

Grundlagen der Wahrscheinlichkeitstheorie und Statistik

Prof. Dr. Viktor Sandor



Köln
28./29. Juni 2006

Inhaltsverzeichnis

1	Notationen	1
1.1	Die Menge $\overline{\mathbb{R}}$	1
1.2	Monotonie	1
1.2.1	Mengensysteme	1
1.2.2	Funktionenfolgen	1
I	Maßtheorie und Integration	2
2	Grundlagen der Maß- und Integrationstheorie	2
2.1	Motivation	2
2.1.1	Maßtheorie: Das Programm	2
2.1.2	Integrationstheorie: Das Programm	5
2.2	Mengensysteme	7
2.3	Eigenschaften von Maßen	9
2.4	Konstruktion von Maßen	9
2.4.1	Existenz und Eindeutigkeit von Fortsetzungen	9
2.4.2	Das Lebesgue-Borel Maß	10
2.5	Messbare Funktionen	11
2.6	Integration	15
2.6.1	Konstruktion	15
2.6.2	Nullmengen	18
2.6.3	Konvergenzsätze	21
2.6.4	$L^1(\Omega)$ und $L^2(\Omega)$	21
2.7	Lebesgue- und Riemann-Integral	22
2.8	Produktmaße	22
2.8.1	Konstruktion	22
2.8.2	Integration bezüglich Produktmaßen	23
2.9	Dichten	24
2.10	Bildmaße	26
II	Wahrscheinlichkeitstheorie	28
3	Grundlagen der Wahrscheinlichkeitsrechnung	28
3.1	Zufällige Ereignisse	28
3.2	Axiome der Wahrscheinlichkeitstheorie	29
3.3	Kombinatorik	32

4	Zufallsvariablen und Wahrscheinlichkeitsverteilungen	33
4.1	Zufallsvariablen	33
4.2	Verteilungsfunktionen	34
4.2.1	Maßerzeugende Funktionen	34
4.2.2	Grundlagen	35
4.2.3	Diskrete und stetige Zufallsvariablen	37
4.3	Maßzahlen für Verteilungen	38
4.3.1	Erwartungswert	38
4.3.2	Varianz	41
4.3.3	Schiefe	42
4.3.4	Höhere Momente	42
4.3.5	Quantile	42
4.4	Ungleichungen	43
4.5	Diskrete Verteilungen	45
4.5.1	Die Binomialverteilung	45
4.5.2	Die negative Binomial-Verteilung	45
4.5.3	Die Poisson-Verteilung	46
4.5.4	Die hypergeometrische Verteilung	47
4.6	Stetige Verteilungen	48
4.6.1	Die Gleichverteilung	48
4.6.2	Die Exponentialverteilung	49
4.6.3	Die Normalverteilung	49
4.6.4	Die Gamma-Verteilung	50
4.7	Erzeugende Funktionen	51
4.7.1	Die wahrscheinlichkeitserzeugende Funktion	51
4.7.2	Momentenerzeugende Funktion	52
4.7.3	Charakteristische Funktion	54
4.8	Mehrdimensionale Zufallsvariablen	56
4.8.1	Kovarianz und Korrelation	58
4.8.2	Die n -dimensionale Normalverteilung	59
4.9	Hilfssaussagen	60
4.9.1	Konvexe Funktionen	60
4.9.2	Die Γ -Funktion	61
5	Unabhängigkeit und bedingte Wahrscheinlichkeiten	62
5.1	Bedingte Wahrscheinlichkeit	62
5.2	Unabhängige Ereignisse	64
5.3	Die Formel von Bayes	66
5.4	Unabhängige σ -Algebren	68
5.5	Unabhängige Zufallsvariablen	68
5.5.1	Grundlagen	68

5.5.2	Produkte von Zufallsvariablen	70
5.5.3	Summen von Zufallsvariablen	71
5.5.4	Erzeugende Funktionen	74
5.5.5	Beispiele Faltung	75
5.5.6	Eigenschaften der mehrdimensionalen Normalverteilung	75
6	Bedingte Verteilungen und Momente	77
6.1	Diskrete Zufallsvariablen	77
6.2	Definition im allgemeinen Fall	78
6.3	Eigenschaften	81
6.4	Darstellung	82
6.5	Berechnungen durch Bedingen	87
6.5.1	Erwartungswerte	87
6.5.2	Wahrscheinlichkeiten	88
6.6	Bedingte Wahrscheinlichkeiten	89
6.7	Bedingte Unabhängigkeit	92
6.8	Gestutzte Verteilungen	92
6.9	Hilfssätze	94
7	Grenzwertsätze	95
7.1	Gesetze der großen Zahlen	95
7.2	Der Zentrale Grenzwertsatz	97
7.3	Der Zentralsatz der Statistik	99
7.3.1	Empirische Verteilungsfunktion	99
7.3.2	Satz von Glivenko Cantelli	100
III	Statistik	101
8	Punktschätzer	103
8.1	Begriff	103
8.2	Eigenschaften von Schätzern	104
8.3	Schätzen von Erwartungswert und Varianz	105
8.4	Momenten-Schätzer	107
8.5	Das Maximum-Likelihood-Prinzip	109
8.6	Zusammenfassung	112
9	Intervallschätzer	114
9.1	Testverteilungen	114
9.1.1	χ^2 -Verteilung	115
9.1.2	t -Verteilung	116

9.1.3	F -Verteilung	117
9.2	Konfidenzintervalle für Parameter der Normalverteilung	118
9.2.1	Konfidenzintervall für μ mit bekannter Varianz	118
9.2.2	Konfidenzintervall für μ mit unbekannter Varianz	121
9.2.3	Konfidenzintervall für σ^2	123
9.3	Konfidenzintervalle für eine Wahrscheinlichkeit	125
10	Hypothesentests	127
10.1	Konzepte	127
10.1.1	Fehlerwahrscheinlichkeiten	128
10.1.2	Gütefunktion	128
10.1.3	p -Werte	128
10.1.4	Zusammenhang mit Konfidenzintervallen	129
10.2	Test bei Normalverteilungsannahme	130
10.2.1	Gauß-Tests	130
10.2.2	t -Tests	132
10.2.3	χ^2 -Streuungstest	134
10.2.4	F -Test	135
10.3	Asymptotischer Binomialtest	135
10.4	Anpassungstests	136
10.4.1	Kolmogorov-Smirnov	136
10.4.2	χ^2 -Anpassungstest	137
10.5	χ^2 -Unabhängigkeitstest	142
11	Lineare Regression	145
11.1	Modell und Parameterschätzung	145
11.2	Tests und Konfidenzintervalle	146
11.2.1	Testen der Regressionsparameter	147
11.2.2	Konfidenzintervalle	147
11.2.3	Prognoseintervalle	148

1 Notationen

1.1 Die Menge $\overline{\mathbb{R}}$

Es sei $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ mit der Ordnung

$$-\infty < a < \infty \text{ für alle } a \in \mathbb{R}.$$

Mit dieser Festlegungen besitzt jedes $A \subset \overline{\mathbb{R}}$ ein Supremum und ein Infimum in $\overline{\mathbb{R}}$. Intervalle werden mit $[a, b]$, $[a, b)$, $(a, b]$, (a, b) bezeichnet wobei $a \leq b$, $a, b \in \overline{\mathbb{R}}$. Es wird folgenderweise gerechnet:

(a) Addition

$$\begin{aligned} a + (\pm\infty) &= (\pm\infty) + a = \pm\infty, & a \in \mathbb{R} \\ (+\infty) + (+\infty) &= (+\infty), (-\infty) + (-\infty) = (-\infty) \end{aligned}$$

Nicht definiert sind $+\infty + (-\infty)$ und $-\infty + (+\infty)$.

(b) Multiplikation

$$\begin{aligned} a \cdot (\pm\infty) &= (\pm\infty) \cdot a = \pm\infty, & a \in (0, \infty] \\ a \cdot (\pm\infty) &= (\pm\infty) \cdot a = \mp\infty, & a \in [-\infty, 0) \\ 0 \cdot (\pm\infty) &= (\pm\infty) \cdot 0 := 0 \end{aligned}$$

1.2 Monotonie

1.2.1 Mengensysteme

Seien Mengen $A_1, A_2, \dots \subset \Omega$. Dann definieren wir

$$\begin{aligned} A_n \uparrow A &:\Longleftrightarrow A_1 \subset A_2 \subset \dots, A = \bigcup_{n \in \mathbb{N}} A_n \\ A_n \downarrow A &:\Longleftrightarrow A_1 \supset A_2 \supset \dots, A = \bigcap_{n \in \mathbb{N}} A_n \end{aligned}$$

1.2.2 Funktionenfolgen

Sei $(f_n)_{n \in \mathbb{N}}$ eine Folge von Funktionen $f_n : \Omega \longrightarrow \mathbb{R}$.

$$\begin{aligned} f_n \uparrow f &:\Longleftrightarrow f_1 \leq f_2 \leq \dots, f(x) = \lim_{n \rightarrow \infty} f_n(x) \\ f_n \downarrow f &:\Longleftrightarrow f_1 \geq f_2 \geq \dots, f(x) = \lim_{n \rightarrow \infty} f_n(x) \end{aligned}$$

Teil I

Maßtheorie und Integration

2 Grundlagen der Maß- und Integrationstheorie

In diesem Kapitel sollen die grundlegenden Fragen der Maß- und Integrationstheorie aufgezeigt werden.

2.1 Motivation

2.1.1 Maßtheorie: Das Programm

Die Maßtheorie entstand aus der Notwendigkeit, anschauliche Verfahren wie Längen-, Flächen- und Volumenmessung auf eine mathematische Grundlage zu stellen.

Nehmen wir zum Beispiel das Problem, Teilmengen des $\mathbb{R}^2$ eine „Fläche“ zuzuordnen. Bei der Lösung dieses Problems sollten einige elementare Eigenschaften erhalten bleiben wie:

- Ist eine Fläche F aus zwei Teilstücken F_1 und F_2 zusammengesetzt, so sollten sich die Flächeninhalte addieren.
- Der leeren Menge muss der Inhalt Null zugeordnet werden.
- Für gewisse Teilmengen, z.B. Rechtecke und Dreiecke, ist völlig klar, was ihr Flächeninhalt ist.
- ...

Wir spalten das Problem in zwei Teilprobleme auf, nämlich

- (a) Vorgelegt sei eine Menge Ω (im vorigen Beispiel war $\Omega = \mathbb{R}^2$). Welche Eigenschaften soll das System der Teilmengen haben, denen wir einen Inhalt zuordnen wollen?
- (b) Falls (a) geklärt ist, welche Eigenschaften soll die eigentliche Inhaltsmessung haben?

Zunächst zu (a)

Man könnte zunächst versuchen, alle Teilmengen von Ω zu behandeln. Eine derart weitreichende Theorie ist im Allgemeinen nicht zu erwarten (Banach

Tarski Paradoxon s. S. 4). Was aber ist wesentlich zu fordern, was sind „plausible“ Bedingungen? Wie in anderen Bereichen der Mathematik auch wird die Beantwortung auf einen Kompromiss hinauslaufen: die Forderungen sollen einerseits möglichst schwach sein, um die Theorie für viele Anwendungsfälle verfügbar zu machen; andererseits sind bei sehr schwachen Bedingungen keine interessanten Ergebnisse zu erwarten. Heute herrscht allgemein die Meinung vor, dass das System der mit einem Inhalt zu versehenen Mengen mit den in der nachstehenden Definition zusammengefassten Eigenschaften ausgestattet sein sollte.

Definition 2.1.1 Sei Ω eine Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$ ein System von Teilmengen. $\mathcal{A}$ heißt **σ -Algebra**, falls

- (a) $\emptyset, \Omega \in \mathcal{A}$
- (b) $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$
- (c) Sind $A_1, A_2, \dots$ Elemente von $\mathcal{A}$ so auch $\bigcup_{n \geq 1} A_n$

Ist $\mathcal{A}$ eine σ -Algebra auf Ω , so wird statt $A \in \mathcal{A}$ auch A **ist messbar** gesagt. Das Tupel $(\Omega, \mathcal{A})$ wird **Messraum** genannt.

Nun zum Problem (b).

Gegeben sei dazu ein Messraum $(\Omega, \mathcal{A})$. Unter welchen Bedingungen wollen wir eine auf $\mathcal{A}$ definierte Zuordnungsvorschrift als „Inhaltsmessung“ interpretieren? Offensichtlich wird der „Inhalt“ eines $A \in \mathcal{A}$ eine nicht negative reelle Zahl sein müssen, wobei wir – motiviert am Beispiel des $\mathbb{R}^2$ – auch den Wert $+\infty$ zulassen werden. Es geht also um Abbildungen

$$\mu : \mathcal{A} \longrightarrow [0, +\infty].$$

In der folgenden Definition wird präzisiert, wann so ein μ als „Maß“- d.h. als verallgemeinerte Inhaltsmessung angesehen werden soll.

Definition 2.1.2 Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Abbildung $\mu : \mathcal{A} \longrightarrow [0, +\infty]$ heißt Ein **Maß** auf $(\Omega, \mathcal{A})$, wenn gilt:

- (a) $\mu(\emptyset) = 0$
- (b) Ist $A_1, A_2, \dots$ eine Folge in $\mathcal{A}$ mit $A_i \cap A_k = \emptyset$ für $i \neq k$ (man spricht dann von einer disjunkten Folge), so gilt

$$\mu \left(\bigcup_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n).$$

Dabei ist die rechte Seite als Reihe in $[0, +\infty]$ aufzufassen. μ heißt σ -additiv.

Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt **Maßraum**.

Die Forderungen sind mehr oder weniger plausibel (endliche Additivität statt σ -Additivität wäre zwar naheliegender gewesen, führt aber zu einer wesentlich ärmeren Theorie).

Das Banach-Tarski-Paradoxon

Wir wenden uns kurz der nahe liegenden Frage zu, ob man nicht jeder Teilmenge $A \subset \Omega$ einen „Inhalt“ zuordnen kann. Im $\mathbb{R}^3$ ist das nicht möglich, wenn man zusätzlich voraussetzt, dass der Inhalt von Quadern wie üblich berechnet wird (Grundfläche $\cdot$ Höhe).

Satz 2.1.3 (Banach, 1923) *Sind A und B beschränkte Teilmengen des $\mathbb{R}^3$ mit nicht leerem Inneren, so gibt es ein $m \in \mathbb{N}$ und disjunkte Zerlegungen*

$$A = \bigcup_{i=1}^m A_i, \quad B = \bigcup_{i=1}^m B_i$$

so, dass für $i = 1, \dots, m$ die Menge A_i zu B_i kongruent ist¹.

Der Satz von Banach und Tarski besagt, dass es z.B. möglich ist, eine Kugel vom Radius 1 im $\mathbb{R}^3$ derart in endlich viele disjunkte Teilmengen zu zerlegen, dass diese Teilstücke durch geeignete Bewegungen des $\mathbb{R}^3$ wieder zu zwei disjunkten Kugeln vom Radius 1 zusammengesetzt werden können. Die Mengen A_i bzw. B_i im allgemeinen sehr kompliziert, sie werden mit Hilfe des Auswahlaxioms konstruiert. Für A_i, B_i ist der Begriff des Volumens nicht sinnvoll.

Daraus schließt man leicht, dass es kein Maß $\mu : \mathcal{P}(\mathbb{R}^3) \rightarrow [0, \infty]$ geben kann mit folgenden Eigenschaften:

(a)

$$\mu \left(\prod_{i=1}^3 [a_i, b_i] \right) = \prod_{i=1}^3 (b_i - a_i), \quad a_i < b_i, \quad i = 1, 2, 3.$$

(b) Sind A und B kongruent, so ist

$$\mu(A) = \mu(B).$$

Also ist die Potenzmenge $\mathcal{P}(\mathbb{R}^3)$ zur „vernünftigen Inhaltsmessung“ zu reichhaltig, man muss mit weniger Mengen auskommen.

¹ $A, B \subset \mathbb{R}^3$ sind kongruent, wenn es eine Abbildung $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ gibt, die Produkt von Drehungen und Verschiebungen ist, mit $T(A) = B$

2.1.2 Integrationstheorie: Das Programm

Der Maßbegriff ist grundlegend für die gesamte Wahrscheinlichkeitstheorie. Dort geht man aus von

- einer Menge Ω , die als Menge aller möglichen Ergebnisse eines Zufallsexperiments („Menge der Elementarereignisse“) aufzufassen ist
- einer σ -Algebra $\mathcal{A}$ auf Ω (die $A \in \mathcal{A}$ heißen „Ereignisse“)
- einem Maß $\mathbb{P}$ auf $(\Omega, \mathcal{A})$ mit $\mathbb{P}(\Omega) = 1$ (ein sogenanntes **Wahrscheinlichkeitsmaß**)

Für $A \in \mathcal{A}$ wird dann $\mathbb{P}(A)$ als „die Wahrscheinlichkeit für das Eintreffen von A “ interpretiert.

Gegeben sei nun ein Maßraum $(\Omega, \mathcal{A}, \mu)$ und dazu eine Funktion $f : \Omega \rightarrow [0, \infty)$. (Wir beschränken uns zunächst auf nicht negative f .)

Wir fassen f als Gewichtungsfunktion auf und stellen uns das Problem, Ω bezüglich dieser Gewichtung zu „messen“; hat also f auf gewissen Anteilen von Ω den Wert 0 (oder 1 oder 2), so soll das keinen Beitrag (bzw. Maß dieses Anteils bzw. das Doppelte dieses Maßes) liefern. Das ist noch recht vage, beschreibt aber in verschiedensten Bereichen auftretende Problemstellungen:

- Ist Ω ein Intervall, versehen mit dem Lebesguemaß, so führt die Bestimmung der „Fläche unter einer Kurve“ genau auf dieses Problem.
- Ist Ω wieder ein Intervall, diesmal aufgefasst als Modell eines Stabes inhomogener Masse, f die die Massendichte beschreibende Funktion, so liefert unser Ansatz gerade den Wert der Gesamtmasse.
- Variiert man die vorherige Situation, indem man f als Ladungsdichte vorgibt, so werden wir auf den Begriff der Gesamtladung geführt.
- Ist $(\Omega, \mathcal{A}, \mathbb{P})$ das Modell eines Wahrscheinlichkeitsraumes (also $\mathbb{P}(\Omega) = 1$) und zusätzlich $\Omega \subset \mathbb{R}$, so ist es plausibel, die durch $f(x) = x$ gewichtete „Messung von Ω bezüglich f “ als Erwartungswert zu interpretieren

Wir werden das Problem der „durch f gewichteten Ω -Messung“ mit der Integrationstheorie lösen. Vorher wollen wir kurz zusammenstellen, was ein sinnvoller Ansatz für Eigenschaften haben müsste. Bezeichnen wir mit $I(f)$ (I wie Integral) das „Ergebnis der durch f gewichteten Messung von Ω “ dann sind folgende Probleme zu klären:

- (a) Für welche f ist $I(f)$ überhaupt definiert, welche Eigenschaften hat die Gesamtheit dieser Funktionen?

(b) Welche Eigenschaften hat die Abbildung $f \mapsto I(f)$?

Zu (a)

Sei $\mathcal{F} \subset \{f : \Omega \rightarrow \mathbb{R}\}$ die Gesamtheit der Funktionen, für die $I(f)$ erklärt ist. Wünschenswert ist es, dass $\mathcal{F}$ „möglichst viele“ Funktionen enthält und, dass man mittels der für das Rechnen mit Funktionen wichtigen Operationen (z.B. Addition, Grenzübergänge) wieder Elemente von $\mathcal{F}$ gewinnt.

Eine mögliche Ausfüllung dieses „Programms“ wäre:

- Ist $A \in \mathcal{A}$ mit $\mu(A) < \infty$, so gehört die **Indikatorfunktion**

$$1_A : \Omega \rightarrow \mathbb{R},$$

$$1_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

zu $\mathcal{F}$ (das sind die naheliegendsten Funktionen, die sich bei Vorgabe von $(\Omega, \mathcal{A}, \mu)$ definieren lassen).

- $\mathcal{F}$ ist abgeschlossen unter sinnvollen algebraischen Operationen, also etwa

$$- f, g \in \mathcal{F}, \alpha, \beta \in \mathbb{R} \implies \alpha f + \beta g \in \mathcal{F}$$

$$- f, g \in \mathcal{F} \implies f \cdot g \in \mathcal{F}$$

und der Grenzwertbildung

$$- \text{Ist } (f_n)_{n \in \mathbb{N}} \subset \mathcal{F} \text{ mit } f_n \rightarrow f \text{ punktweise auf } \Omega \text{ so gilt } f \in \mathcal{F}.$$

Die Integrationstheorie wird nicht allen diesen Forderungen genügen. Das ist auch nicht überraschend, denn schon bei uneigentlichen Riemann-Integralen hat man genügend Beispiele dafür, dass für stetige Funktionen Produkte bzw. Limiten integrierbarer Funktionen nicht notwendig integrierbar sind.

Nun zu (b)

Was ist von $f \mapsto I(f)$ sinnvollerweise zu fordern, um die Definition von I als angemessen in Hinblick auf die Verwirklichung unseres Ziels ansehen zu können?

Sicher muss I das Folgende dazu leisten:

Verträglichkeit Ist f so einfach, dass „die durch f gewichtete Ω -Messung“ einen offensichtlichen Wert hat, so sollte $I(f)$ genau diesen Wert liefern, also:

$$\forall A \in \mathcal{A}, \mu(A) < +\infty : I(1_A) = \mu(A) \quad (2.1)$$

„**Linearität**“ Wegen (2.1) und der Additivität von μ muss offensichtlich

$$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$$

gelten, falls alle diese Ausdrücke definiert sind.

Monotonie Gilt punktweise $f \leq g$ (ist die „Gewichtung“ durch f also stets kleiner gleich derjenigen durch g), so sollte auch $I(f) \leq I(g)$ gelten. Da aufgrund der Linearität $I(0) = 0$ gilt, heißt das insbesondere, dass

$$f \geq 0 \implies I(f) \geq 0$$

zu fordern ist.

Zusammenfassend können wir also sagen: Gesucht werden eine Menge $\mathcal{F}$ von Abbildungen $f : \Omega \longrightarrow \mathbb{R}$ und eine Abbildung $I : \mathcal{F} \longrightarrow \mathbb{R}$ so, dass $\mathcal{F}$ die Menge $\{1_A | A \in \mathcal{A}, \mu(A) < \infty\}$ umfasst, weitgehend abgeschlossen ist unter algebraischen Operationen und Grenzübergängen und, dass I im wesentlichen „linear“ und monoton ist sowie der Verträglichkeitsbedingung

$$I(1_A) = \mu(A)$$

genügt.

2.2 Mengensysteme

Bemerkung 2.2.1 • *Aus den Eigenschaften einer σ -Algebra folgt unmittelbar: Mit $A_1, A_2, \dots \in \mathcal{A}$ ist auch $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$.*

- *Allgemein gilt (etwas ungenau): In einer σ -Algebra sind alle mengentheoretischen Operationen möglich, sofern nur höchstens abzählbar viele Elemente der σ -Algebra beteiligt sind und höchstens abzählbar viele Operationen ausgeführt werden.*

Nun stellen wir ein wichtiges Konstruktionsprinzip für σ -Algebren vor.

Lemma 2.2.2 *Sei $\mathcal{E} \subset \mathcal{P}(\Omega)$. Dann gibt es eine kleinste σ -Algebra $\sigma(\mathcal{E})$, die $\mathcal{E}$ umfasst. $\sigma(\mathcal{E})$ ist die von $\mathcal{E}$ erzeugte σ -Algebra. $\mathcal{E}$ heißt **Erzeuger**.*

Beweis Es gilt

$$\sigma(\mathcal{E}) = \bigcap \{ \mathcal{A} | \mathcal{A} \text{ ist } \sigma\text{-Algebra, } \mathcal{E} \subset \mathcal{A} \}$$

Da $\mathcal{P}(\Omega)$ eine σ -Algebra ist, ist die Menge auf der rechten Seite nicht leer. Man beweist leicht, dass die rechte Seite eine σ -Algebra ist. $\square$

Es ist im Allgemeinen hoffnungslos, eine explizite Beschreibung der Elemente von $\sigma(\mathcal{E})$ zu geben. Die für die Anwendungen wichtigen σ -Algebren entstehen aber überwiegend auf diese Weise. Man hat ein System $\mathcal{E}$ von Mengen, das für den Messprozess zur Verfügung stehen soll (etwa: alle Rechtecke und Kreise im $\mathbb{R}^2$) und betrachtet dann $\sigma(\mathcal{E})$, um die für σ -Algebren entwickelte Theorie anwenden zu können.

Beispiel 2.2.3 (a) $\mathcal{P}(\Omega)$ und $\{\emptyset, \Omega\}$ sind stets σ -Algebren auf Ω (offensichtlich handelt es sich gerade um die größte und die kleinste mögliche σ -Algebra).

(b) Ist $A \subset \Omega$, $A \neq \emptyset$, dann gilt $\sigma(\{A\}) = \{\emptyset, A, \Omega \setminus A, \Omega\}$.

(c) Als etwas weniger triviales Beispiel betrachte man

$$\mathcal{A} := \{A \mid A \subset \Omega, A \text{ oder } \Omega \setminus A \text{ ist höchstens abzählbar}\}$$

(d) Sei (T, τ) ein topologischer Raum. $\sigma(\tau)$ (also die von den offenen Teilmengen erzeugte σ -Algebra) heißt die σ -Algebra der **Borelmengen** auf T . Die Borelmengen des $\mathbb{R}^n$ bezeichnen wir mit $\mathcal{B}^n$.

