} \circ g^{-1}(\eta_i)$, $i = 1, \dots, n$. □

Für verallgemeinerte lineare Modelle mit natürlichen Linkfunktionen gilt insbesondere

$$W(\beta) = -I(\beta) = -\frac{1}{\tau^4} \sum_{i=1}^n x_{ij} x_{ik} \sigma_i^2(\beta), \quad (3.3.3)$$

weil in diesem Fall $\nu_i = 0$ für alle $i = 1, \dots, n$. $W(\beta)$ ist also deterministisch. Tatsächlich ist nach Lemma 3.2.1 $\theta_i = x_i^\top \beta = \eta_i$ und somit $\frac{\partial^2 \theta_i}{\partial \eta_i^2} = 0$, $i = 1, \dots, n$.

Aus Bemerkung 3.3.1 außerdem: $u_i^2 = \frac{1}{\tau^4} \sigma_i^4(\beta)$.

Beispiel 3.3.1. Wie sehen $U(\beta)$, $I(\beta)$ und $W(\beta)$ für unsere Modelle aus Beispiel 2.6.2 (natürliche Linkfunktionen) aus?

1. **Normalverteilung:** dieser Fall entspricht der üblichen multivariaten linearen Regression mit normalverteilten Störgrößen. In diesem Fall gilt $\mu = X\beta$, $\tau^2 = \sigma^2$.

Aus Bemerkung 3.3.1 folgt

$$\begin{aligned} U(\beta) &= \frac{1}{\sigma^2} X^\top (Y - X\beta), \\ I(\beta) &= (\mathbb{E} (U_i(\beta) \cdot U_j(\beta)))_{i,j=1,\dots,m} = \frac{1}{\sigma^2} X^\top X, \\ W(\beta) &= -I(\beta). \end{aligned}$$

2. **Logistische Regression:** hier gilt $\tau^2 = 1$, $\mu_i = p_i$, $\sigma_i^2 = p_i(1 - p_i)$, $i = 1, \dots, n$, $p_i \in (0, 1)$ und somit

$$\begin{aligned} U(\beta) &= X^\top(Y - p) , \\ I(\beta) &= X^\top \text{diag}(p_i(1 - p_i))X , \\ W(\beta) &= -I(\beta) , \end{aligned}$$

wobei $p = (p_1, \dots, p_n)^\top$.

3. **Poisson-Regression:** es gilt $\tau^2 = 1$, $\mu_i = \lambda_i = \sigma_i^2$, $i = 1, \dots, n$ und somit

$$\begin{aligned} U(\beta) &= X^\top(Y - \lambda) , \\ I(\beta) &= X^\top \text{diag}(\lambda_i)X , \\ W(\beta) &= -I(\beta) , \end{aligned}$$

wobei $\lambda = (\lambda_1, \dots, \lambda_n)^\top$.

Wann ist die Lösung des Gleichungssystems $U(\beta) = 0$ auch ein Maximum-Punkt der Funktion $\log L(Y, \beta)$?

Mit anderen Worten: Wann existiert der ML-Schätzer $\hat{\beta}$ von β , der eindeutig bestimmt ist?

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \log L(Y, \beta)$$

An der hinreichenden Bedingung eines Maximums folgt, dass die Hesse-Matrix $W(\beta)$ negativ definit sein muss.

Betrachten wir den Spezialfall der natürlichen Linkfunktion.

Dann gilt nach Bemerkung 3.3.1:

- Das Gleichungssystem $U(\beta) = 0$ schreibt sich $U(\beta) = \frac{1}{\tau^2} X^\top(Y - \mu(\beta)) = 0$
- Die Matrix $W(\beta) = -\frac{1}{\tau^4} X^\top \text{diag}(\sigma_i^2(\beta))X$ ist negativ definit, falls zusätzlich $\operatorname{rg}(X) = m$ und $0 < \sigma_i^2(\beta) < \infty$ für alle $i = 1, \dots, n$.

Unter diesen Bedingungen existiert also ein eindeutiger ML-Schätzer $\hat{\beta}$ für β .

Geben wir jetzt Verfahren an, die das (im Allgemeinen nicht lineare) Gleichungssystem $U(\beta) = 0$ numerisch lösen. Diese Ansätze sind iterativ, d.h. sie nähern sich schrittweise dem ML-Schätzer $\hat{\beta}$ an.

1. Newton-Verfahren

Wähle einen geeigneten Startwert $\hat{\beta}_0 \in \mathbb{R}^m$.

Im Schritt $k + 1$, berechne $\hat{\beta}_{k+1}$ aus $\hat{\beta}_k$, $k = 0, 1, \dots$ auf folgende Art und Weise:

- Nimm die Taylor-Entwicklung von $U(\beta)$ bis zur ersten Ordnung an der Stelle $\hat{\beta}_k$: $U(\beta) \approx U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k)$.
- Setze sie gleich Null: $U(\hat{\beta}_k) + W(\hat{\beta}_k)(\beta - \hat{\beta}_k) = 0$
- Die Lösung dieses Gleichungssystems ist $\hat{\beta}_{k+1}$:

$$\hat{\beta}_{k+1} = \hat{\beta}_k - W^{-1}(\hat{\beta}_k) \cdot U(\hat{\beta}_k), \quad k = 0, 1, 2, \dots,$$

vorausgesetzt, dass $W(\hat{\beta}_k)$ invertierbar ist.

Breche den Iterationsprozess ab, sobald $|\hat{\beta}_{k+1} - \hat{\beta}_k| < \delta$ für eine vorgegebene Genauigkeit $\delta > 0$ ist.

Das Konvergenzverhalten dieses Verfahrens hängt entscheidend von der Wahl von $\hat{\beta}_0$ ab, für dessen Konvergenz $\hat{\beta}_0$ nah genug bei $\hat{\beta}$ liegen muss. Ein weiterer Nachteil dieses Verfahrens ist, dass die zufällige Matrix $W(\beta)$ unter Umständen nicht invertierbar ist. Deswegen schlagen wir jetzt eine Modifikation des Newton-Verfahrens vor, bei der $W(\beta)$ durch den Erwartungswert

$$\mathbb{E} W(\beta) = -I(\beta) \tag{3.3.4}$$

ersetzt wird. Dass die Identität (3.3.3) stimmt, folgt aus dem Satz 3.3.2, und der Tatsache, dass $\mathbb{E} Y_i = \mu_i$, $i = 1, \dots, n$. Wenn man voraussetzt, dass $rg(X) = m$ und $u_i \neq 0$, $i = 1, \dots, n$, so ist nach Satz 3.3.1 $I(\beta)$ invertierbar. Dieses Verfahren wird *Fisher Scoring* genannt.

Der einzige Unterschied zu den Schritten des Newton-Verfahrens besteht beim Fisher Scoring darin, dass man in Schritt 2 die iterative Gleichung

$$\hat{\beta}_{k+1} = \hat{\beta}_k + I^{-1}(\hat{\beta}_k) U(\hat{\beta}_k), \quad k = 0, 1, \dots$$

einsetzt.

Im Falle einer natürlichen Linkfunktion gilt nach Bemerkung 3.3.1

$$\begin{aligned} \hat{\beta}_{k+1} &= \hat{\beta}_k + \tau^4 \left(X^\top \text{diag}(\sigma_i^2(\hat{\beta}_k)) X \right)^{-1} \frac{1}{\tau^2} \left(X^\top (Y - \mu(\hat{\beta}_k)) \right) \\ &= \hat{\beta}_k + \tau^2 \left(X^\top \text{diag}(\sigma_i^2(\hat{\beta}_k)) X \right)^{-1} \left(X^\top (Y - \mu(\hat{\beta}_k)) \right). \end{aligned}$$

3.4 Asymptotische Tests für β

Das Ziel dieses Abschnittes ist es, eine Testregel für die Hypothese

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0 \quad \text{mit} \quad \beta = (\beta_1, \dots, \beta_m)^\top, \quad \beta_0 = (\beta_{01}, \dots, \beta_{0m})^\top$$

zu konstruieren. Insbesondere sind die Haupthypothesen $H_0 : \beta = 0$ bzw. $H_0 : \beta_j = 0$ von Interesse, weil sie die Tatsache reflektieren, dass die Zielvariablen $Y = (Y_1, \dots, Y_n)^\top$ von einigen Ausgangsvariablen (z.B. $(x_{1j}, \dots, x_{nj})^\top$ im Falle der Hypothese $\beta_j = 0$) unabhängig sind.

Um solche Hypothesen testen zu können, werden Teststatistiken T_n vorgeschlagen, die asymptotisch (für $n \rightarrow \infty$) eine bekannte Prüfverteilung (z.B. multivariate Normalverteilung oder χ^2 -Verteilung) besitzen. Dafür sind gewisse Vorarbeiten notwendig.

Sei

$$g(\mathbb{E}Y_i) = X_i\beta, \quad i = 1, \dots, n,$$

ein verallgemeinertes lineares Modell mit natürlicher Linkfunktion g . Seien $L(Y, \beta)$, $U(\beta)$ und $I(\beta)$ die Likelihood-Funktion, der Vektor der partiellen Ableitungen von $\log L(Y, \beta)$ bzw. die Fisher-Informationsmatrix in diesem Modell.

Durch $\hat{\beta}_n = \hat{\beta}(Y_1, \dots, Y_n, X)$ bezeichne man eine Folge von Schätzern für β .

Es gelten folgende Voraussetzungen:

1. \exists Kompaktum $K \subset \mathbb{R}^m$, so dass alle Zeilen X_i , $i = 1, \dots, n$, $n \in \mathbb{N}$, von X in K liegen. Dabei soll $\theta = x^\top \beta \in \Theta$ für alle $\beta \in \mathbb{R}^m$ und $x \in K$.
2. Es existiert eine Folge $\{\Gamma_n\}_{n \in \mathbb{N}}$ von diagonalen $(m \times m)$ -Matrizen $\Gamma_n = \Gamma_n(\beta)$ mit positiven Diagonalelementen und den Eigenschaften $\lim_{n \rightarrow \infty} \Gamma_n = 0$, $\lim_{n \rightarrow \infty} \Gamma_n^\top I_n(\beta) \Gamma_n = K^{-1}(\beta)$, wobei $K(\beta)$ eine symmetrische positiv definite $(m \times m)$ -Matrix ist, $\forall \beta \in \mathbb{R}^m$.

Satz 3.4.1. Unter obigen Voraussetzungen gilt:

es existiert eine Γ_n -Konsistente Folge von ML-Schätzern $\{\hat{\beta}_n\}$ für β ,
(d.h. $\mathbb{P}(\Gamma_n^{-1}|\hat{\beta}_n - \beta| \leq \varepsilon, U(\hat{\beta}_n) = 0) \rightarrow 1$ für $n \rightarrow \infty$), so dass

1. $T_n^* = \Gamma_n^{-1}(\hat{\beta}_n - \beta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, K(\beta))$ und
2. $T_n = 2(\log L(Y, \hat{\beta}_n) - \log L(Y, \beta)) \xrightarrow[n \rightarrow \infty]{d} \chi_m^2$, $m = \dim \beta$

Bemerkung 3.4.1. (vgl. [15], S.288-292)

1. Oft wählt man $\Gamma_n = \text{diag}\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$
2. Bisher wurde stets angenommen, dass der Störparameter τ^2 bekannt ist. Falls es nicht der Fall ist, kann τ^2 durch

$$\hat{\tau}^2 = \frac{1}{n-m} \sum_{i=1}^n \frac{(Y_i - \mu_i(\hat{\beta}_n))^2}{b''(\hat{\theta}_{ni})}$$

geschätzt werden, wobei $\hat{\theta}_{ni} = (b')^{-1}(\mu_i(\hat{\beta}_n))$, $i = 1, \dots, n$ ist. Dieser Schätzer ist ein empirisches Analogon der Gleichung $\tau^2 = \frac{\text{Var}Y_i}{b''(\theta_i)}$ aus Lemma 3.1.2.

3. Die Aussage 2. des Satzes 3.4.1 gilt auch, wenn man den unbekannt Parameter τ^2 durch einen konsistenten Schätzer τ_n^2 ersetzt.

Wie verwendet man nun den Satz 3.4.1 zum Testen der Hypothesen

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0 ,$$

oder komponentenweise

$$H_0 : \beta_j = \beta_{j0} , \quad j = 1, \dots, m \quad \text{vs.} \quad H_1 : \exists j_1 : \beta_{j_1} \neq \beta_{j_1 0} \quad ?$$

Sei

$$g(\mathbb{E}Y_i) = \sum_{j=1}^m x_{ij}\beta_j , \quad i = 1, \dots, n ,$$

ein verallgemeinertes lineares Modell mit natürlicher Linkfunktion g .

Nach Bemerkung 3.3.1 gilt

$$\log L(Y, \beta) = \frac{1}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top \beta + a(Y_i, \tau) - b(x_i^\top \beta) \right)$$

wobei $Y = (Y_1, \dots, Y_n)^\top$ und $x_i = (x_{i1}, \dots, x_{im})^\top$. Deshalb gilt

$$T_n = \frac{2}{\tau^2} \sum_{i=1}^n \left(Y_i x_i^\top (\hat{\beta}_n - \beta_0) - b(x_i^\top \hat{\beta}_n) + b(x_i^\top \beta_0) \right)$$

Bei Vorgabe eines Exponential-Modells (τ, b - bekannt), der Stichprobe der Zielvariablen Y und der Designmatrix X wird H_0 verworfen, falls $T_n > \chi_{m,1-\alpha}^2$, wobei m die Anzahl der Parameter im Modell, $\chi_{m,1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der χ_m^2 -Verteilung und $\alpha \in (0, 1)$ das Signifikanzniveau des asymptotischen Tests ist. Dieser Test ist nur für relativ große n anwendbar. Der Fehler 1. Art hat dabei (für $n \rightarrow \infty$) die asymptotische Wahrscheinlichkeit α . Falls eine einfache Hypothese

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0$$

getestet werden soll, benutzt man die aus der Statistik T_n^* abgeleitete Teststatistik T_n^1 . H_0 wird verworfen, falls

$$|T_n^1| = \frac{|\hat{\beta}_{nj}|}{(\Gamma_n(\hat{\beta}_n))_{jj}} > z_{1-\frac{\alpha}{2}},$$

wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der $\mathcal{N}(0, 1)$ -Verteilung ist. Hierbei ist $\{\Gamma_n\}$ so gewählt worden, dass $K(\beta) = Id$ ist, $\forall \beta \in \mathbb{R}^m$. Dies ist ein asymptotischer Test zum Niveau α , weil

$$\begin{aligned} P_{H_0}(|T_n^1| > z_{1-\frac{\alpha}{2}}) &= 1 - P_{H_0}(|T_n^*| \leq z_{1-\frac{\alpha}{2}}) \xrightarrow{n \rightarrow \infty} 1 - \Phi(z_{1-\frac{\alpha}{2}}) + \underbrace{\Phi(-z_{1-\frac{\alpha}{2}})}_{1-\Phi(z_{1-\frac{\alpha}{2}})} \\ &= 1 - \left(1 - \frac{\alpha}{2}\right) + 1 - \left(1 - \frac{\alpha}{2}\right) = \alpha, \end{aligned}$$

wobei

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

die Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung ist.