Es sollte damit nun leicht sein, die folgenden Tatsachen über die Borelmengen in $\mathbb{R}$ herzuleiten:

- $\mathbb{Q}$ ist eine Borelmenge [$\mathbb{Q} = \cup\{\{q\} \mid q \in \mathbb{Q}\}$ und $\{q\}$ ist Komplement einer offenen Menge]
- Ist $A \subset \mathbb{R}$ eine Borelmenge, so auch $x + A$ für jedes $x \in \mathbb{R}$. [s. Beispiel 2.5.5, S. 12]

$\mathcal{B}^n$ kann durch unterschiedliche Mengensysteme erzeugt werden. Man kann zeigen:

$$\mathcal{B}^n = \sigma\left(\left\{\prod_{i=1}^n (-\infty, r_i] \mid r_1, \dots, r_n \in \mathbb{R}\right\}\right) \quad (2.2)$$

$$\mathcal{B}^n = \sigma\left(\left\{\prod_{i=1}^n [a_i, b_i] \mid a_i < b_i, a_i, b_i \in \mathbb{R}, i = 1, \dots, n\right\}\right) \quad (2.3)$$

2.3 Eigenschaften von Maßen

Satz 2.3.1 Sei $\Omega, \mathcal{A}, \mu$ ein Maßraum. Dann gilt:

- (a) $A, B \in \mathcal{A}, A \subset B \implies \mu(A) \leq \mu(B)$. Gilt zusätzlich $\mu(A) < \infty$ dann folgt $\mu(B \setminus A) = \mu(B) - \mu(A)$.
- (b) Sei $A \in \mathcal{A}$ und für $n \in \mathbb{N}$ $A_n \in \mathcal{A}$ mit $A_n \uparrow A$. Dann gilt $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$.
- (c) Sei $A \in \mathcal{A}$ und für $n \in \mathbb{N}$ $A_n \in \mathcal{A}$ mit $A_n \downarrow A$, $\mu(A_1) < \infty$. Dann folgt $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$.

Die Konvergenz in (b) bzw. (c) heißt **Stetigkeit von unten** bzw. **oben**.

Beweis (a) Wegen der Additivität von μ gilt $\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A)$ und es folgen beide Behauptungen.

(b) Sei $B_1 := A_1$, $B_{n+1} := A_{n+1} \setminus A_n$ für $n \in \mathbb{N}$. Dann sind die B_n paarweise disjunkt und $\bigcup_{k=1}^n B_k = A_n$. Somit erhalten wir wegen der σ -Additivität von μ

$$\mu(A) = \mu\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \sum_{n \in \mathbb{N}} \mu(B_n) = \lim_{n \rightarrow \infty} \underbrace{\sum_{k=1}^n \mu(B_k)}_{=\mu(A_n)} = \lim_{n \rightarrow \infty} \mu(A_n).$$

(c) Die Folge $(\mu(A_n))_{n \in \mathbb{N}}$ ist monoton fallend und nach unten durch 0 beschränkt, also konvergent. Es gilt $A_1 \setminus A_n \uparrow A_1 \setminus A$ und somit wegen (b)

$$\mu(A_1 \setminus A) = \lim_{n \rightarrow \infty} \mu(A_1 \setminus A_n) = \lim_{n \rightarrow \infty} (\mu(A_1) - \mu(A_n)) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n).$$

Somit erhalten wir mit (a) $\mu(A_1) - \mu(A) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n)$. $\square$

2.4 Konstruktion von Maßen

2.4.1 Existenz und Eindeutigkeit von Fortsetzungen

Aus einem Mengensystem $\mathcal{E}$ kann die σ -Algebra $\sigma(\mathcal{E})$ erzeugt werden. Wir gehen der Frage nach unter welchen Voraussetzungen eine Mengenfunktion $\mu : \mathcal{E} \rightarrow [0, \infty)$ zu einem Maß auf $\sigma(\mathcal{E})$ fortgesetzt werden kann und ob diese Fortsetzung eindeutig bestimmt ist.

Satz 2.4.1 Es sei $\mathcal{E} \subset \mathcal{P}(\Omega)$ und $\mu : \mathcal{E} \rightarrow [0, \infty]$, σ -additiv auf $\mathcal{E}$ und $\mu(\emptyset) = 0$, falls $\emptyset \in \mathcal{E}$.

(a) $\mathcal{E}$ besitze folgende Eigenschaften²:

- $\emptyset \in \mathcal{E}$
- $A, B \in \mathcal{E} \implies A \cap B \in \mathcal{E}$ [$\mathcal{E}$ heißt in diesem Fall $\cap$ -stabil]
- Für $A, B \in \mathcal{E}, B \subset A$ gibt es paarweise disjunkte $C_1, \dots, C_m \in \mathcal{E}$ mit $A \setminus B = \bigcup_{i=1}^m C_i$.

Dann gibt es ein Maß $\tilde{\mu}$ auf $\sigma(\mathcal{E})$ mit $\tilde{\mu}|_{\mathcal{E}} = \mu$.

(b) Zusätzlich zu (a) gelte ³:

$$\exists (E_k)_{k \in \mathbb{N}} \subset \mathcal{E} : \quad \mu(E_k) < \infty, \quad E_n \uparrow \Omega, \quad (2.4)$$

Dann ist die Fortsetzung $\tilde{\mu}$ eindeutig.

Beweis [El], S.61, 5.7 Korollar.

Korollar 2.4.2 Es seien μ, ν Maße auf der $(\Omega, \mathcal{A})$ und $\mathcal{E}$ ein $\cap$ -stabiler Erzeuger von $\mathcal{A}$ mit folgende Eigenschaften:

(a) $\mu|_{\mathcal{E}} = \nu|_{\mathcal{E}}$

(b) $\mathcal{E}$ erfüllt (2.4).

Dann ist $\mu = \nu$.

Beweis [El], S.60, 5.6 Eindeutigkeitssatz. □

2.4.2 Das Lebesgue-Borel Maß

Die Menge aller Quader $\mathcal{Q}^n$ des $\mathbb{R}^n$

$$\mathcal{Q}^n := \left\{ \prod_{i=1}^n]a_i, b_i] \mid a_i \leq b_i, i = 1, \dots, n, a_i, b_i \in \mathbb{R} \right\}$$

erfüllt die Bedingungen in Satz 2.4.1 (a) [$\emptyset = (a, a]^n \in \mathcal{Q}^n$, der Durchschnitt zweier Quader ist ein Quader und das Komplement eines Quaders $B \subset A$ in A ist Vereinigung von höchstens 2^n disjunkten Quadern]. Die Mengenfunktion $\lambda^n : \mathcal{Q}^n \longrightarrow [0, \infty]$,

$$\lambda^n \left(\prod_{i=1}^n]a_i, b_i] \right) = \prod_{i=1}^n (b_i - a_i). \quad (2.5)$$

² $\mathcal{E}$ heißt in diesem Fall ein Semiring

³ μ heißt dann σ -endlich

ist σ -additiv (s. [B-WT] S. 21, Satz 4.4 bzw. [El], S. 38, 2.2 Satz) erfüllt die Bedingung (2.4) von Satz 2.4.1 (b) [$E_k := (-k, k]^n, \lambda(E_k) = (2k)^n < \infty$]. Somit gibt es laut Satz 2.4.1 genau eine Fortsetzung von λ^n auf $\mathcal{B}^n = \sigma(\mathcal{Q}^n)$. Dieses Maß bezeichnen wir wieder mit λ^n .

Definition 2.4.3 Das Maß λ^n auf $\mathcal{B}^n$ mit der Eigenschaft (2.5) heißt **Lebesgue-Borel Maß (L-B-Maß)**.

Wenn die Dimension des Maßraums klar ist bezeichnen wir das L-B Maß auch mit nur λ

2.5 Messbare Funktionen

Definition 2.5.1 Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ Messräume. $f : \Omega \longrightarrow \Omega'$ heißt $\mathcal{A} - \mathcal{A}'$ **messbar**, wenn gilt:

$$\forall A_2 \in \mathcal{A}' : f^{-1}(A_2) \in \mathcal{A}. \quad (2.6)$$

In den meisten Fällen ist klar auf welche σ -Algebren sich die Messbarkeit bezieht. Aus diesem Grund werden wir oft statt $\mathcal{A} - \mathcal{A}'$ messbar nur messbar sagen, insbesondere wenn $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}^1)$.

Meist ist es schwierig (2.6) nachzuweisen. Kennt man einen Erzeuger von $\mathcal{A}$, dann reicht es (2.6) für den Erzeuger zu überprüfen.

Lemma 2.5.2 Sei $\mathcal{E}' \subset \mathcal{P}(\Omega')$ mit $\sigma(\mathcal{E}') = \mathcal{A}'$. Dann gilt:

$$f : \Omega \longrightarrow \Omega' \text{ messbar} \iff \forall E' \in \mathcal{E}' : f^{-1}(E') \in \mathcal{A}. \quad (2.7)$$

Beweis „ $\implies$ “ gilt laut der Definition der Messbarkeit und $\mathcal{E}' \subset \sigma(\mathcal{E}') = \mathcal{A}'$. „ $\impliedby$ “: Das System

$$\mathcal{T} := \{A' \subset \Omega' \mid f^{-1}(A') \in \mathcal{A}\}$$

ist eine σ -Algebra [$f^{-1}(\Omega') = \Omega \in \mathcal{A}$, also $\Omega' \in \mathcal{T}$. Ferner gilt $f^{-1}(\Omega' \setminus A') = \Omega \setminus f^{-1}(A')$ und $f^{-1}(\bigcup_{n \in \mathbb{N}} A'_n) = \bigcup_{n \in \mathbb{N}} f^{-1}(A'_n)$]. Somit folgt

$$(2.7) \implies \mathcal{E}' \subset \mathcal{T} \implies \mathcal{A}' = \sigma(\mathcal{E}') \subset \sigma(\mathcal{T}) = \mathcal{T}$$

also $\mathcal{A}' \subset \mathcal{T}$ und somit (2.6). $\square$

Korollar 2.5.3 Sei $f : \Omega \longrightarrow \Omega'$ messbar und $\mathcal{E}'$ ein Erzeuger von $\mathcal{A}'$. Dann ist $f^{-1}(\mathcal{A}')$ eine σ -Algebra und $f^{-1}(\mathcal{E}')$ ein Erzeuger von $f^{-1}(\mathcal{A}')$, also

$$\sigma(f^{-1}(\mathcal{E}')) = f^{-1}(\mathcal{A}').$$

Beweis Die erste Aussage ist klar. Sei

$$\mathcal{T}' := \{A' \subset \Omega' \mid f^{-1}(A') \in \sigma(f^{-1}(\mathcal{E}'))\}.$$

Wie im Beweis von Lemma 2.5.2 sieht man, dass $\mathcal{T}'$ eine σ -Algebra auf Ω' ist. Da f messbar ist, gilt $\mathcal{E}' \subset \mathcal{T}'$, also auch $\mathcal{A}' = \sigma(\mathcal{E}') \subset \mathcal{T}'$. $\square$

Lemma 2.5.2 wenden wir auf Funktionen $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n)$ an:

Lemma 2.5.4 $f : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n)$ ist messbar, wenn gilt

$$\forall r_1, \dots, r_n \in \mathbb{R} : f^{-1} \left(\prod_{i=1}^n (-\infty, r_i] \right) \in \mathcal{A}.$$

Beweis Die Behauptung folgt mit (2.2) und Lemma 2.5.2. $\square$

Beispiel 2.5.5 (a) Für $A \in \mathcal{A}$ ist $f : \Omega \longrightarrow \mathbb{R}$, $f(x) = 1_A(x)$ $\mathcal{A} - \mathcal{B}^1$ messbar.

(b) Jede stetige Funktion $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ ist $\mathcal{B}^n - \mathcal{B}^m$ messbar. Damit folgt insbesondere $x + A \in \mathcal{B}^n$ für alle $A \in \mathcal{B}^n$ und $x \in \mathbb{R}^n$.

Wir wenden uns nun dem Raum der messbaren Funktionen

$$\mathcal{M}(\Omega, \mathcal{A}) := \{f : \Omega \longrightarrow \mathbb{R} \mid f \text{ ist } \mathcal{A} - \mathcal{B}^1 \text{ messbar}\}$$

zu. Ziel ist es zu zeigen, dass $\mathcal{M}(\Omega, \mathcal{A})$ abgeschlossen ist bezüglich algebraischer Operationen und Grenzwertbildung. Dazu vorweg noch ein Ergebnis über die Messbarkeit der Verkettung zweier messbarer Funktionen:

Lemma 2.5.6 (a) Seien $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$, $g : (\Omega', \mathcal{A}') \longrightarrow (\Omega'', \mathcal{A}'')$ messbar. Dann ist $g \circ f$ messbar.

(b) $f = (f_1, \dots, f_n) : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}^n, \mathcal{B}^n)$ ist messbar $\iff f_i$ ist für alle $i = 1, \dots, n$ messbar.

Beweis (a) Sei $A'' \in \mathcal{A}''$:

$$(g \circ f)^{-1} = f^{-1}(\underbrace{g^{-1}(A'')}_{\in \mathcal{A}'}) \implies f^{-1}(g^{-1}(A'')) \in \mathcal{A}.$$

(b) „ $\implies$ “ folgt aus der Stetigkeit der Projektionen $\pi_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $x \longmapsto x_i$ aus (a) und Beispiel 2.5.5, (b).

„ $\Leftarrow$ “: Sei $r_i \in \mathbb{R}^n$, $i = 1, \dots, n$. Dann gilt

$$f^{-1} \left(\prod_{i=1}^n (-\infty, r_i] \right) = \bigcap_{i=1}^n f^{-1} ((-\infty, r_i]). \quad (2.8)$$

Da f_i messbar ist, folgt $f^{-1}((-\infty, r_i]) \in \mathcal{A}$ für alle i und somit ist die rechte Seite in $\mathcal{A}$. Mit Lemma 2.5.4 folgt die Behauptung. $\square$

Es folgen nun die Abgeschlossenheitseigenschaften von $\mathcal{M}(\Omega, \mathcal{A})$.

Satz 2.5.7 (a) Sei $f_i \in \mathcal{M}(\Omega, \mathcal{A})$, $i = 1, \dots, n$. Dann gilt

$$\forall \alpha_1, \dots, \alpha_n \in \mathbb{R} : \sum_{i=1}^n \alpha_i f_i \in \mathcal{M}(\Omega, \mathcal{A}) \quad (2.9)$$

$$\prod_{i=1}^n f_i \in \mathcal{M}(\Omega, \mathcal{A}). \quad (2.10)$$

(b) Sei $f_i \in \mathcal{M}(\Omega, \mathcal{A})$, $i \in \mathbb{N}$. Dann gilt

(i) Existiert $\sup_i f_i$ (bzw. $\inf_i f_i$) punktweise, dann gilt $\sup_i f_i \in \mathcal{M}(\Omega, \mathcal{A})$ (bzw. $\inf_i f_i \in \mathcal{M}(\Omega, \mathcal{A})$).

(ii) Existiert $\lim_{i \rightarrow \infty} f_i$ punktweise, dann gilt $\lim_{i \rightarrow \infty} f_i \in \mathcal{M}(\Omega, \mathcal{A})$.

Beweis (a) Die Funktion $f = (f_1, \dots, f_n) : \Omega \rightarrow \mathbb{R}^n$ ist nach Lemma 2.5.6

(b) messbar. Die Funktionen $g, h : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$g(x) := \sum_{i=1}^n \alpha_i x_i, \quad h(x) = \prod_{i=1}^n x_i$$

sind stetig, also messbar. Dann folgt mit Lemma 2.5.6, (a), dass $g \circ f$ und $h \circ f$ messbar sind, also die Behauptung.

(b), (i): Für alle $r \in \mathbb{R}$ gilt

$$\{x \in \Omega \mid \sup_i f_i(x) \leq r\} = \bigcap_{i \in \mathbb{N}} \underbrace{f^{-1}((-\infty, r])}_{\in \mathcal{A}} \in \mathcal{A}.$$

Damit ist nach Lemma 2.5.4 $\sup_i f_i$ messbar. Wegen $\inf_i f_i = -\sup_i (-f_i)$ ist laut (a) auch $\inf_i f_i$ messbar.

(b), (ii) folgt aus (i) wegen

$$\lim_{i \rightarrow \infty} f_i = \sup_k \left(\underbrace{\inf_{i \geq k} f_i}_{\in \mathcal{M}(\Omega, \mathcal{A})} \right).$$

$\square$

Zum Abschluss des Abschnitts folgt nun eine weitere Charakterisierung messbarer Funktionen.

Definition 2.5.8 Eine messbare Abbildung $f : \Omega \longrightarrow \mathbb{R}$ heißt **Treppenfunktion**, wenn sie nur endlich viele Werte annimmt.

Ist f eine Treppenfunktion dann gibt es $\alpha_i \in \mathbb{R}$, $A_i \in \mathcal{A}$, $i = 1, \dots, n$ mit

$$\forall \omega \in \Omega : f(\omega) = \sum_{i=1}^n \alpha_i 1_{A_i}(\omega)$$

Satz 2.5.9 Äquivalent sind:

(a) $f \in \mathcal{M}(\Omega, \mathcal{A})$

(b) $\exists (f_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega, \mathcal{A}) : f_n \text{ Treppenfunktion und } \lim_{n \rightarrow \infty} f_n = f.$

Für $f \geq 0$ kann $(f_n)_{n \in \mathbb{N}}$ monoton wachsend und $f_n \geq 0, n \in \mathbb{N}$ gewählt werden.

Beweis „(b) $\implies$ (a)“: Klar wegen Satz 2.5.7 (b).

„(a) $\implies$ (b)“: Sei zunächst $f \geq 0$. Für $n \in \mathbb{N}$ definiere

$$A_{k,n} := f^{-1} \left([(k-1)2^{-n}, k2^{-n}) \right), k = 1, \dots, n2^n$$

und

$$f_n(x) := \begin{cases} n & \text{falls } x \in f^{-1}([n, \infty)) \\ \sum_{k=1}^{n2^n} (k-1)2^{-n} 1_{A_{k,n}}(x) & \text{sonst} \end{cases}$$

(Idee: f_n entsteht aus f indem der Wertebereich $[0, n)$ in Intervalle der Länge 2^{-n} zerlegt wird und f durch die jeweils untere Intervallgrenze approximiert wird).

f_n ist Treppenfunktion und es gilt $f_n \uparrow f$: Sei $x \in \Omega$. Dann gibt es ein $n_0 \in \mathbb{N}$ mit

$$\forall n \geq n_0 : f(x) \leq n$$

Laut Definition gilt auch

$$\forall n \geq n_0 : |f_n(x) - f(x)| \leq 2^{-n}$$

also $f_n(x) \rightarrow f(x)$. Wir zeigen nun, dass $f_n \leq f_{n+1}$ gilt. Sei $x \in \Omega$.

1. Fall $f(x) < n$. Dann gibt es $k \in \{1, \dots, n2^n\}$ mit

$$f(x) \in [(k-1)2^{-n}, k2^{-n})$$

also laut Definition von f_n

$$f_n(x) = (k-1)2^{-n}.$$

Weiter gilt

$$f(x) \in \left[(k-1)2^{-n}, \frac{1}{2}((k-1)2^{-n} + k2^{-n}) \right) = \left[(k-1)2^{-n}, (2k-1)2^{-n-1} \right)$$

oder

$$f(x) \in \left[(2k-1)2^{-n-1}, k2^{-n} \right)$$

Laut Definition von f_{n+1} folgt

$$\begin{aligned} f_{n+1}(x) &= 2(k-1)2^{-n-1} = f_n(x) \text{ oder} \\ f_{n+1}(x) &= (2k-1)2^{-n-1} > 2(k-1)2^{-n-1} = f_n(x). \end{aligned}$$

2. Fall $n \leq f(x) < n+1$. Dann ist $f_n(x) = n$ und $f_{n+1}(x) = (k-1)2^{-n-1}$ mit

$$f(x) \in \left[(k-1)2^{-n-1}, k2^{-n-1} \right).$$

Wegen der Voraussetzung ist dann $(k-1)2^{-n-1} \geq n$, also $f_{n+1}(x) = (k-1)2^{-n-1} \geq n = f_n(x)$.

3. Fall $n+1 \leq f(x)$ Dann ist $f_n(x) = n$ und $f_{n+1}(x) = n+1$.

Für allgemeines f gilt

$$f = f_+ - f_-$$

mit $f_-(x) := \max(0, -f(x))$, $f_+(x) := \max(0, f(x))$. Es gilt wegen Satz 2.5.7 $f_+, f_- \in \mathcal{M}(\Omega, \mathcal{A})$ und $f_+, f_- \geq 0$. Man kann nun den ersten Teil des Beweises auf f_+ und f_- anwenden und erhält Folgen (f_n) und (g_n) von Treppenfunktionen mit

$$f_n \longrightarrow f_+, \quad g_n \longrightarrow f_-$$

also gilt $f_n - g_n \rightarrow f_+ - f_- = f$. □

2.6 Integration

2.6.1 Konstruktion

Wir stellen die Konstruktion des Integrals bezüglich Maßen dar. Wir gehen von einem Maßraum $(\Omega, \mathcal{A}, \mu)$ aus.

Schritt 1: $f \in \mathcal{M}(\Omega, \mathcal{A})$, $f \geq 0$, **Treppenfunktion**

Laut Voraussetzung gibt es $\alpha_i \geq 0$, $A_i \in \mathcal{A}$, $i = 1, \dots, n$ mit

$$f(\omega) = \sum_{i=1}^n \alpha_i 1_{A_i}(\omega). \tag{2.11}$$

Dann sei

$$\int_{\Omega} f d\mu := \int_{\Omega} f(\omega) d\mu(\omega) := \sum_{i=1}^n \alpha_i \mu(A_i).$$

Man kann zeigen, dass diese Zahl nicht von der speziellen Darstellung von f abhängt.

Schritt 2: $f \in \mathcal{M}(\Omega, \mathcal{A})$, $f \geq 0$

Laut Satz 2.5.9 gibt es eine Folge $(f_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega, \mathcal{A})$ von Treppenfunktionen mit $f_n \uparrow f$, $f_n \geq 0$, $n \in \mathbb{N}$. Dann sei

$$\int_{\Omega} f d\mu := \int_{\Omega} f(\omega) d\mu(\omega) := \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu(\omega).$$

Man kann zeigen, dass der Grenzwert auf der rechten Seite existiert und, dass die so definierte Zahl nicht von der Auswahl der Folge $(f_n)_{n \in \mathbb{N}}$ abhängt. Ferner gilt für $f, g \in \mathcal{M}(\Omega, \mathcal{A})$, $f, g \geq 0$

$$\int_{\Omega} (f + g) d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu$$

Schritt 3: $f \in \mathcal{M}(\Omega, \mathcal{A})$

Für f gilt

$$f = f_+ - f_-$$

mit $f_-(\omega) := -\min(0, f(\omega))$, $f_+(\omega) := \max(0, f(\omega))$. Es gilt wegen Satz 2.5.7 $f_+, f_- \in \mathcal{M}(\Omega, \mathcal{A})$. Außerdem ist $|f| \geq 0$ und $|f| \in \mathcal{M}(\Omega, \mathcal{A})$ wegen

$$|f| = f_+ + f_-.$$

Somit ist folgende Definition sinnvoll.

Definition 2.6.1 Sei $f \in \mathcal{M}(\Omega, \mathcal{A})$.

(a) f heißt **μ -integrierbar**, wenn

$$\int_{\Omega} |f(\omega)| d\mu(\omega) < \infty$$

gilt.

(b) f heißt **über $A \in \mathcal{A}$ integrierbar**, wenn $f \cdot 1_A$ integrierbar ist. Wir setzen

$$\int_A f d\mu := \int_A f(\omega) d\mu(\omega) := \int f 1_A d\mu.$$

Ist f μ -integrierbar, dann sei

$$\int_{\Omega} f d\mu := \int_{\Omega} f(\omega) d\mu(\omega) := \int_{\Omega} f_+(\omega) d\mu(\omega) - \int_{\Omega} f_-(\omega) d\mu(\omega).$$

Beispiel 2.6.2 Sei $\omega_0 \in \Omega$ und $\delta_{\omega_0} : \mathcal{A} \longrightarrow [0, 1]$

$$\delta_{\omega_0}(A) := \begin{cases} 1 & \omega_0 \in A \\ 0 & \omega_0 \notin A. \end{cases}$$

δ_{ω_0} heißt **Dirac-Maß** in Ω_0 . Alle $f \in \mathcal{M}(\Omega, \mathcal{A})$ sind δ_{ω_0} -integrierbar und es gilt

$$\int_{\Omega} f d\delta_{\omega_0} = f(\omega_0).$$

Beweis Es werden die Schritte 1-3 durchgeführt. Ist $f \geq 0$ Treppenfunktion wie in (2.11), dann gilt

$$\int_{\Omega} f(\omega) d\delta_{\omega_0} = \sum_{i=1}^n \alpha_i \underbrace{\delta_{\omega_0}(A_i)}_{=1_{A_i}(\omega_0)} = \sum_{i=1}^n \alpha_i 1_{A_i}(\omega_0) = f(\omega_0).$$

Ist nun $f \geq 0$ dann sei $(f_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega, \mathcal{A})$ eine Folge von Treppenfunktionen mit $f_n \uparrow f$, $f_n \geq 0$, $n \in \mathbb{N}$. Dann gilt:

$$\int_{\Omega} f(\omega) d\delta_{\omega_0}(\omega) := \lim_{n \rightarrow \infty} \underbrace{\int_{\Omega} f_n d\delta_{\omega_0}(\omega)}_{=f_n(\omega_0)} = \lim_{n \rightarrow \infty} f_n(\omega_0) = f(\omega_0).$$

Schließlich gilt

$$\int_{\Omega} f_+(\omega) d\delta_{\omega_0}(\omega) = f_+(\omega_0) \text{ und } \int_{\Omega} (f_-(\omega)) d\delta_{\omega_0}(\omega) = f_-(\omega_0)$$

also

$$\begin{aligned} \int_{\Omega} f(\omega) d\delta_{\omega_0}(\omega) &= \int_{\Omega} f_+(\omega) d\delta_{\omega_0}(\omega) - \int_{\Omega} f_-(\omega) d\delta_{\omega_0}(\omega) = f_+(\omega_0) - f_-(\omega_0) \\ &= f(\omega_0). \end{aligned}$$

□

Das so definierte Integral besitzt folgende Eigenschaften:

Satz 2.6.3 f, g seien μ -integrierbar. Es gilt:

$$\int_{\Omega} \alpha f + \beta g d\mu = \int_{\Omega} \alpha f d\mu + \beta \int_{\Omega} g d\mu \quad (2.12)$$

$$f \geq g \implies \int_{\Omega} f d\mu \geq \int_{\Omega} g d\mu \quad (2.13)$$

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu \quad (2.14)$$

Obiger Satz gilt auch für die Integration über Teilmengen $A \subset \Omega$. (Beweis: Übung)

Korollar 2.6.4 Sei $f \in \mathcal{M}(\Omega, \mathcal{A})$ integrierbar. Dann gilt für alle $A, B \in \mathcal{A}$, A, B disjunkt

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu. \quad (2.15)$$

Beweis Es gilt $f1_{A \cup B} = f1_A + f1_B$ und die Behauptung folgt aus (2.12) $\square$

2.6.2 Nullmengen

Definition 2.6.5 Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. $N \in \mathcal{A}$ heißt μ -**Nullmenge**, wenn $\mu(N) = 0$ gilt. Die λ^n Nullmengen von $(\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$ heißen **Lebesgue Nullmengen**.