Beispiel 3.4.1. (Kreditrisikoprüfung)

vgl. Fahrmeir, L., Kneib, T., Lang, S. - Regression, S.208ff

Es liegt folgender Datensatz einer süddeutschen Bank aus den 1990er Jahren vor:

Es werden Ergebnisse der Kreditrisikoprüfung von $n = 1000$ Kreditanträgen (ca. 700 gute und 300 schlechte Kredite) analysiert:

Zielvariable $Y_i = \begin{cases} 0, & \text{falls das Darlehen vom Kunden } i \text{ zurückgezahlt wurde} \\ 1, & \text{falls das Darlehen vom Kunden } i \text{ nicht zurückgezahlt wurde} \end{cases}$

Die Designmatrix X enthält folgende Zusatzinformationen über den Kunden:

x_{i1} - Kontoführung des Kontos bei der Bank: $= \begin{cases} 1, & \text{kein Konto} \\ 0, & \text{sonst} \end{cases}$

x_{i2} - Bewertung der Kontoführung: $= \begin{cases} 1, & \text{gutes Konto} \\ 0, & \text{kein oder schwaches Konto} \end{cases}$

x_{i3} - Laufzeit des Kredits in Monaten

x_{i4} - Höhe des Kredits in DM

x_{i5} - Zahlungsverhalten beim Kunden: $= \begin{cases} 1, & \text{gut} \\ 0, & \text{sonst} \end{cases}$

x_{i6} - Verwendungszweck: $= \begin{cases} 1, & \text{privat} \\ 0, & \text{geschäftlich} \end{cases}$

Frage: Wie soll $\hat{\beta}$ geschätzt werden?

Als Modell wird das Logit-Modell gewählt mit $p_i = P(Y_i = 1)$, $i = 1, \dots, n$:

		$Y = 1$	$Y = 0$
x_1	kein Konto	45.0	20.0
x_2	gut	15.3	49.8
	schlecht	39.7	30.2
x_4	Kredithöhe	$Y = 1$	$Y = 0$
	$0 < \dots \leq 500$	1.00	2.14
	$500 < \dots \leq 1000$	11.33	9.14
	$1000 < \dots \leq 1500$	17.00	19.86
	$1500 < \dots \leq 2500$	19.67	24.57
	$2500 < \dots \leq 5000$	25.00	28.57
	$5000 < \dots \leq 7500$	11.33	9.71
	$7500 < \dots \leq 10000$	6.67	3.71
	$10000 < \dots \leq 15000$	7.00	2.00
	$15000 < \dots \leq 20000$	1.00	0.29
x_5	Frühere Kredite	$Y = 1$	$Y = 0$
	gut	82.33	94.95
	schlecht	17.66	5.15
x_6	Verwendungszweck	$Y = 1$	$Y = 0$
	privat	57.53	69.29
	beruflich	42.47	30.71

Tabelle 3.1: Auszug aus dem Originaldatensatz

\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{x}_5	\bar{x}_6
0.274	0.393	20.903	3271	0.911	0.657

Tabelle 3.2: Mittelwerte \bar{x}_j von x_{ij} im Datensatz

$$\log \frac{p_i}{1-p_i} = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + x_{i4}\beta_4 + x_{i5}\beta_5 + x_{i6}\beta_6 \quad \text{für } i = 1, \dots, n,$$

wobei $\beta = (\beta_0, \dots, \beta_6)^\top$, $m = 7$.

Ziel: Schätze β_0, \dots, β_6 und prüfe, welche Faktoren für die künftige Kreditvergabe relevant sind.

$H_0 : \beta_i = 0$ (Merkmal x_i beeinflusst die Kreditvergabe nicht) wird abgelehnt, falls p-Wert $\leq \alpha$. Man sieht, dass u.a. auch β_4 für die Kreditvergabe nicht relevant ist, was der Intuition widerspricht. Eine Verfeinerung des Modells ist notwendig:

	Wert	$\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$	T_n^1	p-Wert
β_0	0.281	0.303	-0.94	0.347
β_1	0.618	0.175	3.53	< 0.001
β_2	-1.338	0.201	-6.65	< 0.001
β_3	0.033	0.008	4.29	< 0.001
β_4	0.023	0.033	0.72	0.474
β_5	-0.986	0.251	-3.93	< 0.001
β_6	-0.426	0.266	-2.69	0.007

Tabelle 3.3: Ergebnis zur ML-Schätzung durch das Fisher Scoring Verfahren, wobei $\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$ als asymptotische Standardabweichung von $\hat{\beta}_i$ interpretiert wird. Signifikanzniveau: $\alpha = 0.001$

Neues Modell:

$$g(\mathbb{E} Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3^1 x_{i3} + \beta_3^2 x_{i3}^2 + \beta_4^1 x_{i4} + \beta_4^2 x_{i4}^2 + \beta_5 x_{i5} + \beta_6 x_{i6}$$

	Wert	$\sqrt{(I_n^{-1}(\hat{\beta}))_{ii}}$	T_n^1	p-Wert
β_0	-0.488	0.390	-1.25	0.211
β_1	0.618	0.176	3.51	< 0.001
β_2	-1.337	0.202	-6.61	< 0.001
β_3^1	0.092	0.025	3.64	< 0.001
β_3^2	-0.001	< 0.001	-2.20	0.028
β_4^1	-0.264	0.099	-2.68	0.007
β_4^2	0.023	0.007	3.07	0.002
β_5	-0.995	0.255	-3.90	< 0.001
β_6	-0.404	0.160	-2.52	0.012

Tabelle 3.4: p -Werte für die Regressionskoeffizienten des neuen Modells

Frage: Welches Modell ist besser?

Mit anderen Worten, wir testen

$$H_0 : \beta_3^2 = 0 \text{ (lineares Modell) vs. } H_1 : \beta_3^2 \neq 0 \text{ (quadratisches Modell) bzw.}$$

$$H_0 : \beta_4^2 = 0 \text{ (lineares Modell) vs. } H_1 : \beta_4^2 \neq 0 \text{ (quadratisches Modell) .}$$

Dabei verallgemeinern wir die Art der statistischen Hypothesen wie folgt: es wird

$$H_0 : C\beta = d \text{ vs. } H_1 : C\beta \neq d$$

getestet, wobei C eine $(r \times m)$ -Matrix mit $rg C = r \leq m$ ist und $d \in \mathbb{R}^r$.

Zum Vergleich: früher haben wir

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0, \quad \beta, \beta_0 \in \mathbb{R}^m$$

getestet. Natürlich ist $\beta = \beta_0$ ein Spezialfall von $C\beta = d$ mit $C = \text{Id}$, $d = \beta_0$. Die neuen Hypothesen beinhalten Aussagen über die Linearkombinationen der Parameterwerte. Wie soll H_0 vs. H_1 getestet werden?

Sei $\tilde{\beta}_n$ der ML-Schätzer von β unter H_0 , d.h. $\tilde{\beta}_n = \underset{\beta \in \mathbb{R}^m: C\beta=d}{\operatorname{argmax}} \log L(Y, \beta)$

Sei $\hat{\beta}_n$ der ML-Schätzer von β unrestringiert, d.h. $\hat{\beta}_n = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmax}} \log L(Y, \beta)$.

Die Idee der folgenden Tests ist es, $\tilde{\beta}_n$ mit $\hat{\beta}_n$ zu vergleichen. Falls die Abweichung $\hat{\beta}_n - \tilde{\beta}_n$ groß ist, soll H_0 abgelehnt werden.

Satz 3.4.2. Sei $\log L(Y, \beta)$ die Log-Likelihood-Funktion der Stichprobe der Zielvariablen $Y = (Y_1, \dots, Y_n)^\top$, $I_n(\beta)$ die Fisher-Informationsmatrix, $U(\beta)$ die Score-Funktion des verallgemeinerten linearen Modells mit natürlicher Linkfunktion

$$g : g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n.$$

Wir führen folgende Teststatistiken ein:

1. **Likelihood-Ratio-Teststatistik:**

$$\tilde{T}_n = 2(\log L(Y, \hat{\beta}_n) - \log L(Y, \tilde{\beta}_n))$$

2. **Wald-Statistik:**

$$\tilde{T}_n^* = (C\hat{\beta}_n - d)^\top (CI_n^{-1}(\hat{\beta}_n)C^\top)^{-1} (C\hat{\beta}_n - d)$$

3. **Score-Statistik:**

$$\bar{T}_n^* = U(\tilde{\beta}_n)^\top I_n^{-1}(\tilde{\beta}_n) U(\tilde{\beta}_n)$$

Unter gewissen Bedingungen an die Schätzer $\hat{\beta}$ und $\tilde{\beta}$ (vgl. Satz 3.4.1) sind die Teststatistiken 1 - 3 asymptotisch χ_m^2 -verteilt: z.B. gilt für die Likelihood-Ratio-Teststatistik

$$\tilde{T}_n \xrightarrow[n \rightarrow \infty]{d} \chi_m^2.$$

Folgerung 3.4.1. Der Satz 2.6.4 liefert uns folgende Entscheidungsregel: H_0 wird abgelehnt, falls

$$\tilde{T}_n(\tilde{T}_n^*, \bar{T}_n) > \chi_{m, 1-\alpha}^2.$$

Dies ist ein asymptotischer Test zum Signifikanzniveau α .

Beispiel 3.4.2 (Fortsetzung). Es ergeben sich folgende Werte für die Teststatistiken:

$$\begin{aligned} \tilde{T}_n &= 12.44, & \text{p-Wert: } 0.0020 \\ \tilde{T}_n^* &= 11.47, & \text{p-Wert: } 0.0032. \end{aligned}$$

Für $\alpha = 0.005$ gilt p-Wert $\leq \alpha$, somit wird $H_0 : \beta_4^2 = 0$ abgelehnt \Rightarrow das quadratische verallgemeinerte lineare Modell ist besser.

3.5 Kriterien zur Modellwahl bzw. Modellanpassung

Es ist bekannt, dass die Güte der Anpassung eines parametrischen Modells an die Daten im Allgemeinen steigt, wenn die Anzahl der Parameter erhöht wird. Die Aufgabe eines Statistikers ist es aber, ein gut passendes Modell mit einer möglichst kleinen Anzahl an Parametern zu finden. Deshalb verwendet man folgendes Informationskriterium von Akaike, um Modelle mit (möglicherweise) unterschiedlichen Parametersätzen zu vergleichen.

Informationskoeffizient von Akaike:

$$\text{AIC} = -2 \log L(Y, \hat{\beta}) + 2m ,$$

wobei $Y = (Y_1, \dots, Y_n)$ die Stichprobe der Zielvariablen im verallgemeinerten linearen Modell und $\hat{\beta}$ der dazugehörige ML-Schätzer sei. Der Wert von AIC berücksichtigt einerseits die Forderung der Maximalität der Log-Likelihood-Funktion $\log L(Y, \hat{\beta})$, andererseits bestraft er Modelle mit einer großen Anzahl von Parametern m . Das Modell mit dem kleineren AIC ist als besseres Modell einzustufen. Manchmal verwendet man statt AIC den normierten Koeffizienten AIC/n .

Beispiel 3.5.1 (Fortsetzung). Berechnen wir den Informationskoeffizienten von Akaike für das lineare und quadratische Logit-Modell im Beispiel der Kreditrisikoprüfung:

$$\text{Lineares Modell : AIC} = 1043.815$$

$$\text{Quadratisches Modell : AIC} = 1035.371$$

Man sieht anhand des AIC, dass die Wahl zu Gunsten des quadratischen Modells ausfällt.

Der Nachteil der oben beschriebenen AIC-Regel liegt darin, dass die endgültige Entscheidung dem Statistiker überlassen bleibt. Deshalb ist es wünschenswert, einen statistischen Test zu konstruieren, der die Güte der Modellanpassung beurteilen kann.

Wir werden jetzt den χ^2 -Test beschreiben.

Sei

$$g(\mathbb{E} Y_i) = X_i \beta , \quad i = 1, \dots, n ,$$

ein verallgemeinertes lineares Modell mit Linkfunktion g und Parametervektor $\beta = (\beta_1, \dots, \beta_m)^\top$. Teilen wir die Zielvariablen Y_1, \dots, Y_n in k Gruppen auf, so dass sie möglichst homogen in Bezug auf die zu schätzenden Parameter sind. So liegt z.B. eine solche Aufteilung vor, wenn der Wertebereich der Zielvariablen Y_i „geschickt“ in $k > m$ ¹ Intervalle $(a_l, b_l]$ unterteilt wird:

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = a_3 < \dots < b_{k-1} = a_k < b_k \leq +\infty$$

¹ $k \leq m \Rightarrow D \xrightarrow[n \rightarrow \infty]{d} \underbrace{\chi_{k-m-1}^2}_{<0}$

In die Gruppe l fallen alle Beobachtungen Y_i , die zu $(a_l, b_l]$ gehören. Dabei müssen $(a_l, b_l]$ so gewählt werden, dass $\hat{\mu}_j = g^{-1}(X_j \hat{\beta})$ innerhalb einer Gruppe konstant wird: $\hat{\mu}_j \equiv \hat{\mu}_l \forall j$ aus Gruppe l .² Sei

- $n_l = \# \{Y_j : Y_j \in (a_l, b_l]\}$ die Klassenstärke der Klasse l
- $\bar{Y}_l = \frac{1}{n_l} \sum Y_j$ das arithmetische Mittel innerhalb der Klasse l
- $\hat{\beta}$ der ML-Schätzer von β , der aus Y gewonnen wurde
- $l_l(\beta) = \sum \log f_\theta(Y_j)$ die Log-Likelihood-Funktion der Zielvariablen Y_i innerhalb der Gruppe l
- $\hat{\mu}_l = g^{-1}(X_l \hat{\beta})$ und $v(\hat{\mu}_l)$ der Erwartungswert- bzw. der Varianzschätzer von $\mu_l = \mathbb{E} Y_l$, die aus dem ML-Schätzer $\hat{\beta}$ gewonnen wurden

Dabei ist $v(\hat{\mu}_l) = \tau^2 b''(b'^{-1}(\hat{\mu}_l))$, wobei $b(\cdot)$ der entsprechende Koeffizient in der Dichte f_θ aus der Exponentialfamilie ist. Man bildet folgende Teststatistiken:

$$\chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{\mu}_l)^2}{v(\hat{\mu}_l)/n_l}$$

$$D = -2\tau^2 \sum_{l=1}^k (l_l(\hat{\mu}_l) - l_l(\bar{Y}_l))$$

Satz 3.5.1.

Falls $n \rightarrow \infty$ und die Anzahl $n_l \rightarrow \infty \forall l$, dann gilt unter gewissen Voraussetzungen Folgendes:

$$\chi^2 \xrightarrow[n \rightarrow \infty]{d} \chi_{k-m-1}^2$$

$$D \xrightarrow[n \rightarrow \infty]{d} \chi_{k-m-1}^2$$

²Dies ist eine informelle Beschreibung des Vorgangs, bei dem für jedes Y_i n_i unabhängige Kopien von Y_i erzeugt werden, die die i -te Klasse bilden.

Folgerung 3.5.1.

Mit Hilfe der Behauptungen des Satzes 2.6.5 können die Hypothesen

$$H_0 : Y = (Y_1, \dots, Y_n) \text{ stammt aus dem Modell } g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n$$

vs.

$$H_1 : Y = (Y_1, \dots, Y_n) \text{ stammt nicht aus dem Modell } g(\mathbb{E} Y_i) = X_i \beta, \quad i = 1, \dots, n$$

folgendermaßen getestet werden:

H_0 wird (für große n) zum asymptotischen Signifikanzniveau α verworfen, falls

$$\chi^2 > \chi_{k-m-1, 1-\alpha}^2 \quad \text{bzw.} \quad D > \chi_{k-m-1, 1-\alpha}^2.$$

Diese Tests sollten aber nicht verwendet werden, falls die Klassenstärken n_l klein sind.

Beispiel 3.5.2.

Wie sehen die oben beschriebenen Tests im Falle der Logit- bzw. Poisson-Regression aus?

1. **Logit-Modell:** $Y_i \sim \text{Bernoulli}(p_i)$, $i = 1, \dots, n$

$$\Rightarrow \text{verallgemeinertes lineares Modell} \quad \log \frac{p_i}{1-p_i} = X_i \beta, \quad i = 1, \dots, n$$

Wir teilen Y_1, \dots, Y_n in k Klassen auf, so dass die Wahrscheinlichkeit des Auftretens von 1 in jeder Klasse möglichst gut durch $\bar{Y}_l = \frac{1}{n_l} \sum Y_i$ geschätzt wird. Somit gilt mit $\hat{\mu}_l = \hat{p}_l = g^{-1}(X_l \hat{\beta}) = \frac{e^{X_l^\top \hat{\beta}}}{1 + e^{X_l^\top \hat{\beta}}}$, $v(\hat{p}_l) = \hat{p}_l(1 - \hat{p}_l)$

$$\Rightarrow \chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{p}_l)^2}{\hat{p}_l(1 - \hat{p}_l)/n_l}$$

2. **Poisson-Modell:** $Y_i \sim \text{Poisson}(\lambda)$,

$$\Rightarrow \text{verallgemeinertes lineares Modell} \quad \log \lambda_i = X_i \beta, \quad i = 1, \dots, n$$

Somit gilt mit $\hat{\mu}_l = \hat{\lambda}_l = e^{X_l \hat{\beta}}$, $v(\hat{\lambda}_l) = \hat{\lambda}_l$

$$\Rightarrow \chi^2 = \sum_{l=1}^k \frac{(\bar{Y}_l - \hat{\lambda}_l)^2}{\hat{\lambda}_l/n_l}$$

4 Hauptkomponentenanalyse

In diesem Kapitel werden Methoden zur Reduktion der Komplexität von sehr großen statistischen Datensätzen vorgestellt, die als Hauptkomponentenanalyse (HKA) bekannt sind (engl. Principal Component Analysis, PCA). Mit ihrer Hilfe ist es möglich einen sehr hochdimensionalen Datensatz $X = (X_1, \dots, X_n)^T \in \mathbb{R}^n$ auf wenige wirklich wichtige Komponenten $\varphi = AX \in \mathbb{R}^d$ zurückzuführen, $d \ll n$, die aber dabei die meiste Variabilität des originalen Datensatzes X beibehalten. A ist dabei eine $(d \times n)$ -Matrix, die zu finden ist, wenn gewisse (in 4.2.1 angegebene) Nebenbedingungen erfüllt sind. Andere Beispiele von Anwendungen sind Visualisierung von komplexen Datensätzen, Ausreißer-Erkennung, Cluster-Analyse u.s.w.. Für eine Übersicht siehe z.B. [8].

4.1 Einführung

Um nachfolgende Problemstellungen zu motivieren, betrachten wir ein Beispiel des Text Mining aus der Autoindustrie:

Beispiel 4.1.1. Ein Autohersteller ist daran interessiert, seine Verluste, die in Folge von Betrug und Inkompetenz seitens seiner Niederlassungen bei Garantie-Reparaturen auftreten, zu minimieren. Deshalb möchte er eine Auffälligkeitsanalyse von Reparaturbesichtigungen aus Garantie-Werkstätten betreiben, die dazu führen sollte, computergestützt, verdächtige Meldungen zu finden, die nachher manuell und einzeln weiter geprüft werden. Ein weiterer Anreiz für die automatischen Früherkennung von Auffälligkeiten besteht darin, dass flächendeckende Prüfungen nur für wenige Niederlassungen und in unregelmäßigen Zeitabständen (aus Kostengründen) möglich sind, und selbst die könnte man sich sparen. Ein typischer Text, der eine Garantie-Reparatur beschreibt, verwendet maximal 300.000 Wörter aus einem Fachwortschatz. Daher werden solche Texte als Vektoren $x = (x_1, \dots, x_n)^T$ der Länge $n = 300.000$ dargestellt, wobei

$$x_i = \begin{cases} 1 & , \text{ falls das Wort } i \text{ im Text } x \text{ vorkommt} \\ 0 & , \text{ sonst} \end{cases}$$

Diese Vektoren x werden normiert, so dass sie auf der Sphäre S^{n-1} liegen. Innerhalb eines Jahres entsteht dadurch eine riesige Datenbank solcher Vektoren x mit mehreren Millionen Einträgen. Die Aufgabe eines Statistikers besteht in der drastischen Reduktion der Dimension $n - 1$ des Datensatzes, so dass eine Visualisierung des Datensatzes möglich wird. Eine mögliche Lösung liegt in der Verwendung von HKA. Die HKA geht in ihren Ursprüngen auf die Arbeiten von Beltran (1873) und Jordan (1874) zurück, die

die Single Value Decomposition verwendeten. In der mehr oder minder modernen Form (vgl. 4.2.1) erscheint sie erst in den Arbeiten von K. Pearson (1901) und H. Hotelling (1933). Auch der Name HKA stammt von Hotelling. Eine Weiterentwicklung der Methoden ist Girshick (1939), Anderson (1963), Rao (1964) und anderen zu verdanken. Erst nach der Einführung der PCs ist aber diese Methodologie richtig angewandt geworden. Denn ohne Computer ist die Berechnung von Hauptkomponenten für $n > 4$ sehr schwierig. Seit den 1980er Jahren gibt es einen rasanten Anstieg der Anwendungen von HKA in allen Wissensbereichen (vor allem in Ingenieurwissenschaften), wo multivariate Datensätze analysiert werden sollen.

4.2 Hauptkomponentenanalyse auf Modellebene

In diesem Abschnitt wollen wir das Hauptproblem der HKA für Zufallsstichproben $X = (X_1, \dots, X_n)^T$ mit bekannter Kovarianzstruktur einführen. Sei $X = (X_1, \dots, X_n)^T$ eine Zufallsstichprobe von Zufallszahlen X_i mit bekannter Kovarianzmatrix Σ und $\text{Var} X_i \in (0, \infty)$, $i = 1, \dots, n$. Seien $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ die Eigenwerte von Σ , die in absteigender Reihenfolge geordnet und alle von einander verschieden sind. Wir suchen Linearkombinationen $\alpha^T X$ von X_i , die die maximale Varianz besitzen, wobei der Vektor α entsprechend normiert ist z.B., so dass $\alpha \in S^{n-1}$ in der Euklidischen Norm.

Definition 4.2.1. Die Linearkombination $\alpha_i^T X$, $i = 1, \dots, n$, heißt i -te Hauptkomponente von X , falls sie die maximale Varianz besitzt unter der Bedingung, dass $\alpha_i \in S^{n-1}$ und $\alpha_1^T X, \alpha_2^T X, \dots, \alpha_{i-1}^T X$ und $\alpha_i^T X$ unkorreliert sind:

$$\begin{cases} \text{Var } \alpha^T X \rightarrow \max_{\alpha} \\ |\alpha| = 1 \\ \text{Cov}(\alpha^T X, \alpha_j^T X) = 0, \quad j = 1, \dots, i-1 \end{cases} \quad (4.2.1)$$

Dabei heißt α_i der Koeffizientenvektor der i -ten Hauptkomponente $\alpha_i^T X$.

Satz 4.2.1. Die i -te Hauptkomponente von X ist gegeben durch

$$Y_i = \alpha_i^T X,$$

wobei α_i der Eigenvektor von Σ mit Eigenwert λ_i ist. Dabei gilt

$$\text{Var}(Y_i) = \lambda_i, \quad i = 1, \dots, n.$$

Beweis. Zeigen wir, dass die Aussage des Satzes gilt für $i = 1, 2$. Für $i > 2$ ist der Beweis analog.

Für $i = 1$ gibt es eine Nebenbedingung $|\alpha| = 1$ in (4.2.1), die in die Lagrange-Zielfunktion

$$f(\alpha) = \text{Var}(\alpha^T X) + \lambda(|\alpha|^2 - 1)$$

übernommen wird. Dabei gilt

$$\begin{aligned}\text{Var}(\alpha^T X) &= \mathbb{E}(\alpha^T X - \mathbb{E}\alpha^T X)^2 = \mathbb{E}(\alpha^T (X - \mathbb{E}X))^2 = \mathbb{E}\alpha^T (X - \mathbb{E}X)(X - \mathbb{E}X)^T \alpha \\ &= \alpha^T \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T \alpha = \alpha^T \Sigma \alpha,\end{aligned}$$

$$|\alpha|^2 = \alpha^T \cdot \alpha, \text{ und } f(\alpha) = \alpha^T \Sigma \alpha + \lambda(\alpha^T \alpha - 1).$$

Die notwendige Bedingung des Maximums ist

$$\frac{\partial f}{\partial \alpha} = 0, \quad \frac{\partial f}{\partial \lambda} = 0,$$

wobei die zweite Gleichung einfach die Nebenbedingung $|\alpha| = 1$ repräsentiert.

$\frac{\partial f}{\partial \alpha} = \left(\frac{\partial f}{\partial \alpha^1}, \dots, \frac{\partial f}{\partial \alpha^n} \right)$, wobei $\alpha = (\alpha^1, \dots, \alpha^n)^T$ und $\frac{\partial f}{\partial \alpha} = 0$ schreibt sich $\Sigma \alpha - \lambda \alpha = 0$ in Vektorform oder $\Sigma \alpha = \lambda \alpha$, was heißt, dass α ein Eigenvektor von Σ mit dem Eigenwert λ ist. Da $\text{Var}(\alpha^T X) = \alpha^T \Sigma \alpha$ maximal sein soll, gilt

$$\text{Var}(\alpha^T X) = \alpha^T \lambda \alpha = \lambda \underbrace{\alpha^T \alpha}_1 = \lambda$$

und $\lambda = \lambda_1 > \lambda_2 > \dots > \lambda_n \Rightarrow \lambda = \lambda_1$ und $\alpha = \alpha_1$.

Für $i = 2$, soll die Maximierungsaufgabe

$$\left\{ \begin{array}{l} \alpha^T \Sigma \alpha \rightarrow \max_{\alpha} \\ \alpha^T \cdot \alpha = 1 \\ \text{Cov}(\alpha_1^T X, \alpha^T X) = 0 \end{array} \right.$$

bezüglich α gelöst werden, wobei

$$\text{Cov}(\alpha_1 X, \alpha^T X) = \alpha_1^T \Sigma \alpha = \alpha^T \Sigma \alpha_1 = \alpha^T \lambda_1 \alpha_1 = \lambda_1 \alpha^T \alpha_1.$$

Das heißt, folgende Funktion soll maximiert werden:

$$f(\alpha) = \alpha^T \Sigma \alpha + \lambda(\alpha^T \alpha - 1) + \delta \alpha^T \alpha_1.$$

Genau wie oben bekommt man

$$\frac{\partial f}{\partial \alpha} = \Sigma \alpha + \lambda \alpha + \delta \alpha_1 = 0$$

Durch die Nebenbedingungen $\alpha_1^T \Sigma \alpha = 0$ und $\alpha_1^T \alpha = 0$ (siehe oben) bekommt man

$$\alpha_1^T \frac{\partial f}{\partial \alpha} = \delta \underbrace{\alpha_1^T \alpha_1}_1 = \delta = 0,$$

was bedeutet, dass $\Sigma \alpha = \lambda \alpha$ und α ist wieder ein Eigenvektor von Σ mit Eigenwert λ . Da α orthogonal zu α_1 sein soll und $\text{Var}(\alpha^T X) = \lambda$ maximal sein soll, bekommt man

$$\alpha = \alpha_2 \text{ und } \lambda = \lambda_2 \Rightarrow Y_2 = \alpha_2^T X.$$

□

Übungsaufgabe 4.2.1. Führen Sie den Beweis für $i > 2$ durch!

Sei nun $A = (\alpha_1, \dots, \alpha_n)$. Dies ist eine orthogonale $(n \times n)$ -Matrix, für die gilt (aus dem Satz 4.2.1), dass

$$\Sigma A = A\Lambda, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n),$$

oder, äquivalent dazu,

$$A^T \Sigma A = \Lambda, \quad \Sigma = A\Lambda A^T \quad (4.2.2)$$

Satz 4.2.2. Für eine $(n \times m)$ -Matrix B , mit orthogonalen Spalten b_i , $i = 1, \dots, m$, $m \leq n$, sei $Y = B^T X$ und $\Sigma_Y = \text{Cov}(Y) = B^T \Sigma B$ die Kovarianzmatrix von Y . Dann gilt

$$A_m = \underset{B}{\text{argmax}} \text{Spur}(\Sigma_Y),$$

wobei $A_m = (\alpha_1, \dots, \alpha_m)$.