Ist das Maß μ aus dem Zusammenhang klar, werden wir auch kurz Nullmenge statt μ -Nullmenge verwenden.

Beispiel 2.6.6 (a) Für $x \in \mathbb{R}$ ist $\{x\}$ eine λ^1 - Nullmenge.

(b) $\mathbb{Q}$ ist $\{x\}$ eine λ^1 - Nullmenge.

Beweis (a) folgt direkt aus (2.5).

Zu (b): $\mathbb{Q}$ ist abzählbar, also

$$\mathbb{Q} = \{q_1, q_2, \dots\} = \bigcup_{n \in \mathbb{N}} \{q_n\}$$

und somit folgt die Behauptung aus der σ -Additivität von Maßen. $\square$

Lemma 2.6.7 (a) Die abzählbare Vereinigung von μ -Nullmengen ist wieder eine Nullmenge.

(b) Messbare Teilmengen von μ -Nullmengen sind μ -Nullmengen.

Beweis (a) Seien $N_1, N_2, \dots$ Nullmengen. Dann gilt

$$\mu \left(\bigcup_{k \in \mathbb{N}} N_k \right) \leq \sum_{k \in \mathbb{N}} \mu(N_k) = 0.$$

(b) Ist N eine μ -Nullmenge und $M \subset N$ dann folgt aus $0 \leq \mu(M) \leq \mu(N) = 0$ die Behauptung. $\square$

Definition 2.6.8 Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Eine Eigenschaft E gilt **für fast alle** $\omega \in \Omega$ bzw. **fast überall (f.ü.)**, wenn es eine μ -Nullmenge N gibt mit

$$\forall \omega \in \Omega \setminus N : E(\omega) \text{ ist wahr.}$$

In der Wahrscheinlichkeitstheorie wird **fast sicher (f.s.)** an Stelle von f.ü. verwendet.

Beispiel 2.6.9 (a) Konvergenz f.ü.: Sei $f_n : [0, 1] \longrightarrow [0, 1]$, $f_n(x) := x^n$. Dann gilt

$$f_n \longrightarrow 0 \text{ f.ü.}$$

(b) Sei $f : \mathbb{R} \longrightarrow [0, 1]$,

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \text{sonst} \end{cases}.$$

Es gilt $f = 0$ f.ü.

Beweis (a) $(f_n(x))_{n \in \mathbb{N}}$ konvergiert gegen 0 für alle $x \in [0, 1]$. $\{1\}$ ist eine Nullmenge.

(b) $\mathbb{Q}$ ist eine Nullmenge. $\square$

Satz 2.6.10 Sei $f \in \mathcal{M}(\Omega, \mathcal{A})$, $f \geq 0$. Es gilt

$$f = 0 \text{ f.ü.} \iff \int_{\Omega} f d\mu = 0.$$

Beweis „ $\implies$ “: Laut Voraussetzung gibt es eine Nullmenge $N \subset \Omega$ mit

$$\forall \omega \in \Omega \setminus N : f(\omega) = 0.$$

(i) f Treppenfunktion, d.h. es gibt $A_i \in \mathcal{A}$ und $\alpha_i > 0$ mit

$$f = \sum_{i=1}^n \alpha_i 1_{A_i}.$$

Dann gilt $A_i \subset N$ für alle $i = 1, \dots, n$: $\omega \in A_i \implies f(\omega) > 0 \implies \omega \notin \Omega \setminus N \implies \omega \in N$. Somit folgt

$$\int_{\Omega} f d\mu = \sum_{i=1}^n \alpha_i \underbrace{\mu(A_i)}_{\leq \mu(N)=0} = 0.$$

(ii) Sei $f_n \uparrow f$ eine Folge von Treppenfunktionen, $f_n \geq 0$. Dann gilt

$$0 \leq f_n \leq f$$

also gilt $f_n = 0$ μ -f.ü. für alle $n \in \mathbb{N}$ und folglich laut (i) $\int f_n d\mu = 0$. Damit folgt aus der Konstruktion des Integrals

$$\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \underbrace{\int_{\Omega} f_n d\mu}_{=0} = 0$$

„ $\Leftarrow$ “: Sei $A_n := \left\{ \omega \in \Omega \mid f(\omega) > \frac{1}{n} \right\}$ und $A := \{ \omega \in \Omega \mid f(\omega) > 0 \}$. Dann $A_n \downarrow A$ also $0 \leq \mu(A) \leq \mu(A_n)$. Ferner gilt

$$0 = \int_{\Omega} f d\mu \geq \int_{\Omega} f 1_{A_n} d\mu \geq \int_{\Omega} \frac{1}{n} 1_{A_n} d\mu = \frac{1}{n} \mu(A_n) \geq 0 \implies \mu(A_n) = 0$$

und es folgt $\mu(A) = 0$ wegen der Stetigkeit von oben. $\square$

Integrierbare Funktionen kann man auf einer Nullmenge beliebig abändern, ohne dass ihre Integrierbarkeit verloren geht.

Korollar 2.6.11 $f \in \mathcal{M}(\Omega, \mathcal{A})$ ist auf jeder μ -Nullmenge integrierbar und $\int_N f d\mu = 0$.

Beweis Es gilt $f 1_N = 0$ f.ü. also folgt die Behauptung. $\square$

Korollar 2.6.12 f, g seien integrierbar. Es gilt:

$$(a) \quad f = g \text{ f.ü.} \implies \int_{\Omega} f d\mu = \int_{\Omega} g d\mu.$$

$$(b) \quad f \leq g \text{ f.ü.} \implies \int_{\Omega} f d\mu \leq \int_{\Omega} g d\mu.$$

Beweis (a) $f - g = 0$ f.ü., also $0 = \int_{\Omega} (f - g) d\mu$.

(b) Es sei $f \leq g$ überall auf $\Omega \setminus N$, $N \in \mathcal{A}$ eine Nullmenge. Dann

$$\int_{\Omega} (f - g) d\mu \stackrel{(2.15)}{=} \underbrace{\int_{\Omega \setminus N} (f - g) d\mu}_{\geq 0} + \underbrace{\int_N (f - g) d\mu}_{=0} \geq 0.$$

In beiden Fällen folgt nun die Behauptung aus (2.12). $\square$

2.6.3 Konvergenzsätze

Die Grenzwertbildung kann in folgenden Fällen mit der Integration vertauscht werden:

Satz 2.6.13 (Monotone Konvergenz) Sei $(f_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega, \mathcal{A})$, $f_n \geq 0$ μ -f.ü. für alle $n \in \mathbb{N}$, $f_n \uparrow f$ μ -f.ü.. Dann gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

Satz 2.6.14 (Dominierte Konvergenz) Sei $(f_n)_{n \in \mathbb{N}} \subset \mathcal{M}(\Omega, \mathcal{A})$, μ -f.ü. konvergent gegen f und es gebe eine μ -integrierbare Funktion g mit

$$\forall n \in \mathbb{N} : |f_n| \leq g \quad \mu\text{-f.ü.}$$

Dann ist f μ -integrierbar und es gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

2.6.4 $L^1(\Omega)$ und $L^2(\Omega)$

Besonders wichtig ist die Menge der integrierbaren bzw. quadratintegrierbaren Funktionen

$$\begin{aligned} \mathcal{L}^1(\Omega) &:= \{f : \Omega \longrightarrow \mathbb{R} \mid f \text{ ist integrierbar}\} \\ \mathcal{L}^2(\Omega) &:= \{f : \Omega \longrightarrow \mathbb{R} \mid f^2 \text{ ist integrierbar}\} \end{aligned}$$

Bemerkenswert ist folgende Eigenschaft:

$$f, g \in \mathcal{L}^2(\Omega) \implies fg \in \mathcal{L}^1(\Omega).$$

f, g heißen äquivalent, wenn $f = g$ f.ü. gilt. Die Menge aller Äquivalenzklassen in $\mathcal{L}^1(\Omega)$ bzw. $\mathcal{L}^2(\Omega)$ wird mit $L^1(\Omega)$ bzw. $L^2(\Omega)$ bezeichnet. Ein Element $f \in L^1(\Omega)$ kann als eine integrierbare Funktion angesehen werden, die bis auf eine Nullmenge eindeutig bestimmt ist, analog $f \in L^2(\Omega)$.

Man kann zeigen, dass $L^1(\Omega)$ ein Banachraum ist mit der Norm

$$\|f\|_1 := \int_{\Omega} |f| d\mu$$

und, dass $L^2(\Omega)$ ein Hilbertraum ist mit dem Skalarprodukt

$$\langle f, g \rangle := \int_{\Omega} fg d\mu.$$

Die induzierte Norm bezeichnen wir mit

$$\|f\|_2 := \sqrt{\langle f, f \rangle} = \left(\int f^2 d\mu \right)^{1/2}$$

Insbesondere gilt also für $f, g \in L^2(\Omega)$

$$\forall f, g \in L^2(\Omega) : |\langle f, g \rangle| = \left| \int fg d\mu \right| \leq \|f\|_2 \cdot \|g\|_2 \quad (2.16)$$

2.7 Lebesgue- und Riemann-Integral

Satz 2.7.1 *Ist f messbar und Riemann-integrierbar, dann ist f λ^1 -integrierbar und das Riemann-Integral und das λ^1 -Integral von f stimmen überein.*

Beweis [El], S. 151, Satz 6.1. □

Somit ist jede messbare, eigentlich Riemann-integrierbare Funktion auch λ^1 -integrierbar und man kann die üblichen Rechentechniken verwenden.

Ist das Integrationsintervall unbeschränkt, dann gilt der

Satz 2.7.2 *Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ messbar und $f|_{[a,b]}$ Riemann-integrierbar für jedes $[a, b] \subset \mathbb{R}$. f ist genau dann λ^1 -integrierbar, wenn $|f|$ uneigentlich Riemann-integrierbar ist. In diesem Fall stimmen beide Integrale überein.*

Beweis [El], S. 153, Satz 6.3. □

Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) := \frac{\sin x}{x} 1_{(0,\infty)}(x)$ ist uneigentlich Riemann-integrierbar, $|f|$ jedoch nicht. Damit ist f auch nicht λ^1 -integrierbar.

2.8 Produktmaße

2.8.1 Konstruktion

Gegeben seien Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$. Ziel ist es auf $\Omega_1 \times \Omega_2$ eine σ -Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ und ein Maß $\mu_1 \otimes \mu_2 : \mathcal{A}_1 \otimes \mathcal{A}_2 \rightarrow [0, \infty]$ zu definieren so, dass für $A_i \in \mathcal{A}_i$ gilt:

$$\begin{aligned} A_1 \times A_2 &\in \mathcal{A}_1 \otimes \mathcal{A}_2 \\ \mu_1 \otimes \mu_2(A_1 \times A_2) &= \mu_1(A_1) \mu_2(A_2) \end{aligned} \quad (2.17)$$

Wir wenden zur Konstruktion Satz 2.4.1 ein. Sei $\mathcal{E} := \mathcal{A}_1 \times \mathcal{A}_2 := \{A_1 \times A_2 | A_i \in \mathcal{A}_i, i = 1, 2\}$. Auf $\mathcal{E}$ wird mit (2.17) eine Mengenfunktion definiert. Die σ -Additivität kann man mit Integration führen. Das System $\mathcal{E}$ erfüllt die Bedingungen von Satz 2.4.1 (a). Erfüllen $\mu_i, \mathcal{A}_i$ auch die Bedingung (2.4) dann kann die Mengenfunktion (2.17) mit Satz 2.4.1 in eindeutiger Weise auf $\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\mathcal{E})$ zu einem Maß fortgesetzt werden. Entsprechend konstruiert man das Produktmaß endlich vieler Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, \dots, n$.

2.8.2 Integration bezüglich Produktmaßen

Wir greifen die Notation und die Voraussetzungen des vorangehenden Abschnitts 2.8.1 auf. Im Folgenden befassen wir uns mit dem Integral bezüglich des Produktmaßes und dessen Berechnung.

Satz 2.8.1 (Fubini) *Sei $f \in \mathcal{M}(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, $f \geq 0$ f.ü. Dann sind die Funktionen $f(\cdot, \omega_2)$ bzw. $f(\omega_1, \cdot)$ für alle $\omega_1 \in \Omega_1$ bzw. $\omega_2 \in \Omega_2$ messbar, die Funktionen*

$$\begin{aligned}\Omega_1 \ni \omega_1 &\longmapsto \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \in [0, \infty] \\ \Omega_2 \ni \omega_2 &\longmapsto \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \in [0, \infty]\end{aligned}$$

sind $(\Omega_1, \mathcal{A}_1)$ bzw. $(\Omega_2, \mathcal{A}_2)$ messbar und es gilt

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).\end{aligned}$$

Ist einer der Terme endlich, dann auch die beiden anderen.