Beweis. Da $\alpha_1, \dots, \alpha_n$ eine Basis in \mathbb{R}^n bilden, gilt

$$b_k = \sum_{i=1}^n c_{ik} \alpha_i, \quad k = 1, \dots, m,$$

wobei $B = (b_1, \dots, b_m)$, oder, in Matrixform, $B = AC$, mit $C = (c_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, m$. Daher gilt

$$\Sigma_Y = B^T \Sigma B = C^T \underbrace{A^T \Sigma A}_{\Lambda} C = C^T \Lambda C = \sum_{j=1}^n \lambda_j c_j c_j^T,$$

wobei c_j^T die j -te Zeile von C ist. Deshalb gilt

$$\text{Spur}(\Sigma_Y) = \sum_{j=1}^n \lambda_j \text{Spur}(c_j c_j^T) = \sum_{j=1}^n \lambda_j \text{Spur}(c_j^T c_j) = \sum_{j=1}^n \lambda_j |c_j|^2.$$

Da $C = A^{-1}B = A^T B$, gilt

$$C^T C = B^T \underbrace{A A^T}_{I_n} B = \underbrace{B^T B}_{I_m} = I_m,$$

wobei

$$I_k = \text{diag}(\underbrace{1, \dots, 1}_k).$$

Somit

$$\sum_{i=1}^n \sum_{j=1}^m c_{ij}^2 = m,$$

und die Spalten von C sind orthonormal. Daher kann C als ein Teil (erste m Spalten) einer orthonormalen $(n \times n)$ -Matrix D gesehen werden. Da auch die Zeilen von D orthonormale Vektoren sind und c_i^T die ersten m Elemente der Zeilen von D bilden, gilt

$$c_i^T c_i = \sum_{j=1}^m c_{ij}^2 \leq 1, \quad i = 1, \dots, n.$$

Da

$$\text{Spur}(\Sigma_Y) = \sum_{i=1}^n \lambda_i \underbrace{\sum_{j=1}^m c_{ij}^2}_{\beta_i} = \sum_{i=1}^n \beta_i \lambda_i,$$

wobei $\beta_i \leq 1$, $i = 1, \dots, n$, $\sum_{i=1}^n \beta_i = m$, und

$$\lambda_1 > \lambda_2 > \dots > \lambda_n, \quad \sum_{i=1}^n \beta_i \lambda_i \rightarrow \max$$

für $\beta_1 = \dots = \beta_m = 1$, $\beta_{m+1} = \dots = \beta_n = 0$. Aber wenn $B = A_m$, dann gilt

$$c_{ij} = \begin{cases} 1 & , 1 \leq i = j \leq m \\ 0 & , \text{sonst} \end{cases},$$

woraus $\beta_1 = \dots = \beta_m = 1$, $\beta_{m+1} = \dots = \beta_n = 0$ folgt. Somit ist A_m die Lösung von $\text{Spur}(\Sigma_Y) \rightarrow \max_B$. \square

Die Behauptung des Satzes 4.2.2 bedeutet, dass

$$\text{Var} \left(\sum_{i=1}^m Y_i \right) = \text{Var} \left(\sum_{i=1}^m \alpha_i^T X \right)$$

maximal ist für $\forall m = 1, \dots, n$, falls Y_i Hauptkomponenten von X sind.

Folgerung 4.2.1. (*Spektraldarstellung von Σ*). Es gilt

$$\Sigma = \sum_{i=1}^n \lambda_i \cdot \alpha_i \cdot \alpha_i^T \quad (4.2.3)$$

Beweis. Die Darstellung folgt aus (4.2.2), weil

$$\Sigma = (\alpha_1, \dots, \alpha_n) \cdot \text{diag}(\lambda_1, \dots, \lambda_n) \cdot (\alpha_1, \dots, \alpha_n)^T$$

\square

Bemerkung 4.2.1. 1. Da $\lambda_1 > \lambda_2 > \dots > \lambda_n$ mit $|\alpha_i| = 1, \forall i$, folgt aus der Darstellung (4.2.3), dass die ersten Hauptkomponenten nicht nur den Hauptbeitrag zur Varianz von X_i , sondern auch zu den Kovarianzen liefern. Dieser Beitrag wird mit steigendem $i = 1, \dots, n$ immer geringer.

2. Falls $\text{Rang}(\Sigma) = r < n$, dann bedeutet (4.2.3), dass Σ komplett aus ihren ersten r Hauptkomponenten und Koeffizientenvektoren bestimmt werden kann.

Lemma 4.2.1. Sei Σ eine positiv definite symmetrische $(n \times n)$ -Matrix mit Eigenwerten $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ und entsprechenden Eigenvektoren $\alpha_1, \dots, \alpha_n, |\alpha_i| = 1, i = 1, \dots, n$. Dann gilt

$$\lambda_k = \sup_{\alpha \in S_k, \alpha \neq 0} \frac{\alpha^T \Sigma \alpha}{|\alpha|^2},$$

wobei $S_k = \langle \alpha_1, \dots, \alpha_{k-1} \rangle^\perp$ für beliebige $k = 1, \dots, n$.

Beweis. Sei

$$c = \sup_{\alpha \in S_k} \frac{\alpha^T \Sigma \alpha}{|\alpha|^2}.$$

Zeigen wir, dass $\lambda_k \leq c \leq \lambda_k$.

1. $c \geq \lambda_k$: Für $\alpha = \alpha_k$ beweist man

$$c \geq \frac{\alpha_k^T \Sigma \alpha_k}{\alpha_k^T \alpha_k} = \frac{\lambda_k \alpha_k^T \alpha_k}{\alpha_k^T \alpha_k} = \lambda_k$$

2. $c \leq \lambda_k$: Es ist zu zeigen, dass

$$\alpha^T \Sigma \alpha \leq \lambda_k |\alpha|^2, \quad \forall \alpha \in S_k, \quad \alpha \neq 0, \quad \forall \alpha \in \mathbb{R}^n \quad \alpha = \sum_{i=1}^n c_i \alpha_i,$$

weil $\{\alpha_i\}_{i=1}^n$ eine orthonormale Basis bilden.

$$\alpha \in S_k \quad \Rightarrow \quad c_1 = \dots = c_{k-1} = 0,$$

dass heißt

$$\begin{aligned} \alpha &= \sum_{i=k}^n c_i \alpha_i, & \Sigma \alpha &= \sum_{i=1}^n c_i \Sigma \alpha_i = \sum_{i=1}^n c_i \lambda_i \alpha_i, & \alpha^T \Sigma \alpha &= \left(\sum_{i=1}^n c_i \alpha_i \right)^T \left(\sum_{i=1}^n \lambda_i c_i \alpha_i \right) \\ &= \sum_{i,j=1}^n c_i c_j \lambda_i \underbrace{\alpha_j^T \alpha_i}_{\delta_{ij}} = \sum_{i=1}^n c_i^2 \lambda_i, & |\alpha|^2 &= \sum_{i=1}^n c_i^2 \end{aligned}$$

Deshalb gilt für $\alpha \in S_k$

$$\alpha^T \Sigma \alpha = \sum_{i=k}^n c_i^2 \lambda_i \leq \sum_{i=k}^n \lambda_k c_i^2 = \lambda_k \sum_{i=k}^n c_i^2 = \lambda_k |\alpha|^2,$$

und $c \leq \lambda_k$ weil $\lambda_k > \lambda_j$, $j > k$.

□

Satz 4.2.3. Seien B , Y und Σ_Y wie in Satz 4.2.2. Dann gilt

$$A_m = \operatorname{argmax}_B \det(\Sigma_Y),$$

wobei $A_m = (\alpha_1, \dots, \alpha_m)$.

Beweis. Sei $k \in \{1, \dots, m\}$ fixiert. Führen wir $S_k = \langle \alpha_1, \dots, \alpha_{k-1} \rangle^\perp \subset \mathbb{R}^k$ ein (wie in Lemma 4.2.1). Seien $\mu_1 > \mu_2 > \dots > \mu_m$ Eigenwerte von $\Sigma_Y = B^T \Sigma B$ mit entsprechenden Eigenvektoren $\gamma_1, \dots, \gamma_m$, die orthonormiert sind. Sei $T_k = \langle \gamma_{k+1}, \dots, \gamma_m \rangle \subset \mathbb{R}^m$. Es gilt offensichtlich

$$\operatorname{Dim}(S_k) = n - k + 1, \quad \operatorname{Dim}T_k = k.$$

Genau wie in Lemma 4.2.1 kann gezeigt werden, dass $\forall \gamma \neq 0, \gamma \in T_k$ gilt

$$\frac{\gamma^T \Sigma \gamma}{|\gamma|^2} \geq \mu_k.$$

Betrachten wir $\tilde{S}_k = B(T_k) \subset \mathbb{R}^n$. Da B eine orthonormale Transformation ist, ist sie eindeutig und somit $\operatorname{Dim}(S_k) = \operatorname{Dim}(T_k) = k$. Aus der Formel

$$\operatorname{Dim}(S_k \cup \tilde{S}_k) + \operatorname{Dim}(S_k \cap \tilde{S}_k) = \operatorname{Dim}S_k + \operatorname{Dim}\tilde{S}_k$$

folgt

$$\operatorname{Dim}(S_k \cap \tilde{S}_k) = \underbrace{\operatorname{Dim}S_k}_{n-k+1} + \underbrace{\operatorname{Dim}\tilde{S}_k}_k - \underbrace{\operatorname{Dim}(S_k \cup \tilde{S}_k)}_{\leq n} \geq n - k + 1 + k - n = 1$$

das heißt, $\exists \alpha \in S_k \cap \tilde{S}_k$, $\alpha \neq 0$. Für dieses α gilt $\alpha = B\gamma$, $\gamma \in T_k$ und deshalb

$$\mu_k \leq \frac{\gamma^T \Sigma \gamma^2}{|\gamma|^2} = \frac{\gamma^T B^T \Sigma B \gamma}{\underbrace{\gamma^T \gamma}_{\gamma^T B^T B \gamma}} = \frac{\alpha^T \Sigma \alpha}{\alpha^T \alpha} \leq \lambda_k$$

nach $|\gamma| = |B\gamma|$, weil B Distanzen beibehält. Deshalb gilt $\mu_k \leq \lambda_k$ für alle $k = 1, \dots, m$, und

$$\det(\Sigma_Y) = \prod_{i=1}^m \mu_k \leq \prod_{k=1}^m \lambda_k \quad \Rightarrow \quad \max_B \det(\Sigma_Y) \leq \prod_{k=1}^m \lambda_k.$$

Allerdings gilt für $B = A_m$, $\mu_k = \lambda_k$, $k = 1, \dots, m$, deshalb

$$A_m = \operatorname{argmax}_B \det(\Sigma_Y).$$

□

Nun betrachten wir geometrische Eigenschaften von Hauptkomponenten.

Proposition 4.2.1. Die Hauptkomponentenkoeffizienten $\alpha_1, \dots, \alpha_n$ sind die Hauptachsen des Ellipsoids $x^T \Sigma^{-1} x = c$, mit Halbachsenlängen $\sqrt{c\lambda_i}$, $i = 1, \dots, n$.

Beweis. Die Hauptkomponenten von X sind gegeben durch $Z = A^T X$, wobei $A = (\alpha_1, \dots, \alpha_n)$ eine orthonormale Transformation ist, deshalb $A^T = A^{-1}$, $X = AZ$. Daher gilt für unser Ellipsoid

$$x^T \Sigma^{-1} x \quad \underbrace{=}_{\text{Subst. } x=Az} \quad z^T A^T \Sigma^{-1} A z = z^T \Lambda^{-1} z = c,$$

wobei

$$A^T \Sigma^{-1} A = \Lambda^{-1} = \operatorname{diag} \left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n} \right), \quad \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

weil Σ^{-1} dieselben Eigenvektoren mit Eigenwerten $\frac{1}{\lambda_i}$ hat. Daher kann das Ellipsoid $z^T \Lambda^{-1} z = c$ in seiner normierten Form als

$$\sum_{k=1}^n \frac{z_k^2}{c\lambda_k} = 1$$

dargestellt werden. Daraus folgt, dass α_i in die Richtungen seiner Hauptachsen zeigen und, dass seine Halbachsenlängen gleich $\sqrt{c\lambda_i}$ sind. □

Bemerkung 4.2.2. (*Multivariate Normalverteilung*). Falls $X \sim N(0, \Sigma)$ gilt, dann ist $x^T \Sigma^{-1} x = c$ ein Ellipsoid der konstanten Wahrscheinlichkeit für X , weil die Dichte von X

$$f_X(x) = \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} x^T \Sigma^{-1} x \right\} \cdot \frac{1}{(2\pi)^{\frac{n}{2}}}, \quad x \in \mathbb{R}^n,$$

auf diesem Ellipsoid konstant bleibt. Sonst definiert $x^T \Sigma^{-1} x = c$ Konturen der konstanten Wahrscheinlichkeit für X . Dabei zeigt der Vektor α_1 in die Richtung der größten Varianz von $\alpha^T X$ (es ist die größte Hauptachse mit Länge $\sqrt{c\lambda_1}$ des Ellipsoids); α_2 zeigt in die Richtung der zweit größten Varianz (Halbachse $\sqrt{c\lambda_2}$), usw. (vgl. Bedingung 4.2.1).

Bemerkung 4.2.3. Eine andere Form von Hauptkomponentenanalyse ist möglich, wenn man statt $X = (X_1, \dots, X_n)^T$ die normierte Stichprobe $X_\omega = (X_1/\omega_1, \dots, X_n/\omega_n)^T$ benutzt, wobei Gewichte $\omega = (\omega_1, \dots, \omega_n)^T$ eine gewisse Präferenz in der Analyse zum Ausdruck bringen und somit Vorinformationen enthalten. Eine häufige Wahl ist

$$\omega_i = \sqrt{\sigma_{ii}} = \sqrt{\text{Var}X_i},$$

was zur HKA von $X^* = (X_1^*, \dots, X_n^*)$, $X_i^* = \frac{X_i}{\sqrt{\text{Var}X_i}}$, $i = 1, \dots, n$ mit Hilfe der Korrelationsmatrix $\Sigma^* = (\text{Corr}(X_j, X_i))_{i,j=1}$ führt

$$\text{Corr}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}X_i \text{Var}X_j}} = \text{Cov}(X_i^*, X_j^*), \quad i, j = 1, \dots, n.$$

Dabei kommt man auf andere Hauptkomponenten $\alpha_i^{*T} X^*$, für die $\alpha_i^* \neq \alpha_i$ gilt, $i = 1, \dots, n$.

Was sind dann Vor- bzw. Nachteile von HKA basierend auf (X, Σ) und (X^*, Σ^*) ?

Nachteile von (X, Σ) -HKA:

1. Die HKA basierend auf (X^*, Σ^*) hängt nicht von der Wahl der Maßeinheiten von X ab. Somit sind Vergleiche der Ergebnisse von HKA von mehreren Stichproben unterschiedlicher Herkunft möglich.
2. Falls die Varianzen von X_i sehr unterschiedlich sind, so werden die Variablen X_i mit größten Varianzen auch die ersten HK bestimmen, was eindeutig einen Nachteil darstellt. Die HKA basierend auf (X^*, Σ^*) ist frei von diesem Nachteil. Die (X, Σ) -HKA ist in solchen Fällen nicht aussagekräftig, weil sie (in leicht veränderter Form) einfach die Variablen X_i in der Reihenfolge absteigender Varianzen ordnet.