Insbesondere gilt für $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$:

$$\mu_1 \otimes \mu_2(C) = \int_{\Omega_1} \mu_2(C_{\omega_1}) d\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(C_{\omega_2}) d\mu_2(\omega_2). \quad (2.18)$$

Hierbei sei für $\omega_i \in \Omega_i$

$$\begin{aligned}C_{\omega_1} &:= \{\omega_2 \in \Omega_2 \mid (\omega_1, \omega_2) \in C\} \\ C_{\omega_2} &:= \{\omega_1 \in \Omega_1 \mid (\omega_1, \omega_2) \in C\}\end{aligned}$$

Dies folgt aus dem Satz von Fubini (Satz 2.8.1) indem man $f := 1_C$ wählt. Dann ist

$$\begin{aligned}\int_{\Omega_1} 1_C(\omega_1, \omega_2) d\mu_1(\omega_1) &= \mu_1(C_{\omega_2}) \\ \int_{\Omega_2} 1_C(\omega_1, \omega_2) d\mu_2(\omega_2) &= \mu_2(C_{\omega_1}).\end{aligned}$$

2.9 Dichten

$(\Omega, \mathcal{A}, \mu)$ sei stets ein Maßraum.

Dichten spielen in der Wahrscheinlichkeitstheorie eine entscheidende Rolle.

Lemma 2.9.1 Sei $f \in \mathcal{M}(\Omega, \mathcal{A})$, $f \geq 0$ f.ü. Dann ist $\nu : \mathcal{A} \longrightarrow [0, \infty]$,

$$\nu(A) := \int_A f d\mu$$

ein Maß auf $(\Omega, \mathcal{A})$. Jede μ -Nullmenge ist auch ν -Nullmenge.

Beweis Zu zeigen ist die σ -Additivität. Seien $A_n \in \mathcal{A}, n \in \mathbb{N}$ paarweise disjunkt, $A := \bigcup_{n \in \mathbb{N}} A_n$. Dann gilt

$$0 \leq g_n := \sum_{k=1}^n 1_{A_k}, \quad g_n \uparrow 1_A$$

und somit auch $g_n f \uparrow 1_A f$. Mit dem Satz von dem monotonen Konvergenz Satz 2.6.13 folgt

$$\lim_{n \rightarrow \infty} \underbrace{\int_{\Omega} g_n f d\mu}_{\sum_{k=1}^n \nu(A_k)} = \int_{\Omega} 1_A f d\mu = \nu(A) = \sum_{k=0}^{\infty} \nu(A_k).$$

Ist $N \in \mathcal{A}$ eine μ -Nullmenge, dann ist laut Korollar 2.6.11

$$\int_N f d\mu = 0$$

also $\nu(N) = 0$. □

Definition 2.9.2 Seien μ, ν Maße auf $(\Omega, \mathcal{A})$. Eine Abbildung $f \in \mathcal{M}(\Omega, \mathcal{A})$ heißt μ -**Dichte** von ν , wenn gilt

$$\forall A \in \mathcal{A} : \nu(A) = \int_A f d\mu.$$

Bezeichnung: $f\mu$

Beispiel 2.9.3 Sei $c > 0$, $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) := ce^{-cx} 1_{[0, \infty)}$. Bestimme $\nu([0, a])$ für $a \in \mathbb{R}$.

Lösung $\nu([0, a]) = 1 - \exp(-ca)$ für $a > 0$, sonst 0. □

Beispiel 2.9.4 Das Dirac-Maß δ_0 auf $\mathcal{B}^1$ besitzt keine λ^1 -Dichte.

Lösung $\{0\}$ ist eine λ^1 -Nullmenge, aber wegen $\delta_0(\{0\}) = 1$ keine δ_0 -Nullmenge. Somit folgt die Behauptung mit Lemma 2.9.1. $\square$

Für jede Dichte f gilt $f \geq 0$ f.ü. (Warum?) Die Integration bezüglich $f\mu$ behandeln wir im nächsten Satz:

Satz 2.9.5 Sei $f \in \mathcal{M}(\Omega, \mathcal{A})$, $f \geq 0$ f.ü. eine Dichte. Ist $g \in \mathcal{M}(\Omega, \mathcal{A})$, $g \geq 0$, dann gilt

$$\int_{\Omega} g d(f\mu) = \int_{\Omega} gf d\mu.$$

Beweis (i) Ist g eine Treppenfunktion mit $g = \sum_{i=1}^n \alpha_i 1_{A_i}$, $\alpha_i > 0$, $A_i \in \mathcal{A}$, dann

$$\begin{aligned} \int g d(f\mu) &= \sum_{i=1}^n \alpha_i \int 1_{A_i} d(f\mu) = \sum_{i=1}^n \alpha_i (f\mu)(A_i) \\ &= \sum_{i=1}^n \alpha_i \int_{A_i} f d\mu = \sum_{i=1}^n \alpha_i \int 1_{A_i} f d\mu = \int \sum_{i=1}^n \alpha_i 1_{A_i} f d\mu \\ &= \int fg d\mu. \end{aligned}$$

(ii) Sei $g_n \uparrow g$ eine Folge von Treppenfunktionen, $g_n \geq 0$. Dann folgt mit (i)

$$\int g_n d(f\mu) = \int g_n f d\mu.$$

Die linke Seite konvergiert nach Konstruktion des Integrals gegen $\int g d(f\mu)$. Die rechte Seite konvergiert mit dem Satz von der Monotonen Konvergenz wegen $fg_n \uparrow fg$:

$$\lim_{n \rightarrow \infty} \int g_n f d\mu = \int fg d\mu.$$

$\square$

Man kann auch ein Kriterium angeben, wann ν eine μ -Dichte besitzt:

Satz 2.9.6 (Radon-Nykodym) Seien μ, ν Maße auf $(\Omega, \mathcal{A})$, die die (2.4) erfüllen und

$$\forall A \in \mathcal{A} : \mu(A) = 0 \implies \nu(A) = 0.$$

Dann besitzt ν eine μ -Dichte.

Beweis [B-MT], S. 116, 17.10 Satz.

2.10 Bildmaße

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\Omega', \mathcal{A}')$ ein Messraum und $T : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ messbar. Dann definieren wir

$$\mu_T : \mathcal{A}' \longrightarrow [0, \infty], \quad \mu_T(A') := \mu(T^{-1}(A')).$$

Satz und Definition 2.10.1 μ_T ist ein Maß auf $(\Omega', \mathcal{A}')$ und $f \in \mathcal{M}(\Omega', \mathcal{A}')$. Gilt $f \geq 0$ oder $f \circ T$ μ -integrierbar, dann folgt

$$\int_{\Omega'} f d\mu_T = \int_{\Omega} f \circ T d\mu \quad (2.19)$$

Das Maß μ_T heißt **Bildmaß von μ bezüglich T** .

Beweis Seien nun $A'_1, A'_2, \dots \in \mathcal{A}'$ paarweise disjunkt. Dann gilt

$$T^{-1}\left(\bigcup_{i \geq 1} A'_i\right) = \bigcup_{i \geq 1} T^{-1}(A'_i)$$

und die $T^{-1}(A'_i)$ sind paarweise disjunkt. Somit folgt wegen der σ -Additivität von μ

$$\mu_T\left(\bigcup_{i \geq 1} A'_i\right) = \mu\left(\bigcup_{i \geq 1} T^{-1}(A'_i)\right) = \sum_{i \geq 1} \mu(T^{-1}(A'_i)) = \sum_{i \geq 1} \mu_T(A'_i).$$

Wir beweisen (2.19).

(i) f ist Treppenfunktion mit $f = \sum_{i=1}^n \alpha_i 1_{A'_i}$. Dann

$$\begin{aligned} \int_{\Omega'} f d\mu_T &= \sum_{i=1}^n \alpha_i \mu_T(A'_i) = \sum_{i=1}^n \alpha_i \mu(T^{-1}(A'_i)) \\ &= \sum_{i=1}^n \alpha_i \int_{\Omega} 1_{T^{-1}(A'_i)}(\omega) d\mu(\omega) \\ &= \sum_{i=1}^n \alpha_i \int_{\Omega} 1_{A'_i}(T(\omega)) d\mu(\omega) \\ &= \int_{\Omega} \sum_{i=1}^n \alpha_i 1_{A'_i}(T(\omega)) d\mu(\omega) \\ &= \int_{\Omega} f \circ T d\mu. \end{aligned}$$

(ii) Sei $f_n \uparrow f$, $f_n \geq 0$ Treppenfunktionen. Wegen (i) gilt für alle $n \in \mathbb{N}$

$$\int_{\Omega'} f_n d\mu_T = \int_{\Omega} f_n \circ T d\mu. \quad (2.20)$$

Gemäß Konstruktion des Integrals gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega'} f_n d\mu_T = \int_{\Omega'} f d\mu_T,$$

und aus $f_n \circ T \uparrow f \circ T$ folgt mit dem Satz von der monotonen Konvergenz (Satz 2.6.13)

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \circ T d\mu = \int_{\Omega} f \circ T d\mu.$$

(2.19) folgt nun durch Grenzübergang (2.20) auf beiden Seiten. $\square$

Teil II

Wahrscheinlichkeitstheorie

3 Grundlagen der Wahrscheinlichkeitsrechnung

Wir geben eine kurze Motivation und Überblick zum Aufbau der Wahrscheinlichkeitstheorie.

3.1 Zufällige Ereignisse

Ein Vorgang, der beliebig oft wiederholbar ist, und dessen Endzustand oder *Ergebnis* vom Zufall abhängt, heißt *Zufallsexperiment*. Wir wollen annehmen, dass die Menge der möglichen Ergebnisse bekannt ist, d.h. jedes möglichen Ergebnis entspricht einem Element ω einer Menge Ω . Die Menge Ω heißt *Ergebnismenge*, die Elemente ω von Ω heißen auch *Elementarereignisse*.

Beispiel 3.1.1 (a) Werfen eines Würfels: die Elementarereignisse sind

$$\omega_1 = 1, \omega_2 = 2, \omega_3 = 3, \omega_4 = 4, \omega_5 = 5, \omega_6 = 6,$$

und die Ergebnismenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(b) Ermitteln einer Fahrstrecke (in km), die ein PKW mit 10 Litern Treibstoff zurücklegen kann. Die Ergebnismenge ist $\Omega = [0, \infty)$.

(c) Überprüfung von n Geräten, ob sie defekt sind oder intakt. Die Elementarereignisse sind Vektoren mit n Komponenten,

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i = \text{intakt oder defekt}, i = 1, \dots, n\}.$$

Eine andere Möglichkeit ist hier den defekten Geräten die Zahl 1 und den intakten Geräten die Zahl 0 zuzuordnen. Welche Ergebnismenge würde sich dann ergeben?

(d) Ermittlung einer Messreihe durch Bestimmung der Lebensdauern (in Stunden) von n Geräten. $\Omega = \mathbb{R}^n$

Neben den Elementarereignissen sind noch Mengen von Elementarereignissen von Interesse, die wir als **Ereignisse** bezeichnen.

Beispiel 3.1.2 (Fortsetzung von Beispiel 3.1.1) (a) Bei einmaligem Wurf eines Würfels „ungerade Augenzahl“: $A = \{1; 3; 5\}$

(b) „Mindestens 70 km“: $A = \{\omega : \omega \geq 70 \text{ km}\}$

(c) „Mindestens zwei Geräte sind defekt“: Wir verwenden $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i = 0 \text{ oder } \omega_i = 1, i = 1, \dots, n\} = \{0, 1\}^n$.

$$A = \{\omega : \sum_{i=1}^n \omega_i \geq 2\}.$$

(d) „Die durchschnittliche Lebensdauer der untersuchten Geräte liegt zwischen 100 und 120 Stunden“:

$$A = \left\{ \omega : 100 \leq \frac{\sum_{i=1}^n \omega_i}{n} \leq 120 \right\}$$

Man verwendet die Sprechweise „das Ereignis A tritt ein“, wenn ein Ergebnis ω mit $\omega \in A$ auftritt.

Wir stellen einige weitere Sprechweisen für Ereignisse vor:

Bezeichnung	Beschreibung	Mengen
A ist sicheres Ereignis	A tritt sicher ein	$A = \Omega$
A ist unmögliches Ereignis	A tritt sicher nicht ein	$A = \emptyset$
A ist Teilergebnis von B	wenn A eintritt dann auch B	$A \subset B$
A und B sind disjunkte Ereignisse	A und B können nicht gleichzeitig eintreten	$A \cap B = \emptyset$
A und B sind komplementär	entweder A oder B tritt ein	$A = \Omega \setminus B$
A ist Vereinigung der A_i mit $i \in \mathbb{N}$,	eines der Ereignisse A_i tritt ein	$A = \bigcup_{i \in \mathbb{N}} A_i$
A ist Durchschnitt der A_i mit $i \in \mathbb{N}$,	alle Ereignisse A_i treten ein	$A = \bigcap_{i \in \mathbb{N}} A_i$
A ist die Differenz zweier Ereignisse B und C	es tritt B aber nicht C ein	$B \setminus C$

Die oben zusammengestellten Operationen mit Ereignissen sollen stets zu weiteren Ereignissen führen und sollen zur Gesamtheit aller in Frage kommenden Ereignisse gehören. Das Ereignissystem sollte also eine σ -Algebra $\mathcal{A}$ sein. Als **Ereignisse** bezeichnen wir die Elemente einer σ -Algebra.

3.2 Axiome der Wahrscheinlichkeitstheorie

Nun wollen wir jedem Ereignis A auch eine Zahl zuordnen, die ausdrückt, wie wahrscheinlich es ist, dass A eintritt. Ein naheliegender Ansatz ist der

Folgende: das Zufallsexperiment wird n -mal unter den gleichen Bedingungen durchgeführt. Bezeichnet man mit n_A die Anzahl der Versuche, bei denen das Ereignis A eingetreten ist und mit n die Gesamtanzahl der Versuche, so lehrt die Erfahrung, dass die relativen Häufigkeiten $h_A := \frac{n_A}{n}$ mit wachsendem n einem für A charakteristischen Zahlenwert $\mathbb{P}(A)$ zustrebt.