Beispiel 4.2.1. Sei $X = (X_1, X_2)$, wobei X_1 die Länge darstellt und X_2 das Gewicht. X_1 kann in cm oder m gemessen werden, X_2 allerdings nur in kg. In diesen zwei Fällen seien die Kovarianzmatrizen von X gegeben durch

$$\Sigma_1 = \begin{pmatrix} 80 & 44 \\ 44 & 80 \end{pmatrix} \quad \text{bzw.} \quad \Sigma_2 = \begin{pmatrix} 8000 & 4400 \\ 4400 & 8800 \end{pmatrix}.$$

Die Berechnung der ersten HK ergibt in beiden Fällen

$$\alpha_1^T X = 0,707X_1 + 0,707X_2 \quad \text{für } \Sigma_1 \quad \text{bzw.} \quad \alpha_1^T X = 0,998X_1 + 0,055X_2 \quad \text{für } \Sigma_2.$$

Zu bemerken ist, dass im ersten Fall X_1 und X_2 gleiche Beiträge zur 1. HK besitzen, wobei im 2. Fall X_1 den dominierenden Einfluss ausübt. Dazu gilt $\frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot 100\% = 77,5\%$ im ersten Fall und $\frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot 100\% = 99,3\%$ im 2. Fall (es ist der Anteil der Variation der ersten HK von der gesamten Varianz).

3. Falls Zufallsvariable X_i in X unterschiedlicher Herkunft sind (wie im obigen Beispiel), dann ist die Interpretation des Anteils der Variation problematisch, weil in der Summe $\lambda_1 + \dots + \lambda_n$ m^2 , kg^2 , usw. aufsummiert werden. Die HKA basierend auf (X^*, Σ^*) dagegen betrachtet maßlose Größen, so dass die Summe $\lambda_1 + \dots + \lambda_n$ durchaus interpretierbar ist.

Vorteile von (X, Σ) -HKA:

1. Falls statt Σ bzw. Σ^* ihre empirische Analoga $\hat{\Sigma}$ bzw. $\hat{\Sigma}^*$ benutzt werden (wenn $\Sigma(\Sigma^*)$ nicht bekannt sind, müssen sie aus den Daten geschätzt werden), dann hat $(X, \hat{\Sigma})$ -HKA Vorteile, weil die statistischen Methoden hier einfacher sind als bei $(X^*, \hat{\Sigma}^*)$ -HKA.
2. Wenn X_i in X alle dieselbe Maßeinheit besitzen, dann ist die HKA basierend auf (X, Σ) manchmal vorteilhafter, weil bei der Standardisierung von (X, Σ) auf (X^*, Σ^*) der Bezug zu den Einheiten, in denen X gemessen wurde, verloren geht.

Bemerkung 4.2.4. Manchmal wird in Definition 4.2.1 statt $|\alpha| = 1$ die Normierung $|\alpha_k| = \sqrt{\lambda_k}$, $k = 1, \dots, n$ benutzt (siehe Optimierungsaufgabe (4.2.1)). Dies ist insbesondere der Fall in der korrelationsbasierten HKA.

Bemerkung 4.2.5. (*Gleiche Eigenwerte λ_i*). Falls einige Eigenwerte von Σ gleich sind, z.B. $\lambda_1 = \lambda_2 = \dots = \lambda_k > \lambda_{k+1} > \dots > \lambda_m$, bedeutet dies, dass es einen linearen Unterraum der Dimension k gibt, in denen eine beliebige Basis die ersten k Eigenvektoren darstellt. Dies bedeutet, dass für die HKA die ersten k Eigenvektoren nicht eindeutig definiert werden können. Geometrisch interpretiert: Die ersten k Halbachsen von $x^T \Sigma^{-1} x = c$ sind gleich, d.h., das Ellipsoid $x^T \Sigma^{-1} x = c$ hat einen sphärischen k -dimensionalen Durchschnitt durch den Ursprung, in dem die Richtungen der Halbachsen beliebig (orthogonal zueinander) gewählt werden können.

Bemerkung 4.2.6 ($\lambda_i = 0$). Wenn $\lambda_1 > \dots > \lambda_{n-k} > \lambda_{n-k+1} = \dots = \lambda_n = 0$, dann gibt es in der Stichprobe X lediglich $n-k$ linear unabhängige Zufallsvektoren X_i . Deshalb sollten nur diese $n-k$ Variablen zur Analyse benutzt werden.

4.3 Hauptkomponentenanalyse auf Datenebene

Bei diesem Abschnitt wird nicht mehr vorausgesetzt, dass die Kovarianzmatrix Σ bekannt ist. Deshalb soll sie durch die empirische Kovarianzmatrix $\hat{\Sigma}$ ersetzt werden. Seien X^1, X^2, \dots, X^m unabhängige Realisierungen eines n -dimensionalen Zufallsvektors $X = (X_1, \dots, X_n)^T$, $X^i = (X_1^i, \dots, X_n^i)^T$, $i = 1, \dots, m$. X^i wird als Beobachtung von X interpretiert.

Definition 4.3.1. Definiere den n -dimensionalen Zufallsvektor a_k durch

$$a_k = \operatorname{argmax}_{a \in \mathbb{R}^n} \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$$

mit Nebenbedingungen $|a| = 1$, a unkorreliert mit a_1, \dots, a_{k-1} für alle $k = 1, \dots, n$, wobei

$$Y_i = a^T X^i, \quad i = 1, \dots, m, \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i.$$

So definiert $a_k^T X$ die k -ten Hauptkomponenten von X mit Koeffizientenvektor a_k , $Y_{ik} = a_k^T X^i$ ist die Auswertung der k -ten HK auf der i -ten Beobachtung X^i von X , $i = 1, \dots, m$, $k = 1, \dots, n$.

Lemma 4.3.1. Es gilt

$$\frac{1}{m-1} \sum_{i=1}^m (Y_{ik} - \bar{Y}_k)^2 = l_k, \quad k = 1, \dots, n,$$

wobei

$$\bar{Y}_k = \frac{1}{m} \sum_{i=1}^m Y_{ik}, \quad \bar{X}_k = \frac{1}{m} \sum_{i=1}^m X_k^i, \quad k = 1, \dots, n$$

und l_k der Eigenwert der empirischen Kovarianzmatrix $\hat{\Sigma} = (\hat{\sigma}_{ij})_{i,j=1}^n$ ist,

$$\hat{\sigma}_{ij} = \frac{1}{m-1} \sum_{t=1}^m (X_t^i - \bar{X}_i)(X_t^j - \bar{X}_j), \quad i, j = 1, \dots, n, \quad l_1 > l_2 > \dots > l_n.$$

a_k ist der Eigenvektor von $\hat{\Sigma}$ mit Eigenwert l_k , $k = 1, \dots, n$.

Beweis.

Übungsaufgabe 4.3.1. Vergleiche den Beweis des Satzes 4.2.1. □

Im Folgenden werden wir X^i durch $X^i - \bar{X}$ ersetzen und dabei die Bezeichnung X^i beibehalten, $i = 1, \dots, n$.

Bemerkung 4.3.1. Die Eigenschaften der HKA formuliert in Satz 4.2.2, Folgerung 4.2.1, Satz 4.2.3, Proposition 4.2.1 bleiben auch in ihrer statistischen Version (Definition 4.3.1) erhalten, mit folgenden offensichtlichen Modifikationen: Σ wird ersetzt durch $\hat{\Sigma}$, $A = (\alpha_1, \dots, \alpha_n)$ durch $A = (a_1, \dots, a_n)$, $A_m = (\alpha_1, \dots, \alpha_m)$ durch $A_m = (a_1, \dots, a_m)$, Σ_Y durch die empirische Kovarianzmatrix $\hat{\Sigma}_Y$ von Y . So benutzt beispielsweise die Spektraldarstellung von $\hat{\Sigma}$

$$\hat{\Sigma} = \sum_{i=1}^n l_i a_i a_i^T \quad (4.3.1)$$

Übungsaufgabe 4.3.2. Zeigen Sie es!

Zeigen wir eine weitere Eigenschaft der empirischen HKA, die auch als eine äquivalente Definition betrachtet werden kann:

Satz 4.3.1. Sei B eine $(n \times p)$ -Matrix, $p \leq n$, mit orthogonalen Spalten. Seien $Z_i = B^T X^i$, $i = 1, \dots, m$ Projektionen von X^i , $i = 1, \dots, m$, auf einen p -dimensionalen Unterraum L_B . Definiere

$$G(B) = \sum_{i=1}^m |X^i - Z_i|^2.$$

Dann gilt

$$A_p = (a_1, \dots, a_p) = \underset{B}{\operatorname{argmin}} G(B).$$

Beweis. Nach dem Satz von Pythagoras gilt $|X^i|^2 = |Z_i|^2 + |X^i - Z_i|^2$, deshalb

$$G(B) = \sum_{i=1}^m |X^i|^2 - \sum_{i=1}^m |Z_i|^2 \rightarrow \min$$

falls

$$\tilde{G}(B) = \sum_{i=1}^m |Z_i|^2 = \sum_{i=1}^m Z_i^T Z_i = \sum_{i=1}^m X^{iT} B B^T X^i \rightarrow \max_B.$$

Es gilt

$$\begin{aligned} \tilde{G}(B) &= \operatorname{Spur} \left(\sum_{i=1}^m (X^{iT} B B^T X^i) \right) = \sum_{i=1}^m \operatorname{Spur} (X^{iT} B B^T X^i) = \sum_{i=1}^m \operatorname{Spur} (B^T X^i X^{iT} B) \\ &= \operatorname{Spur} \left(B^T \underbrace{\left(\sum_{i=1}^m X^i X^{iT} \right)}_{\mathbb{1}_{(m-1)\hat{\Sigma}}} B \right) = (m-1) \operatorname{Spur} (B^T \hat{\Sigma} B) \end{aligned}$$

Zusammengefasst gilt

$$\tilde{G}(B) = (m-1) \operatorname{Spur} (B^T \hat{\Sigma} B),$$

die nach Bemerkung 4.3.1 und Satz 4.2.2 maximal wird, falls $B = A_p$. \square

¹Da X^i durch $X^i - \bar{X}$ ersetzt wurde.

Bemerkung 4.3.2. Wie kann Satz 4.3.1 als äquivalente Definition der empirischen HKA benutzt werden? a_i werden als orthogonale Vektoren definiert, die einen linearen Unterraum $L_p = \langle a_1, \dots, a_p \rangle$ aufspannen, $p = 1, \dots, n-1$, mit der Eigenschaft, dass die Summe der quadratischen orthogonalen Abstände von X^i zu L_p minimal wird. So wäre es z.B. für $p = 1$ L_1 die beste Gerade, die den Datensatz X^1, \dots, X^m approximiert, für $p = n-1$ wäre L_{n-1} die beste Hyperebene mit derselben Eigenschaft (vgl. lineare Regression).

Der folgende Satz gibt uns gleichzeitig eine effiziente Berechnungsmethode und eine neue Interpretation der HK an.

Satz 4.3.2 (Singulärwertzerlegung). Sei $\tilde{X} = (X^1 - \bar{X}, X^2 - \bar{X}, \dots, X^m - \bar{X})^T$ eine $(m \times n)$ -Matrix, die zentrierte Beobachtungen X^i von X enthält. Sei $\text{Rang}(\tilde{X}) = r \leq n, m$. Es gilt folgende Zerlegung:

$$\tilde{X} = ULA_r^T, \quad (4.3.2)$$

wobei U eine $(m \times r)$ -Matrix mit orthonormalen Spalten ist

$$L = \text{diag}(\tilde{l}_1, \dots, \tilde{l}_r) \quad \text{wobei} \quad \tilde{l}_i = \sqrt{(m-1)l_i}$$

die Wurzel aus dem i -ten (nicht trivialen) Eigenwert von $\tilde{X}^T \tilde{X} = (m-1)\hat{\Sigma}$ ist, $i = 1, \dots, r$. $A_r = (a_1, \dots, a_r)$ ist die $(n \times r)$ -Matrix mit Spalten a_i

Beweis. Definiere $U = (u_1, \dots, u_r)$ mit Spalten $u_i = \frac{\tilde{X} a_i}{\tilde{l}_i}$, $i = 1, \dots, r$. Zeigen wir, dass die Darstellung (4.3.2) gilt. Laut Spektraldarstellung (4.3.1) gilt

$$(m-1)\hat{\Sigma} = \tilde{X}^T \tilde{X} = \sum_{i=1}^r \tilde{l}_i^2 a_i a_i^T, \quad \text{weil} \quad l_i = 0, i = r+1, \dots, n.$$

Deshalb

$$ULA_r^T = U \begin{pmatrix} \tilde{l}_1 a_1^T \\ \vdots \\ \tilde{l}_r a_r^T \end{pmatrix} = \sum_{i=1}^r \tilde{X} \frac{a_i \tilde{l}_i a_i^T}{\tilde{l}_i} = \sum_{i=1}^r \tilde{X} a_i a_i^T \stackrel{l_i=0, i>r}{=} \sum_{i=1}^n \tilde{X} a_i a_i^T$$

$\tilde{X} a_i = 0$, $i = r+1, \dots, n$, wegen $\text{rang}(\tilde{X}) = r$ und Zentrierung der Spalten von \tilde{X} durch \bar{X} . Da die Vektoren a_i orthonormal sind, gilt

$$ULA_r^T = \tilde{X} \sum_{i=1}^n a_i a_i^T = \tilde{X} I = \tilde{X}.$$

□

Bemerkung 4.3.3. Die Matrix U liefert folgende Versionen von Auswertungen

$$Y_{ik} = a_k^T X^i = X^{iT} a_k, \quad Y_{ik} = u_{ik} \tilde{l}_k, \quad i = 1, \dots, m, \quad k = 1, \dots, n$$

Es gilt

$$\text{Var}(u_{ik}) = \frac{\text{Var}(Y_{ik})}{\tilde{l}_k^2} = \frac{l_k}{(m-1)l_k} = \frac{1}{m-1}, \quad \forall i, k$$

4.4 Asymptotische Verteilung von HK bei normalverteilten Stichproben

Sei nun $X \sim N(\mu, \Sigma)$, Σ habe Eigenwerte $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ und entsprechende Eigenvektoren α_k , $k = 1, \dots, n$. Berechne

$$\lambda = (\lambda_1, \dots, \lambda_n)^T, \quad l = (l_1, \dots, l_n)^T, \quad \alpha_k = (\alpha_{k1}, \dots, \alpha_{kn})^T, \quad a_k = (a_{k1}, \dots, a_{kn})^T, \\ k = 1, \dots, n$$

Satz 4.4.1. 1. l ist asymptotisch (für $m \rightarrow \infty$) unabhängig von a_k , $k = 1, \dots, n$.

2. l und a_k , $k = 1, \dots, n$ sind asymptotisch $m \rightarrow \infty$ multivariat normalverteilt, mit asymptotischen Erwartungswerten

$$\lim_{m \rightarrow \infty} \mathbb{E}(l) = \lambda \quad \text{und} \quad \lim_{m \rightarrow \infty} \mathbb{E}(a_k) = \alpha_k, \quad k = 1, \dots, n.$$

3. Es gilt

$$\text{Cov}(l_k, l_{k'}) \sim \begin{cases} \frac{2\lambda_k^2}{m-1}, & k = k' \\ 0, & k \neq k' \end{cases} \quad \text{für } m \rightarrow \infty$$

$$\text{Cov}(a_{kj}, a_{k'j'}) \sim \begin{cases} \frac{\lambda_k}{m-1} \sum_{l=1, l \neq k}^n \frac{\lambda_l \alpha_{lj} \alpha_{lj'}}{(\lambda_l - \lambda_k)^2}, & k = k' \\ -\frac{\lambda_k \lambda_{k'} \alpha_{kj} \alpha_{k'j'}}{(m-1)(\lambda_k - \lambda_{k'})^2}, & k \neq k' \end{cases} \quad \text{für } m \rightarrow \infty.$$

Ohne Beweis!

Die Aussagen von Satz 4.4.1 können dazu benutzt werden, ML-Schätzer sowie Konfidenzintervalle für λ und α_k zu konstruieren.

Übungsaufgabe 4.4.1. 1. Zeige, dass ein ML-Schätzer für Σ durch $\frac{m-1}{m} \hat{\Sigma}$ gegeben ist.

2. Zeige, dass der ML-Schätzer

$$\begin{cases} \text{für } \lambda \text{ ist} & \hat{\lambda} = \frac{m-1}{m} l. \\ \text{für } \alpha_k \text{ ist} & \hat{\alpha}_k = a_k, k = 1, \dots, n. \end{cases}$$

3. Zeige, dass die ML-Schätzer in 2. mit Momenten-Schätzern für λ und α_k übereinstimmen, die aus dem Satz 4.4.1 gewonnen werden können.

Folgerung 4.4.1 (Konfidenzintervalle für λ_k). Ein asymptotisches Konfidenzintervall für λ_k ($m \rightarrow \infty$) zum Niveau $1 - \alpha$ ist gegeben durch

$$\left[l_k \left(1 - \sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}} \right)^{-1}, l_k \left(1 + \sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}} \right)^{-1} \right],$$

wobei m so groß ist, dass $-\sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}} < 1$.

Beweis. Da $l_k \sim N\left(\lambda_k, \frac{2\lambda_k^2}{m-1}\right)$ für $m \rightarrow \infty$ aus Satz 4.4.1, 2. und 3., gilt

$$\frac{l_k - \lambda_k}{\sqrt{\frac{2}{m-1}} \lambda_k} \sim N(0, 1) \quad \text{für } m \rightarrow \infty.$$

Daraus folgt

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(z_{\frac{\alpha}{2}} \leq \frac{l_k - \lambda_k}{\lambda_k} \sqrt{\frac{m-1}{2}} \leq z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha,$$

oder für $m \rightarrow \infty$

$$\sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}} \leq \frac{l_k}{\lambda_k} - 1 \leq \sqrt{\frac{2}{m-1}} \underbrace{z_{1-\frac{\alpha}{2}}}_{=-z_{\frac{\alpha}{2}}},$$

$$\frac{l_k}{1 - \sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}}} \leq \lambda_k \leq \frac{l_k}{1 + \sqrt{\frac{2}{m-1}} z_{\frac{\alpha}{2}}}$$

mit Wahrscheinlichkeit $1 - \alpha$. □

Da alle l_k , $k = 1, \dots, n$ asymptotisch ($m \rightarrow \infty$) unabhängig sind, kann ein simultaner Konfidenzbereich für l als kartesisches Produkt der Konfidenzintervalle für l_k aus Folgerung 4.4.1 angegeben werden.

Lemma 4.4.1. Es gilt

$$(m-1) \alpha_k^T \left(l_k \hat{\Sigma}^{-1} + l_k^{-1} \hat{\Sigma} - 2I_n \right) \alpha_k \xrightarrow[m \rightarrow \infty]{d} \chi_{n-1}^2$$

Ohne Beweis!

Daraus folgt das (asymptotische) Konfidenzellipsoid für α_k zum Niveau $1 - \beta$

$$\left\{ y \in \mathbb{R}^n : (m-1) y^T \left(l_k \hat{\Sigma}^{-1} + l_k^{-1} \hat{\Sigma} - 2I_n \right) y \leq \chi_{n-1, \beta}^2 \right\}.$$

Bemerkung 4.4.1. Folgerung 4.4.1 bzw. Lemma 4.4.1 können zur Konstruktion von statistischen Tests für λ_k bzw. α_k folgendermaßen verwendet werden:

1. Testen von $H_0 : \lambda_k = \lambda_{k_0}$ v.s. $H_1 : \lambda_k \neq \lambda_{k_0}$
Die Hypothese H_0 wird verworfen, falls

$$\left| \frac{l_k - \lambda_{k_0}}{\sqrt{\frac{2}{m-1} \lambda_{k_0}}} > z_{\frac{\alpha}{2}} \right|.$$

Dies ist ein asymptotischer Test ($m \rightarrow \infty$) zum Niveau α .

2. Testen wir $H_0 : \alpha_k = \alpha_{k_0}$ v.s. $H_1 : \alpha_k \neq \alpha_{k_0}$
Die Hypothese H_0 wird abgelehnt, falls

$$(m-1) \alpha_{k_0}^T \left(l_k \hat{\Sigma}^{-1} + l_k^{-1} \hat{\Sigma} - 2I_n \right) \alpha_{k_0} \geq \chi_{n-1, \alpha}^2.$$

Dies ist ein asymptotischer ($m \rightarrow \infty$) Test zum Niveau α .

4.5 Ausreißerkennung

In diesem Abschnitt gehen wir davon aus, dass unsere Stichprobe X^1, X^2, \dots, X^m einige Ausreißer enthalten kann. Was aber ist ein Ausreißer? In der statistischen Literatur gibt es dazu keine einheitliche Meinung. Allgemein würden wir sagen, dass die Beobachtung X^i ein Ausreißer ist, wenn sie einen untypischen Wert (in Bezug auf die Verteilung von X) annimmt. Es kann z.B. ein ungewöhnlich hoher bzw. niedriger Wert von einigen Koordinaten von X^i sein. Es kann aber auch eine ungewöhnliche Kombination von gewöhnlichen Koordinatenwerten einiger Koordinaten von X^i sein. Der Grund für solche untypischen Werte X^i kann ein Messfehler, aber auch eine Anomalie im Datensatz sein.

Beispiel 4.5.1. Sei $X = (X_1, X_2)$, wobei X_1 = "Körpergröße" (in cm) und X_2 = "Gewicht" (in kg) von Kindern im Alter von 5 bis 15 Jahren sind. Das Merkmal X wird in einer medizinischen Studie n mal gemessen. Dabei sind Beobachtungen $X^i = (250, 80)$ und $X^j = (175, 25)$ als Ausreißer klassifiziert worden, und zwar daher, weil $X^i = 250$ cm eine unvorstellbare Körpergröße ist, bei X^j sind sowohl $X_1^j = 175$ als auch $X_2^j = 25$ im mittleren Wertebereich von X_1 und X_2 , ihre Kombination jedoch ist praktisch unmöglich.

Wie könnte man Ausreißer enttarnen? Normalerweise werden untypische Werte von X^i anhand von Plots des Datensatzes X^1, \dots, X^m als Einzelpunkte, die nicht in der großen Punktwolke liegen, identifiziert. Bei hoher Dimension n von X ist es jedoch schwierig, so einen Datensatz zu visualisieren. Deshalb kann man vorschlagen einen Datenpunkt der ersten 2-3 HK von (X^1, \dots, X^m) zu erstellen. Dann werden dort ungewöhnlich große bzw. kleine Werte von X_k^i sofort erkennbar. Um jedoch eine ungewöhnliche Zusammensetzung von gewöhnlichen Koordinatenwerten X_k^i zu entdecken, bedarf es der letzten HK. Dazu wird die Auswertung folgender Statistiken empfohlen:

Seien a_1, \dots, a_n die Koeffizientenvektoren der HK von (X^1, \dots, X^m) . Seien $Y_{ik} = a_k^T X^i$, $i = 1, \dots, m$, $k = 1, \dots, n$ die Auswertungen der HK zu den Beobachtungen X^i . Seien l_k , $k = 1, \dots, n$ die Eigenwerte der empirischen Kovarianzmatrix $\hat{\Sigma}$ von (X^1, \dots, X^m) . Für ein $1 \leq n_0 \leq n$, definieren wir die Statistiken

$$d_i^{(1)}(n_0) = \sum_{k=n-n_0+1}^n Y_{ik}^2, \quad d_i^{(2)}(n_0) = \sum_{k=n-n_0+1}^n \frac{Y_{ik}^2}{l_k}, \quad d_i^{(3)}(n_0) = \sum_{k=n-n_0+1}^n l_k Y_{ik}^2,$$

$$d_i^{(4)}(n_0) = \max_{n-n_0+1 \leq k \leq n} \frac{|Y_{ik}|}{\sqrt{l_k}}, \quad i = 1, \dots, m.$$

Lemma 4.5.1. Es gilt

$$d_j^{(2)}(n) = (X^i - \bar{X})^T \hat{\Sigma}^{-1} (X^i - \bar{X}), \quad i = 1, \dots, m,$$

wobei Y_{ik} an ihren empirischen Mittel gemessen werden, das heißt, Y_{ik} werden durch $Y_{ik} - \bar{Y}_k$ ersetzt, $k = 1, \dots, n$, $i = 1, \dots, m$.

Beweis. Es gilt

$$\hat{\Sigma} = ALA^T, \quad \text{wobei } L = \text{diag}(l_1, \dots, l_n) \quad \text{und} \quad A = (a_1, \dots, a_n).$$

Daher

$$\hat{\Sigma}^{-1} = AL^{-1}A^T \quad \text{mit} \quad L^{-1} = \text{diag}(l_1^{-1}, \dots, l_n^{-1}).$$

Da zusätzlich $Y_i = A^T X^i$ für $Y_i = (Y_{i1}, \dots, Y_{in})^T$, $i = 1, \dots, n$, es gilt

$$X^i = A^{T^{-1}} Y_i = AY_i, \quad X^{iT} = Y_i^T A^T, \quad i = 1, \dots, n$$

und deshalb

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X^i = A\bar{Y}, \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i, \quad \bar{X}^T = \bar{Y}^T A^T.$$

Daher gilt

$$\begin{aligned} (X^i - \bar{X})^T \hat{\Sigma}^{-1} (X^i - \bar{X}) &= (Y_i - \bar{Y})^T \underbrace{A^T A}_I L^{-1} \underbrace{A^T A}_I (Y_i - \bar{Y}) \\ &= (Y_i - \bar{Y})^T L^{-1} (Y_i - \bar{Y}) = \sum_{k=1}^n \frac{Y_{ik}^2}{l_k} = d_i^{(2)}(n). \end{aligned}$$

□

Um nun Ausreißer in (X^1, \dots, X^m) zu erkennen, werden Werte $d_i^{(j)}(n)$, $i = 1, \dots, m$, $j = 1, \dots, n$ für $n = 1, 2, 3$ berechnet. Beobachtungen X^i mit den größten Werten $d_i^{(j)}(n)$ werden als mögliche Ausreißer eingestuft. Zusätzlich kann ein Plot von einer Punktwolke

$$D = \left\{ \left(d_i^{(2)}(n) - d_i^{(2)}(n_0), d_i^{(2)}(n_0) \right), i = 1, \dots, m \right\}$$

dabei behilflich sein. X^i wird hier als Ausreißer erkannt, wenn

$$\left(d_i^{(2)}(n) - d_i^{(2)}(n_0), d_i^{(2)}(n_0) \right)$$

isoliert von der übrigen Punktwolke D liegt.

Bemerkung 4.5.1. Falls $X \sim N(\mu, \Sigma)$ mit bekannten μ und Σ , und HKA auf Modellebene durchgeführt wird, können Verteilungen von $d_i^{(j)}(n_0)$ explizit angegeben werden. Es sind (außer $d_i^{(4)}$) Gamma-Verteilungen mit bekannten Parametern z.B. $d_i^{(2)}(n_0) \sim \chi_{n_0}^2$, $i = 1, \dots, m$. Die Verteilungsfunktion von $d_j^{(4)}(n_0)$ ist $\Phi^{n_0}(x)$, wobei $\Phi(x)$ die Verteilungsfunktion der $N(0, 1)$ -Verteilung ist. Dann können Konfidenzintervalle für $d_i^{(j)}(n_0)$ eine formale Entscheidungsregel dafür liefern, ob X^i einen Ausreißer darstellt. Diese Vorgehensweise basiert zwar auf einer festen mathematischen Grundlage, ist aber in der Praxis wenig einsetzbar, da der Fall von normalverteilten Daten (und dazu mit bekannten Parametern μ und Σ !) äußerst selten vorliegt.

Bemerkung 4.5.2. Statistiken $d_i^{(2)}, d_i^{(4)}$ betonen die letzten Statistiken mehr als $d_i^{(1)}$ (wegen der entsprechenden Normierung). Deshalb sind sie zur Entdeckung von ungewöhnlichen Korrelationen in den Daten geeignet (wie etwa in Beispiel 4.5.1, Beobachtung $X^j = (175, 25)$). Statistik $d_j^{(3)}$ betont die ersten HK. Daher ist sie anzuwenden, um ungewöhnlich große (kleine) Werte von Koordinaten X_k^i zu entdecken ($X_1^i = 250$ im Beispiel 4.5.1).