Betrachten wir $A = \emptyset$, das unmögliche Ereignis, dann werden wir $h_A = 0$ erhalten, also

$$h_{\emptyset} = 0. \quad (3.1)$$

Für das sichere Ereignis $A = \Omega$ werden wir $h_A = 1$ erhalten, also

$$h_{\Omega} = 1. \quad (3.2)$$

Schließlich werden wir im Falle disjunkter Ereignisse A und B $n_A + n_B = n_{A \cup B}$ zählen, also

$$h_{A \cup B} = h_A + h_B, \text{ für } A \cap B = \emptyset. \quad (3.3)$$

Es liegt nahe, die zur mathematischen Beschreibung zu wählenden Zahlen $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cup B)$ so zu wählen, dass sie unserer Vorstellung in (3.1)-(3.3) bezüglich der relativen Häufigkeiten möglichst nahe kommen.

Definition 3.2.1 *Ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ bestehend aus einer (Ergebnis-)Menge Ω , einer σ -Algebra $\mathcal{A}$ und einem Maß $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ mit $\mathbb{P}(\Omega) = 1$ heißt **Wahrscheinlichkeitsraum**, das Maß $\mathbb{P}$ heißt **Wahrscheinlichkeitsmaß**. Ist Ω höchstens abzählbar, dann heißt der $(\Omega, \mathcal{A}, \mathbb{P})$ **diskret**. In diesen Fällen gilt $\mathcal{A} = \mathcal{P}(\Omega)$.*

Beispiel 3.2.2 *Für einen Münzwurf sei $\Omega = \{Z, W\}$ und die σ -Algebra gegeben durch $\{\emptyset, \Omega, \{Z\}, \{W\}\}$. Dann ist z.B. die Funktion $\mathbb{P}$ mit $\mathbb{P}(Z) = 0,5$, $\mathbb{P}(W) = 0,5$ ein Wahrscheinlichkeitsmaß. Auch die Funktion $\tilde{\mathbb{P}}$ mit $\tilde{\mathbb{P}}(Z) = 0,1$, $\tilde{\mathbb{P}}(W) = 0,9$ ist formal ein Wahrscheinlichkeitsmaß. Welches ist nun das „richtige“?*

Die Bestimmung des „richtigen“ Wahrscheinlichkeitsmaßes ist in der Praxis oft schwierig (wie groß ist z.B. die „richtige“ Wahrscheinlichkeit, dass es morgen regnet?) und ist eins der Themen der Statistik. In einigen Fällen lässt sich das richtige Wahrscheinlichkeitsmaß jedoch bestimmen, z.B. wenn nur endlich viele Elementarereignisse existieren, die alle gleich wahrscheinlich sind (z.B. Münze, Würfel, Roulette). Dann reduziert sich die Bestimmung der Wahrscheinlichkeit auf das „Abzählen“ von Ereignissen.

Definition 3.2.3 (Laplacesche Wahrscheinlichkeit) *Besteht die Menge der Elementarereignisse Ω aus endlich vielen Elementarereignissen, die alle*

gleich wahrscheinlich sind, so ist die Wahrscheinlichkeit jedes Ereignisses A gegeben durch

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

wobei die Betragsstriche die Anzahl der Elementarereignisse bezeichnen.

Beispiel 3.2.4 (a) Bei einem Würfel gilt: $\Omega = \{1, 2, 3, 4, 5, 6\}$ also $|\Omega| = 6$ und

$$\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = \mathbb{P}(5) = \mathbb{P}(6) = \frac{1}{6}.$$

und

$$\mathbb{P}(\{1, 2\}) = \frac{2}{6}.$$

(b) Dreimaliges Werfen einer Münze. Bei jedem Wurf sind die Ergebnisse 0 (Wappen) und 1 (Zahl) möglich. Wie groß ist die Wahrscheinlichkeit des Ereignisses „mindestens einmal Wappen“?

Wir haben

$$\Omega = \{0, 1\}^3 = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{0, 1\}, i = 1, 2, 3\} \implies |\Omega| = 8$$

und

$$A = \Omega \setminus \{(1, 1, 1)\} \implies |A| = 8 - 1 = 7.$$

Damit erhalten wir $\mathbb{P}(A) = \mathbb{P}(\text{„mindestens einmal Wappen“}) = \frac{7}{8}$.

Direkt aus den Eigenschaften des W-Maßes $\mathbb{P}$ leitet man folgende Rechenregeln her:

Satz 3.2.5 (Regeln für Wahrscheinlichkeiten) Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum, $A, B \in \mathcal{A}$. Es gilt:

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c), \quad (3.4)$$

$$\mathbb{P}(A) \leq 1 \quad (3.5)$$

$$\mathbb{P}(\emptyset) = 0 \quad (3.6)$$

$$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B) \quad (3.7)$$

$$A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B) \text{ („Monotonie“)} \quad (3.8)$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \quad (3.9)$$

Beweis Klar. □

3.3 Kombinatorik

Bei der Berechnung von Laplace Wahrscheinlichkeiten müssen oft Mächtigkeiten von Ereignissen berechnet werden. Hier werden die Grundaufgaben dargestellt und zwar am Beispiel von mehrfachen Ziehungen aus einer Urne, die n Kugeln enthält. Diese werden mit $1, 2, \dots, n$ nummeriert. Es werden insgesamt k Kugeln zufällig gezogen. Folgende Stichproben werden unterschieden:

- **Stichprobe mit Zurücklegen:** Kugeln dürfen mehrfach gezogen werden, also wird jede Kugel nach dem Ziehen in die Urne zurückgelegt.
- **Stichprobe ohne Zurücklegen:** Jede Kugel kann nur höchstens einmal auftreten.
- **Stichprobe mit Reihenfolge:** Es kommt auf die Reihenfolge an, in der die Kugeln gezogen wurden. (Wird zweimal gezogen dann sind $(1, 2)$ und $(2, 1)$ unterschiedliche Ergebnisse.)
- **Stichprobe ohne Reihenfolge:** Man interessiert sich nur für die Menge der gezogenen Kugeln. (Die Ergebnisse $(1, 2)$ und $(2, 1)$ werden nicht unterschieden). Es ergeben sich nun vier Ereignisräume, mit folgenden Mächtigkeiten.

	in Reihenfolge	ohne Reihenfolge
mit Zurücklegen	n^k	$\binom{n+k-1}{k}$
ohne Zurücklegen	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Weiteres findet man in [Kr], Kapitel 1.

4 Zufallsvariablen und Wahrscheinlichkeitsverteilungen

Ein wichtiges Konzept der Wahrscheinlichkeitsrechnung ist der Begriff der Zufallsvariablen. Aussagen über Wahrscheinlichkeiten beziehen sich weniger auf konkrete Ereignisse sondern auf Zufallsvariablen.

Generell gehen wir von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ aus.

4.1 Zufallsvariablen

Mit vielen Ereignissen wie z.B. Zahl oder Wappen beim Münzwurf lässt sich nicht rechnen. Während bei einem Würfel die Aussage geleistet werden kann, dass z.B. die 6 größer als die 1 ist, ist eine derartige Aussage bei einem Münzwurf nicht möglich. Deshalb ordnet man jedem Element der Ergebnismenge Ω eines Zufallsexperiments eine reelle Zahl zu.

Beispiel 4.1.1 *Das Gewicht eines in einer Hühnerfarm gelegten Eis kann als Ergebnis eines Zufallsexperiments interpretiert werden, die Ergebnismenge ist dann $\Omega = [0, \infty)$. Die Eier werden in Gewichtsklassen eingeteilt. Bei der Kalkulation ist der Erlös eines Eis von Interesse. Dafür ist nicht das genaue Gewicht des Eis, sondern nur der Preis für seine Gewichtsklasse entscheidend. Also wird jedem Gewicht $\omega \in \Omega$ ein Preis $X(\omega)$ zugeordnet.*

Da das Ergebnis ω eines Zufallsexperiments zufallsabhängig ist, wird auch der ermittelte Zahlenwert $X(\omega)$ vom Zufall abhängen. Wir interessieren uns also für die Wahrscheinlichkeit, dass $X(\omega)$ in einem bestimmten Intervall $I \subset \mathbb{R}$ liegt, also für die Wahrscheinlichkeit von $X(\omega) \in I$.

Beispiel 4.1.2 *Zweimaliges Werfen einer Würfel. Die Ergebnismenge ist $\Omega = \{1, 6\}^2$. Wir interessieren uns für die Wahrscheinlichkeit, dass die Augensumme zwischen 6 und 9 liegt. Dann definieren wir*

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2,$$

und bestimmen also $\mathbb{P}(6 \leq X \leq 9)$.

Wir kommen nun zur Definition der Zufallsvariablen.

Definition 4.1.3 *Eine **Zufallsvariable** ist eine messbare Abbildung $X : \Omega \longrightarrow \mathbb{R}$.*

Bemerkung Die Messbarkeit von X besagt nichts anders, als dass für reelle Zahlen a, b die Wahrscheinlichkeiten

$$\mathbb{P}(a \leq X \leq b), \mathbb{P}(a < X < b), \mathbb{P}(a \leq X), \dots$$

definiert sind. Dabei schreiben wir allgemein für eine Menge $B \subset \mathbb{R}$, $B \in \mathcal{B}$

$$\begin{aligned} \{X \in B\} &:= \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B). \\ \mathbb{P}(X \in B) &:= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)). \end{aligned}$$

Wir unterscheiden generell zwischen Zufallsvariablen $X, Y, Z, \dots$ und den Werten, die sie annehmen $x, y, z, \dots$. Diese Werte werden als **Realisierung** oder **Realisation** bezeichnet. Vor der Durchführung eines Zufallsexperiments ist ungewiss welches Ergebnis eintritt, also ist ungewiss welchen Wert die Zufallsvariable annimmt. Zufallsvariable (ZV) werden mit Großbuchstaben, die Realisationen mit Kleinbuchstaben bezeichnet.

Beispiel 4.1.4 • *Wir ziehen Buchstaben aus dem Alphabet und können somit die Buchstaben von a-z ziehen, also $\Omega = \{a, b, c, \dots, z\}$. Hier könnten wir zum Beispiel die Zufallsvariable X mit $X(a) = 1$, $X(b) = 2$, $X(c) = 3$, usw. definieren.*

- *Ereignisse wären z.B. $X \in [2, 5]$ oder $X > 10$.*
- *Ziehen wir ein zweites Mal mit Zurücklegen, so könnte die Zufallsvariable Y beispielsweise bezüglich des zweiten Zugs genauso definieren wie X beim ersten Zug: $Y(a) = 1$, $Y(b) = 2$, usw.*
- *Wie groß ist nun die Wahrscheinlichkeit für die Realisationen $x = 3$ und $y = 5$?*

Antwort:

$$\frac{1}{26} \cdot \frac{1}{26} \approx 0,00148.$$

- *Welche Realisation nehmen in diesem Fall die Zufallsvariablen $X^2 + 3$ und $4Y^3 - 7$ an?*

Antwort: 12 bzw. 493.

4.2 Verteilungsfunktionen

4.2.1 Maßerzeugende Funktionen

Satz 4.2.1 *Sei $(\mathbb{R}, \mathcal{B}^1, \mu)$ ein Maßraum. Dann ist die Funktion $F_\mu : \mathbb{R} \rightarrow [0, \infty)$*

$$F_\mu(x) := \mu((-\infty, x])$$

monoton wachsend und von rechts stetig.

Beweis Die Monotonie ist klar. Sei $x \in \mathbb{R}$ und $(x_n)_{n \in \mathbb{N}}$ eine Folge mit $x_n \rightarrow x$, $x_n \geq x$. O.B.d.A gilt $x_n \downarrow x$ (sonst gehe man zu einer Teilfolge über). Es gilt

$$(x_n, \infty) \uparrow (x, \infty).$$

Mit der Stetigkeit von unten Satz 2.3.1 folgt

$$\mu((x, \infty)) = \lim_{n \rightarrow \infty} \mu((x_n, \infty)).$$

Geht man zu den Komplementen über ergibt sich

$$F_\mu(x_n) \longrightarrow F_\mu(x) \quad (n \longrightarrow \infty).$$

□

Auch die Umkehrung gilt:

Satz 4.2.2 Sei $F_\mu : \mathbb{R} \longrightarrow [0, \infty)$ *monoton wachsend und von rechts stetig*. Dann gibt es genau ein Maß μ_F auf $(\mathbb{R}, \mathcal{B}^1)$ mit der Eigenschaft

$$\mu_F((a, b]) = F(b) - F(a). \quad (4.1)$$

Beweis [El], S. 38, 2.2 Satz. □

Die Konstruktion von μ_F beruht auf Satz 2.4.1.

Das Maß μ_F ist in diesem Fall Fortsetzung auf $\mathcal{B}^1$ der durch (4.1) auf der Menge der Intervall definierten Mengenfunktion. Für das Integral bezüglich μ_F wird auch oft

$$\int_{\mathbb{R}} h d\mu_F := \int_{\mathbb{R}} h dF := \int_{\mathbb{R}} h(x) dF(x)$$

geschrieben.