4.6 Hauptkomponentenanalyse und Regression

Sei folgendes multivariates lineares Regressionsmodell gegeben: $Y = X\beta + \varepsilon$, wobei $Y = (Y_1, \dots, Y_n)^T$ der Vektor der Zielvariablen ist, $X = (X_{ij})_{i=1, \dots, n, j=1, \dots, m}$ die $(n \times m)$ -Matrix der Ausgangsvariablen, $\text{Rang}(X) = m$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ der Vektor der Störgrößen, wobei ε_i unabhängig sind mit $\mathbb{E}\varepsilon_i = 0$, $\text{Var}\varepsilon_i = \sigma^2$, $i = 1, \dots, n$. O.B.d.A. werden wir voraussetzen, dass X (wie in Satz 4.3.2) zentriert ist, d.h., das empirische Mittel der Zeilen von X ist Null, oder, etwas detaillierter, X_{ij} wird erstellt durch $X_{ij} - \overline{X_j}$, wobei

$$\overline{X_j} = \frac{1}{n} \sum_{i=1}^n X_{ij}, \quad j = 1, \dots, m.$$

Wenn einige Variablen X_{ij} in X nahezu linear abhängig sind, das heißt $\det(X^T X) \approx 0$, dann wirkt es sich auf den Schätzer $\hat{\beta}$ von β als hohe Instabilität in seiner Berechnung

aus, weil $\text{Cov}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ (vgl. Satz 4.3.2) sehr geringe Varianzen von $\hat{\beta}_j$ enthalten wird. Ein Ausweg aus dieser Situation wird die Verwendung von Verallgemeinerungen sein wie in Kapitel 4.3. Eine andere Möglichkeit ist es, die HKA für X zu verwenden, um so lineare Abhängigkeiten in X durch die letzten HK zu detektieren und einige Variablen β_j aus der Regression auszuschließen. Genau diese Möglichkeit werden wir in diesem Abschnitt näher beschreiben

Seien a_1, \dots, a_m die Koeffizientenvektoren der HK (das heißt Eigenvektoren) von $X^T X$. Sei $Z_{ik} = a_k^T X^i$ die Auswertung der k -ten HK der i -ten Zeile X^i von X , $i = 1, \dots, n$, $k = 1, \dots, m$. Mit $Z = (Z_{ik})$ gilt $Z = XA$, wobei $A = (a_1, \dots, a_m)$ eine orthogonale $(m \times m)$ -Matrix ist. Stellen wir die Regressionsgleichung $Y = X\beta + \mathcal{E}$ folgendermaßen dar:

$$Y = X \underbrace{AA^T}_I \beta + \mathcal{E} = \underbrace{XA}_Z \underbrace{A^T}_\gamma \beta + \mathcal{E} = Z\gamma + \mathcal{E}, \text{ wobei } \gamma = A^T \beta \text{ ist.} \quad (4.6.1)$$

Somit hat man die alten Ausgangsvariablen β durch ihre Transformierte $\gamma = A^T \beta$ ersetzt. Nun folgt die Schätzung von γ aus Satz 2.2.1:

$$\hat{\gamma} = (Z^T Z)^{-1} Z^T Y = L^{-1} Z^T Y, \quad (4.6.2)$$

wobei $L = \text{diag}(l_1, \dots, l_m)$ die Eigenwerte l_i von $X^T X$ enthält. Dies gilt, weil Z orthogonale Spalten besitzt. Daher gilt

$$\hat{\beta} = A\hat{\gamma} = AL^{-1}Z^T Y = \underbrace{AL^{-1}A^T}_{(X^T X)^{-1}} X^T Y = \sum_{k=1}^m l_k^{-1} a_k a_k^T X^T Y,$$

wobei wir in der letzten Gleichungsmetrik Formeln (4.6.1), (4.6.2) und die Spektraldarstellung (Folgerung 4.2.1) von $(X^T X)^{-1}$ benutzt haben. Aus Satz 4.2.2 folgt außerdem, dass

$$\text{Cov}(\hat{\beta}) = \sigma^2 \sum_{k=1}^m l_k^{-1} a_k a_k^T.$$

Somit haben wir folgendes Ergebnis bewiesen:

Lemma 4.6.1. Die MKQ-Lösung der Regressionsgleichung $Y = X\beta + \mathcal{E}$ ist gegeben durch

$$\hat{\beta} = \sum_{k=1}^m l_k^{-1} a_k a_k^T X^T Y.$$

Dabei gilt

$$\text{Cov}(\hat{\beta}) = \sigma^2 \sum_{k=1}^m l_k^{-1} a_k a_k^T.$$

Bemerkung 4.6.1. Was sind die Vorteile der in (4.6.1)-(4.6.2) eingeführten Vorgehensweise?

1. Nach dem Bestimmen der HK von $X^T X$ ist die Berechnung von $\hat{\gamma} = L^{-1} Z^T Y$ einfach und schnell, weil (4.6.2) keine Inversen Matrizen mehr enthält ($L^{-1} = \text{diag}(l_1^{-1}, \dots, l_m^{-1})$ ist dann explizit bekannt).
2. Wenn einige l_k sehr nahe bei Null sind oder sogar $\text{Rang}(X) < m$ ist, können einige der letzten HK (mit Varianzen, die sehr klein oder gar Null sind) von $X^T X$ einfach von der Regression ausgeschlossen werden. Dies wird durch den neuen Schätzer

$$\tilde{\beta} = \sum_{k=1}^p l_k^{-1} a_k a_k^T X^T Y$$

erreicht, $p < m$.

Lemma 4.6.2. Sei $\text{Rang}(X) = m$:

1. Der Schätzer $\tilde{\beta}$ ist verzerrt:

$$\mathbb{E}\tilde{\beta} = \left(I - \sum_{k=p+1}^m a_k a_k^T \right) \beta$$

2. Es gilt:

$$\text{Cov}(\tilde{\beta}) = \sigma^2 \sum_{k=1}^p l_k^{-1} a_k a_k^T$$

Beweis. 1. Da

$$\tilde{\beta} = \hat{\beta} - \sum_{k=p+1}^m l_k^{-1} a_k a_k^T X^T Y$$

ist und $\hat{\beta}$ erwartungstreu ist, gilt

$$\begin{aligned} \mathbb{E}\tilde{\beta} &= \mathbb{E}\hat{\beta} - \sum_{k=p+1}^m l_k^{-1} a_k a_k^T X^T \mathbb{E}Y = \beta - \sum_{k=p+1}^m l_k^{-1} a_k \underbrace{a_k^T X^T X}_{l_k a_k^T} \beta = \beta - \sum_{k=p+1}^m a_k a_k^T \beta \\ &= \left(I - \sum_{k=p+1}^m a_k a_k^T \right) \beta \end{aligned}$$

2. Wird gezeigt in:

Übungsaufgabe 4.6.1.

□

Geben wir noch eine äquivalente Formulierung der Regression mit Hilfe der HKA. Statt $\gamma = A^T\beta$ zu verwenden, werden wir diesmal von der Singulärwertzerlegung (Satz 4.3.2) für X Gebrauch machen:

$$X = UL^{\frac{1}{2}}A^T,$$

wobei U eine $(n \times m)$ -Matrix mit orthonormalen Spalten ist (die normierte Auswertungen von HK an Zeilen von X enthalten) und $L^{\frac{1}{2}} = \text{diag}(\sqrt{l_1}, \dots, \sqrt{l_m})$. Führen wir die Bezeichnung

$$\delta = L^{\frac{1}{2}}A^T\beta \tag{4.6.3}$$

ein, so gilt

$$Y = X\beta + \mathcal{E} = U \underbrace{L^{\frac{1}{2}}A^T\beta}_{\delta} + \mathcal{E} = U\delta + \mathcal{E}.$$

Der MKQ-Schätzer für δ wäre

$$\hat{\delta} = \underbrace{(U^TU)^{-1}}_I U^TY,$$

weil U orthonormale Spalten besitzt. Aus (4.6.3) folgt $\beta = AL^{-\frac{1}{2}}\delta$ und deshalb

$$\hat{\beta} = AL^{-\frac{1}{2}}\hat{\delta} = AL^{-\frac{1}{2}}U^TY.$$

Dabei ist der Zusammenhang zwischen γ und δ folgender:

$$\gamma = A^T\beta = A^T \left(AL^{-\frac{1}{2}}\delta \right) = \underbrace{A^T A}_{I} L^{-\frac{1}{2}}\delta = L^{-\frac{1}{2}}\delta$$

Wir haben somit folgendes Lemma bewiesen:

Lemma 4.6.3. Die HK-Form $Y = U\delta + \mathcal{E}$ der Regression $Y = X\beta + \mathcal{E}$ hat die MKQ-Lösung $\hat{\delta} = U^TY$ bzw.

$$\hat{\beta} = AL^{-\frac{1}{2}}U^TY. \tag{4.6.4}$$

Dabei ist der Parametervektor δ einfach eine normierte Version von γ : $\delta = L^{\frac{1}{2}}\gamma$

Bemerkung 4.6.2. 1. Da es effiziente Algorithmen zur Berechnung der Singulärwertzerlegung gibt, bietet die Berechnungsformel (4.6.4) klare Rechenvorteile gegenüber der gewöhnlichen Formulierung $\hat{\beta} = (X^TX)^{-1}X^TY$, in der X^TX invertiert werden muss.

2. Statt die letzten $m - p$ HK von $X^T X$ aus der Regression auszuschließen (vgl. Bemerkung 4.6.1, 2.), ist es allgemeiner möglich den Schätzer $\tilde{\beta}$ über einer Teilmenge M von $\{1, \dots, m\}$ zu berechnen:

$$\tilde{\beta}_M = \sum_{k \in M} l_k^{-1} a_k a_k^T X^T Y.$$

Dies benutzt, dass nur HK mit Varianzen l_k , $k \in M$, für die Schätzung berücksichtigt werden. Dann gilt auch

$$\text{Cov}(\tilde{\beta}_M) = \sigma^2 \sum_{k \in M} l_k^{-1} a_k a_k^T,$$

vgl. Übungsaufgabe 4.6.1. Diese Vorgehensweise benutzt den Ausschluss der Komponenten γ_k , $k \notin M$ von $\gamma = (\gamma_1, \dots, \gamma_m)^T$ aus der MKQ-Schätzung. Äquivalent kann man vom Ausschluss der Komponenten δ_k , $k \notin M$ von $\delta = (\delta_1, \dots, \delta_m)^T$ reden, weil $\delta = L^{\frac{1}{2}}$, also $\delta_k = \sqrt{l_k} \gamma_k \forall k$ ist.

Was sind mögliche Strategien zur Wahl der Indexmenge M ?

1. $M = \{k : l_k > l^*\}$ für einen vorgegebenen Schwellenwert $l^* > 0$. Wenn

$$\bar{l} = \frac{1}{m} \sum_{i=1}^m l_i$$

bei 1 liegt, so kann $l^* \in (0, 0.1; 0, 1)$. Der Nachteil dieses Verfahrens liegt darin, dass manche HK, die wichtig für die Vorhersage von Y sind, oft kleine Varianzen besitzen und somit hier aus der Betrachtung ausgeschlossen wurden.

2. Sei σ_{ii}^2 das i -te Diagonalelement von $(X^T X)^{-1}$. Es gilt offensichtlich $\sigma_{ii}^2 = \frac{\text{Var} \hat{\beta}_i}{\sigma^2}$ (vgl. Satz 4.2.2), $i = 1, \dots, m$. Dann kann man $M = \{k : \sigma_{kk}^2 > \sigma^*\}$ wählen für einen geeigneten Schwellenwert σ^* . Zur Wahl von σ^* siehe [8], S. 174. Diese Methode besitzt denselben Nachteil wie 1..
3. $M = \{1, \dots, p\}$, wobei p ist die größte Zahl $\leq m$, für die eines der folgenden Kriterien erfüllt wird:

a) Es gilt:

$$\sum_{i=1}^m \mathbb{E}(\tilde{\beta}_{M_i} - \beta_i)^2 \leq \sum_{i=1}^m \mathbb{E}(\hat{\beta}_i - \beta_i)^2, \quad \forall \beta = (\beta_1, \dots, \beta_m)^T \in \mathbb{R}^m \quad (4.6.5)$$

b) Es gilt:

$$\mathbb{E}(c^T \tilde{\beta}_M - c^T \beta)^2 \leq \mathbb{E}(c^T \hat{\beta} - c^T \beta)^2 \quad \forall \beta \in \mathbb{R}^m, c \in \mathbb{R}^m$$

c) Es gilt:

$$\mathbb{E} \left| X\tilde{\beta}_M - X\beta \right|^2 \leq \mathbb{E} \left| X\hat{\beta} - X\beta \right|^2$$

Dabei orientiert sich das Kriterium a) an der Aufgabe, β möglichst präzise zu schätzen. Kriterien b) und c) dagegen erzielen das beste Ergebnis bei der Vorhersage von $\mathbb{E}Y = X\beta$ durch $X\hat{\beta}_M$ bzw. $X\hat{\beta}$. Alle Größen in a)-c) sind mittlere quadratische Fehler, die sowohl den Bias als auch die Varianzen von $\tilde{\beta}_M$ berücksichtigen.

In der statistischen Literatur sind viele weitere Strategien beschrieben, die in konkreten Situationen einen verbesserten Schätzer $\tilde{\beta}_M$ im Vergleich zu $\hat{\beta}$ erzielen. Die Fragestellung der optimalen Wahl von M ist jedoch immer noch offen.

Eine Alternative zur Einschränkung der Menge von HK in der Regression (das heißt zum Ausschluss von HK mit $l_k \approx 0$) ist der folgende Schätzer $\tilde{\beta}_R$:

$$\tilde{\beta}_R = \sum_{k=1}^m (l_k + K_k)^{-1} a_k a_k^T X^T Y,$$

wobei $K_1, \dots, K_m > 0$ Gewichte sind, die eine zusätzliche Auswahl von Einflussgrößen in der Regression darstellen. Durch diese Gewichte wird erreicht, dass $l_k \approx 0$ keinen destabilisierenden Einfluss auf die Schätzung mehr ausüben.

Übungsaufgabe 4.6.2. Zeigen Sie, dass 2)

$$\text{Cov}(\tilde{\beta}_R) = \sigma^2 \sum_{k=1}^m \frac{l_k}{(l_k + K_k)^2} a_k a_k^T$$

1) $\tilde{\beta}_R$ ist ein verzerrter Schätzer von β . Finden Sie den Bias von $\tilde{\beta}_R$!

Die Bezeichnung $\tilde{\beta}_R$ steht für (Engl.) *Ridge Regression*. Hier stellt sich die Frage der Wahl von K_k , $k = 1, \dots, m$. In der Praxis wird oft empfohlen, $K_k = K$, $k = 1, \dots, m$, wobei K klein ist, zu wählen.

Noch eine Anwendung der HKA in der Regression wird durch die sogenannte latente Wurzel-Regression (Engl. latent root regression) gegeben. Diese Art der Regression versucht, nur solche HKA zu eliminieren, die gleichzeitig kleine Varianzen l_k besitzen und keinen Wert für die Vorhersage von $\mathbb{E}Y$ durch $X\beta$ darstellen. Dabei wird die HKA an der $(m+1) \times (m+1)$ -Matrix $\tilde{X}^T \tilde{X}$ mit $\tilde{X} = (Y, X)$ durchgeführt. Seien \tilde{a}_k , $k = 0, \dots, m$ die Koeffizienten der HK von $\tilde{X}^T \tilde{X}$, mit entsprechenden Eigenwerten \tilde{l}_k , $k = 0, \dots, m$. Sei dabei $\tilde{a}_k = (a_{k0}, \dots, a_{km})^T$, $k = 0, \dots, m$.

Definieren wir die Indexmenge der auszuschließenden HK als $M_L = \{k = 0, \dots, m : \tilde{l}_k \leq l^*, |a_{k0}| \leq a^*\}$. Dies ist die Indexmenge von solchen HK, die kleine Varianzen besitzen und keinen großen Einfluss auf die Prognose von Y ausüben. Sei $M = \{0, \dots, m\} \setminus M_L$. Definiere

$$\hat{\beta}_L = \sum_{k \in M} \tilde{c}_k \tilde{a}_k, \quad \text{wobei} \quad \{\tilde{c}_k, k \in M\} = \underset{\beta}{\operatorname{argmin}} |Y - X\beta|^2, \quad \text{mit} \quad \beta = \sum_{k \in M} c_k \tilde{a}_k$$

Satz 4.6.1. Es gilt

$$\tilde{c}_k = -\frac{a_{k0} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}{\tilde{l}_k \sum_{i \in M} \frac{a_{i0}^2}{\tilde{l}_i}}, \quad k \in M$$

Ohne Beweis!

Schwellenwerte l^* und a^* sind immer noch empirisch zu wählen.

4.7 Numerische Berechnung der Hauptkomponenten

Um zu verstehen, was statistische Software-Pakete bei der Berechnung von HK tun, ist es wichtig, einige Algorithmen dazu zu kennen. Dabei wird man sich darüber im Klaren, warum manchmal die Ergebnisse schlecht sind (z.B. bei Eigenwerten, die fast gleich sind) oder welche Einschränkungen diese Algorithmen an die Größe der zu bearbeitenden Datensätze (in Speicher und/oder Laufzeit) implizieren. Wir werden hier eine kurze Übersicht dieser Methoden geben. Da die HKA im Wesentlichen darauf basiert, Eigenwerte λ_i und Eigenvektoren α_i einer positiv semi-definiten $(m \times m)$ -Matrix Σ zu berechnen, werden wir uns mit dieser Berechnung beschäftigen.

Sei also Σ eine $(m \times m)$ -Matrix mit den Eigenvektoren $\alpha_1, \dots, \alpha_m$ und Eigenwerten $\lambda_1, \dots, \lambda_m$, die positiv semi-definit ist. In der Fachliteratur sind mindestens 4 Methoden zur Berechnung von α_i und λ_i bekannt:

1. Potenzmethode
2. QR-Zerlegung
3. Singulärwertzerlegung
4. Neuronale Netzwerke

Wir werden hier kurz nur die Essenz der Potenzmethode erwähnen: diese stellt einen iterativen Algorithmus zum Auffinden von λ_1 und α_1 dar, falls $\lambda_1 \gg \lambda_2 > \dots > \lambda_m$. Sei u_0 der Anfangsvektor aus \mathbb{R}^m . Schreibe $u_r = \Sigma u_{r-1} = \Sigma^r u_0$ für alle $r \in \mathbb{N}$. Wenn

$$u_0 = \sum_{i=1}^m c_i \alpha_i$$

in der Orthonormalbasis $\alpha_1, \dots, \alpha_m$ Koordinaten c_1, \dots, c_m besitzt, dann gilt

$$u_r = \Sigma^r u_0 = \sum_{i=1}^m c_i \Sigma^r \alpha_i = \sum_{i=1}^m c_i \lambda_i^r \alpha_i, \quad r \in \mathbb{N}$$

Sei $u_r = (u_{r1}, \dots, u_{rm})^T$, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{im})^T$.

Lemma 4.7.1. Es gilt

$$\frac{u_{ri}}{u_{r-1,i}} \xrightarrow{r \rightarrow \infty} \lambda_1$$

für $i = 1, \dots, m$ und

$$\frac{u_r}{c_i \lambda_1^r} \xrightarrow{r \rightarrow \infty} \alpha_1$$

Beweis. Für $j = 1, \dots, m$ gilt

$$u_{rj} = \sum_{i=1}^m c_i \lambda_i^r \alpha_{ij}$$

und deshalb

$$\begin{aligned} \frac{u_{rj}}{u_{r-1,j}} &= \frac{\sum_{i=1}^m c_i \lambda_i^r \frac{\alpha_{ij}}{\lambda_1^{r-1}}}{\sum_{i=1}^m c_i \lambda_i^{r-1} \frac{\alpha_{ij}}{\lambda_1^{r-1}}} \\ &= \frac{c_1 \alpha_{1j} \lambda_1 + \sum_{i=2}^m c_i \left(\frac{\lambda_i}{\lambda_1}\right)^{r-1} \lambda_i \alpha_{ij}}{c_1 \alpha_{1j} + \sum_{i=2}^m c_i \left(\frac{\lambda_i}{\lambda_1}\right)^{r-1} \alpha_{ij}} \xrightarrow{r \rightarrow \infty} \frac{c_1 \alpha_{1j}}{c_1 \alpha_{1j}} \lambda_1 = \lambda_1, \quad \text{weil } \frac{\lambda_i}{\lambda_1} < 1, i = 2, \dots, n \end{aligned}$$

weiterhin,

$$\frac{u_r}{u \lambda_1^r} = \alpha_1 + \sum_{i=2}^m \frac{c_i}{c_1} \left(\frac{\lambda_i}{\lambda_1}\right)^r \alpha_i \xrightarrow{r \rightarrow \infty} \alpha_1.$$

□

Die Tatsache, dass c_1 unbekannt ist, soll uns nicht stören, denn $\frac{u_r}{\lambda_1^r}$ kann zum Einheitsvektor normiert werden. Aus dem Beweis des Lemmas 4.6.3 wird klar, dass die Konvergenz-geschwindigkeit von $\frac{u_{ri}}{u_{r-1,i}}$ gegen λ_1 und von $\frac{u_r}{c_1 \lambda_1^r}$ gegen α_1 genau dann schlechter wird, wenn $\lambda_1 \approx \lambda_2$, wenn also $\frac{\lambda_2}{\lambda_1} \approx 1$.

Was wäre aber im Fall $\lambda_1 \approx \lambda_2$ zu tun, um die Konvergenz des Verfahrens zu beschleunigen? Statt Σ kann man in den Iterationen $\Sigma - \rho I$ verwenden, um das Verhältnis $\frac{\lambda_2 - \rho}{\lambda_1 - \rho}$ kleiner zu machen. Oder, statt Σ verwendet man $(\Sigma - \rho I)^{-1}$, das heißt, man löst das Gleichungssystem $(\Sigma - \rho I) u_r = u_{r-1}$ für jedes $r \in \mathbb{N}$. Somit ist für die geeignete Wahl von ρ die Konvergenz zu α_k , $k = 1, \dots, m$ möglich (im zweiten Fall).

Übungsaufgabe 4.7.1. Konstruieren Sie diese Vektoren und beweisen Sie die Konvergenz!

Eine Beschleunigung der Konvergenz kann auch erreicht werden, wenn statt $\{u_r\}$ die Folge $\{u_{2^r}\}$ betrachtet wird, $u_{2^r} = T^{2^r} u_0$, $r \in \mathbb{N}$. Weitere Maßnahmen zur Verbesserung des Algorithmus der Potenzmethode findet man in [8], S. 410-411.

Literaturverzeichnis

- [1] BICKEL, P. ; DOKSUM, K.: *Mathematical Statistics: Basic Ideas and Selected Topics*. 2nd edition, volume 1. London : Prentice Hall, 2001
- [2] CASELLA, G. ; BERGER, R. L.: *Statistical Inference*. 2nd edition. Duxbury : Pacific Grove (CA), 2002
- [3] DOBSON, A.J.: *An Introduction to Generalized Linear Models*. Chapman & Hall, Boca Raton, 2002
- [4] FAHRMEIR, L. ; KÜNSTLER, R. ; I. PIGEOT, G. T.: *Statistik. Der Weg zur Datenanalyse*. 3. Auflage. Berlin : Springer, 2001
- [5] GEORGI, H. O.: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Berlin : de Gruyter, 2002
- [6] HARTUNG, J. ; ELPERT, B. ; KLÖSENER, K. H.: *Statistik*. München : R. Oldenbourg Verlag, 1993. – 9. Auflage
- [7] IRLE, A.: *Wahrscheinlichkeitstheorie und Statistik, Grundlagen - Resultate - Anwendungen*. Teubner, 2001
- [8] JOLLIFFE, I. T.: *Principal Component analysis*. 2nd edition. Springer, 2002
- [9] KOCH, K. R.: *Parameter Estimation and Hypothesis Testing in Linear Models*. Berlin : Springer, 1999
- [10] KRENGEL, U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig : Vieweg, 2002. – 6. Auflage
- [11] L. FAHRMEIR, T. K. ; LANG, S.: *Regression. Modelle, Methoden und Anwendungen*. Berlin : Springer, 2007
- [12] LEHMANN, E. L.: *Testing Statistical Hypothesis*. New York : Springer, 1999
- [13] MAINDONALD, J. ; BRAUN, J.: *Data Analysis and Graphics Using R*. Cambridge University Press, 2003
- [14] PRUSCHA, H.: *Angewandte Methoden der Mathematischen Statistik*. Stuttgart : Teubner, 2000
- [15] PRUSCHA, H.: *Vorlesungen über Mathematische Statistik*. Stuttgart : Teubner, 2000

- [16] SACHS, L.: *Angewandte Statistik*. Springer, 1992
- [17] SACHS, L. ; HEDDERICH, J.: *Angewandte Statistik, Methodensammlung mit R*. 12. Auflage. Berlin : Springer, 2006
- [18] SPIEGEL, M. R. ; STEPHENS, L. J.: *Statistik*. 3. Auflage. McGraw-Hill, 1999
- [19] STAHEL, W. A.: *Statistische Datenanalyse*. Vieweg, 1999
- [20] VENABLES, W. ; RIPLEY, D.: *Modern applied statistics with S-PLUS*. 3rd edition. Springer, 1999
- [21] WASSERMAN, L.: *All of Statistics. A Concise Course in Statistical Inference*. Springer, 2004

Index

- Ablehnungsbereich, 4
- AIC-Kriterium, 122
- analysis of variance, *siehe* Varianzanalyse
- Annahmehbereich, 4
- ANOVA, *siehe* Varianzanalyse
- asymptotische Tests, 115

- best linear unbiased estimator (BLUE), 70
- bester linearer erwartungstreuer Schätzer, 70
- Bestimmtheitsmaß, 79
- bilineare Form, 60
- Binomialverteilung, 26
- Bonferroni-Ungleichung, 81

- Design-Matrix, 55, 68

- Effekt, 100
- Eindeutigkeitssatz
 - für charakteristische Funktionen, 57
 - für momenterzeugende Funktionen, 64
- einparametrische Exponentialklasse, 25
- Entscheidungsregel, 3
- Exponentialfamilie, 103

- Faltungstabilität der multivariaten Normalverteilung, 59
- Fehler 1. und 2. Art, 5
- Fisher Scoring, 120
- Fisher-Informationsmatrix, 43, 110, 121

- Gütefunktion, 5
- Satz von Gauß-Markov, 91
- gemischte Momente, 61

- Hauptsatz über zweiseitige Tests, 34
- Hesse-Matrix, 110
- Hypothese, 3
 - Alternative, 3
 - Haupthypothese, 3
 - testbare, 96

- Informationskoeffizient von Akaike, 122
- Informationsmatrix von Fisher, 43
- Iterationstest, 52

- Karl Popper, 4
- klassenspezifische Differenzen, 100
- Klassenstärke, 36
- klassische ANOVA-Hypothese, 101
- kritischer Bereich, *siehe* Ablehnungsbereich

- Likelihood-Ratio-Test, 121
- lineare Form, 60
- lineare Regression, 55
 - einfache, 70
 - multiple, 70
 - ohne vollen Rang, 84
 - multivariate mit vollem Rang, 68
- Lineare Transformation von $N(\mu, K)$, 59
- Linkfunktion, 103
 - natürliche, 107
- Logit-Modell, 108, 118, 124

- Methode der kleinsten Quadrate, 68
- MKQ-Schätzer, 68
- Modelle
 - verallgemeinerte lineare, 103
- Multinomialverteilung, 36

- Newton-Verfahren, 114

- Neyman-Pearson
 - Fundamentallemma, 21
 - Optimalitätssatz, 20
- nicht-zentrale $\chi^2_{n,\mu}$ -Verteilung, 64
- Normalgleichung, 69
- Normalverteilung
 - multivariate, 55
 - Signifikanztests, 12
- Odd, 108
- p -Wert, 9
- Pearson-Teststatistik, 37
- Poisson-Modell, 124
- Poisson-Regression, 114
- Poissonverteilung, 15, 17
 - Neyman-Fisher-Test, 47
 - Neyman-Pearson-Test, 23
- Probit-Modell, 108
- quadratische Form, 60
 - Kovarianz, 61
- Quantilfunktion der Normalverteilung, 108
- Randomisierungsbereich, 4
- Regression
 - binäre kategoriale, 108
 - logistische, 108, 114
- Residuum, 78
- Reststreuung, 79
- Score-Funktion, 121
- Score-Statistik, 121
- Störgrößen, 68
- Stufe eines Einflussfaktors, 100
- Test
 - Anpassungstest, 35
 - Anpassungstest von Shapiro, 48
 - asymptotischer, 7, 14
 - auf Zusammenhang, 78
 - besserer, 18
 - besten, 18
 - Binomialtest, 50
 - χ^2 -Anpassungstest, 36
 - χ^2 -Pearson-Fisher-Test, 42
 - für Regressionsparameter, 78
 - Iterationstest, 52
 - Kolmogorov-Smirnov, 36
 - Macht, 5
 - Monte-Carlo-Test, 7
 - Neyman-Pearson-Test, 19
 - Ablehnungsbereich, 19
 - einseitiger, 24
 - modifizierter, 32
 - Parameter der Poissonverteilung, 23
 - Umfang, 19
 - NP-Test, *siehe* Neyman-Pearson-Test
 - Parameter der Normalverteilung, 12
 - parametrischer, 5
 - einseitiger, 6
 - linksseitiger, 6
 - rechtsseitiger, 6
 - zweiseitiger, 6
 - parametrischer Signifikanztest, 12
 - power, *siehe* Macht
 - randomisierter, 4, 17
 - Schärfe, 5
 - von Shapiro-Francia, 49
 - von Shapiro-Wilk, 50
 - Stärke, 5
 - Umfang, 18
 - unverfälschter, 10
 - Wald-Test, 14
 - von Wald-Wolfowitz, 53
- Variabilität der Erwartungswerte, 100
- Varianzanalyse, 100
 - einfaktorielle, 100
 - zweifaktorielle, 102
- verallgemeinerte Inverse Matrix, 85
- Verfahren von Cramér-Wold, 57
- Verteilung mit monotonem Dichtekoeffizienten, 24
- Wald-Statistik, 121