4.2.2 Grundlagen

Definition 4.2.3 (Verteilungsfunktion einer ZV) Zu einer Zufallsvariablen X nennt man die Funktion,

$$F : \mathbb{R} \longrightarrow [0, 1], \quad F(x) := \mathbb{P}(X \leq x)$$

die *Verteilungsfunktion von X* .

Bezeichnung Ist F die Verteilungsfunktion einer Zufallsvariablen X werden wir die kürzere Schreibweise

$$X \sim F$$

verwenden.

Satz 4.2.4 (Eigenschaften von Verteilungsfunktionen) Sei X eine ZV mit der Verteilungsfunktion F .

(a) Es gilt $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$.

(b) F wächst monoton.

(c) F ist in jedem Punkt x von rechts stetig und die linksseitigen Grenzwerte existieren.

(d) Für $a \in \mathbb{R}$ gilt

$$\mathbb{P}(X = a) = F(a) - F(a - 0) \quad (4.2)$$

$$\mathbb{P}(X > a) = 1 - F(a). \quad (4.3)$$

(e) Für reelle Zahlen a, b mit $a < b$ gilt

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) \quad (4.4)$$

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a - 0) \quad (4.5)$$

$$\mathbb{P}(a \leq X < b) = F(b - 0) - F(a - 0) \quad (4.6)$$

$$\mathbb{P}(a < X < b) = F(b - 0) - F(a). \quad (4.7)$$

Beweis Zu (c): Satz 4.2.1.

Zu (d): Analog zu (c) zeigt man

$$\mathbb{P}(X < x) = F(x - 0).$$

Nun folgt

$$F(a) = \mathbb{P}(X \leq a) = \mathbb{P}(X = a) + \mathbb{P}(X < a) = \mathbb{P}(X = a) + F(a - 0).$$

Zu (e)

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < X) + \mathbb{P}(X \leq b) - \underbrace{\mathbb{P}(\{X > a\} \cup \{X \leq b\})}_{=1}$$

$$= 1 - F(a) + F(b) - 1 = F(b) - F(a).$$

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X) + \mathbb{P}(X \leq b) - \underbrace{\mathbb{P}(\{X \geq a\} \cup \{X \leq b\})}_{=1}$$

$$= 1 - F(a - 0) + F(b) - 1 = F(b) - F(a - 0).$$

Die übrigen Gleichungen beweist man analog. □

Bemerkung

(a) Ist die Verteilungsfunktion in x_0 nicht stetig, dann folgt

$$\mathbb{P}(X = x_0) = F(x_0) - F(x_0 - 0)$$

(b) Ist die Verteilungsfunktion stetig, dann sind in den Gleichungen (4.4)-(4.7) wegen $F(x - 0) = F(x)$ für alle $a < b$ die rechten Seiten identisch:

$$\begin{aligned} & F(b) - F(a) \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b). \end{aligned}$$

(c) Bei stetiger Verteilungsfunktion gilt für alle x

$$\mathbb{P}(X = x) = 0.$$

4.2.3 Diskrete und stetige Zufallsvariablen

Mit jeder Zufallsvariablen kann man ein Maß auf $(\mathbb{R}, \mathcal{B}^1)$ definieren, nämlich $\mathbb{P}_X$, das Bildmaß von $\mathbb{P}$ bezüglich X , also (s. auch Definition und Satz 2.10.1, S. 26)

$$\forall B \in \mathcal{B} : \mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

Definition 4.2.5 (a) Eine Zufallsvariable X heißt **diskret**, wenn X höchstens abzählbar viele Werte annimmt.

(b) Eine Zufallsvariable X heißt **stetig (verteilt)**, wenn $\mathbb{P}_X$ eine λ^1 -Dichte besitzt.

Bei diskreten Zufallsvariablen mit $X(\Omega) = \{x_k | k \in I \subset \mathbb{N}\}$ nennt man die Funktion $\mathbb{R} \ni x \mapsto \mathbb{P}(X = x) \in [0, 1]$ die **Wahrscheinlichkeitsfunktion** von X .

Bei stetigen Zufallsvariablen mit Dichte f gilt für die Verteilungsfunktion F

$$F(x) = \int_{-\infty}^x f(t) dt := \int_{-\infty}^x f(t) d\lambda^1(t), \quad x \in \mathbb{R}$$

denn laut Definition gilt

$$F(x) = \mathbb{P}_X((-\infty, x]) = \int_{-\infty}^x f(t) dt.$$

Aus der Stetigkeit von F_X folgt **nicht** die Existenz einer Dichte von $\mathbb{P}_X$! Dazu muss F_X absolut stetig sein. Der Begriff stetig (verteilte) Zufallsvariable ist zwar ungenau aber dennoch üblich.

4.3 Maßzahlen für Verteilungen

4.3.1 Erwartungswert

Definition 4.3.1 Es sei X eine Zufallsvariable. Ist $\mathbb{P}$ -integrierbar, dann heißt

$$\mathbb{E}(X) := \int_{\Omega} X d\mathbb{P}$$

Erwartungswert von X .

Satz 4.3.2 Sei X eine Zufallsvariable mit Erwartungswert $\mathbb{E}(X) \in \mathbb{R}$. Dann gilt

$$\mathbb{E}(X) = \int_{\mathbb{R}} x d\mathbb{P}_X(x). \quad (4.8)$$

Insbesondere gilt also:

(a) Für X diskret mit $X(\Omega) = \{x_i : i \in I\}$ mit $I \subset \mathbb{N}$

$$\mathbb{E}(X) = \sum_{i \in I} x_i \cdot \mathbb{P}(X = x_i). \quad (4.9)$$

(b) Für X eine stetige Zufallsvariable mit Dichte f

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot f(x) dx. \quad (4.10)$$

Beweis (4.8) folgt aus dem Transformationssatz für die Integration bezüglich Bildmaßen (aus mit dem Satz 2.10.1, (2.19), S. 26):

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\Omega} \text{id} \circ X d\mathbb{P} \stackrel{(2.19))}{=} \int_{\mathbb{R}} \text{id} d\mathbb{P}_X = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

Zu (4.9): Sei $A_i := X^{-1}(x_i) = \{X = x_i\}$, $i \in I$. Dann gilt

$$X = \sum_{i \in I} x_i 1_{A_i}.$$

(i) Ist $X(\Omega) = \{x_1, \dots, x_n\}$ also endlich, dann folgt

$$\int X d\mathbb{P} = \int \sum_{i=1}^n x_i 1_{A_i} d\mathbb{P} = \sum_{i=1}^n x_i \int 1_{A_i} d\mathbb{P} = \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

(ii) Ist $X \geq 0$, $X(\Omega) = \{x_i | i \in \mathbb{N}\}$, also abzählbar, erhalten wir mit dem Satz von der monotonen Konvergenz und (i)

$$\begin{aligned} \int X d\mathbb{P} &= \int \sum_{i=1}^{\infty} x_i 1_{A_i} d\mathbb{P} = \int \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i 1_{A_i} d\mathbb{P} = \lim_{n \rightarrow \infty} \int \sum_{i=1}^n x_i 1_{A_i} d\mathbb{P} \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i \mathbb{P}(A_i) = \sum_{i=1}^{\infty} x_i \mathbb{P}(A_i). \end{aligned}$$

(iii) Ist X integrierbar, dann führen die Überlegungen von (ii) angewendet auf X_- und X_+ zur Behauptung.

(b) Aus (4.8) erhalten wir mit Satz 2.9.5 über die Integration bezüglich Maßen mit Dichten

$$\mathbb{E}(X) = \int_{\mathbb{R}} x d\mathbb{P}_X = \int_{\mathbb{R}} x f(x) dx.$$

□

Korollar 4.3.3 Sei X eine Zufallsvariable, $g : \mathbb{R} \rightarrow \mathbb{R}$ messbar, $g \circ X$ integrierbar. Dann gilt

$$\mathbb{E}(g \circ X) = \int_{\mathbb{R}} g(t) d\mathbb{P}_X(t) \quad (4.11)$$

(a) Ist X diskret dann gilt

$$\mathbb{E}(g(X)) = \sum_{i=1}^{\infty} g(x_i) \cdot \mathbb{P}(X = x_i). \quad (4.12)$$

(b) Ist X eine stetig verteilte Zufallsvariable mit Dichte f , $f \cdot g$ sei integrierbar. Dann gilt

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f(x) dx. \quad (4.13)$$

Beweis Es gilt wegen Satz 2.10.1

$$\mathbb{E}(g \circ X) = \int \Omega g \circ X d\mathbb{P} = \int_{\mathbb{R}} g d\mathbb{P}_X.$$

(a) Ist $X(\Omega) = \{x_i | i \in I\}$, $I \subset \mathbb{N}$ dann gilt $g(X(\Omega)) = \{g(x_i) | i \in I\}$, also ist auch die Zufallsvariable $g \circ X$ diskret und es folgt (4.12) aus (4.9).

(b) Integration bezüglich Bildmaßen (Satz 2.10.1 (2.19)) ergibt

$$\int_{\Omega} g \circ X d\mathbb{P} = \int_{\mathbb{R}} g d\mathbb{P}_X = \int_{\mathbb{R}} g(x) f(x) dx.$$

□

Wir werden im weiteren Verlauf die Gleichungen (4.12), (4.13) anwenden.

Beispiel 4.3.4 Sei X die Zufallsvariable: Ergebnis beim einmaligen Wurf eines Würfels. Dann ergibt sich

$$\mathbb{E}(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \dots + \frac{1}{6} \cdot 6 = 3,5$$

Satz 4.3.5 (Rechenregeln für Erwartungswerte) Seien X, Y bzw. X_i für $i = 1, \dots, n$ Zufallsvariablen, für die die Erwartungswerte existieren.

(a) $\mathbb{E}(\alpha X + \beta) = \alpha \mathbb{E}(X) + \beta$ für alle $\alpha, \beta \in \mathbb{R}$.

(b)

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i). \quad (4.14)$$

(c) Sei $X \geq 0$ eine Zufallsvariable mit Verteilungsfunktion F , dann gilt

$$\mathbb{E}(X) = \int_0^{\infty} (1 - F(x)) \, dx \quad (4.15)$$

falls das Integral existiert.

Beweis (a) und (b) sind klar. Zu (4.15): Wir werden den Satz von Fubini, Satz 2.8.1 verwenden. Dazu sei

$$A := \{(s, t) \in \mathbb{R}^2 \mid t > 0, \quad 0 < s < t\}.$$

Auf $(\mathbb{R}^2, \mathcal{B}^2)$ betrachten wir das Produktmaß $\lambda^1 \otimes \mathbb{P}_X$. A ist offen, also messbar. Die Projektionen von A sind gegeben durch

$$\begin{aligned} A_s &= \{t \in \mathbb{R} \mid (s, t) \in A\} = \{t \in \mathbb{R} \mid t > s\} = (s, \infty), \quad s > 0 \\ A_t &= \{s \in \mathbb{R} \mid (s, t) \in A\} = \{s \in \mathbb{R} \mid 0 < s < t\} = (0, t), \quad t > 0 \end{aligned}$$

Wegen (2.18) gilt

$$\lambda^1 \otimes \mathbb{P}_X(A) = \int_{\mathbb{R}} \mathbb{P}_X(A_s) \, d\lambda^1(s) = \int_{\mathbb{R}} \lambda^1(A_t) \, d\mathbb{P}_X(t). \quad (4.16)$$

und damit folgt

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} t \, d\mathbb{P}_X(t) = \int_0^{\infty} \lambda^1((0, t)) \, d\mathbb{P}_X(t) = \int_{\mathbb{R}} \lambda^1(A_t) \, d\mathbb{P}_X(t) \\ &\stackrel{(4.16)}{=} \int_{\mathbb{R}} \mathbb{P}_X(A_s) \, d\lambda^1(s) = \int_0^{\infty} (1 - F(s)) \, d\lambda^1(s) = \int_0^{\infty} (1 - F(s)) \, ds. \end{aligned}$$

□

Ist X eine stetige Zufallsvariable mit Dichte f dann wird der Beweis des obigen Satzes „handlicher“.

Bemerkung Die Gleichung (4.15) gibt eine geometrische Deutung des Erwartungswerts für Zufallsvariablen $X \geq 0$: Der Erwartungswert ist die Fläche, die von der y -Achse, F und der Geraden $y = 1$ (Parallele zur x -Achse) eingeschlossen wird.

4.3.2 Varianz

Als Streuungsmaß für die Verteilung einer Zufallsvariablen dient die Varianz:

Definition 4.3.6 Sei X eine Zufallsvariable für die $\mathbb{E}(X)$ und $\mathbb{E}([X - \mathbb{E}(X)]^2)$ existieren. Dann heißt

$$\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2),$$

die **Varianz von X** . (die „durchschnittliche“ quadratischen Abweichung vom Erwartungswert). Die **Standardabweichung** ist definiert als

$$\text{Std}(X) := \sqrt{\text{Var}(X)}.$$

Für diskrete bzw. stetige Zufallsvariablen gilt also

$$\begin{aligned} \text{Var}(X) &= \sum_{i \in I} (x_i - \mathbb{E}(X))^2 \text{ bzw.} \\ \text{Var}(X) &= \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx. \end{aligned}$$

Satz 4.3.7 (Eigenschaften von Varianzen) Sei X quadratintegrierbar. Es gilt:

(a) $\text{Var}(X) \geq 0, \text{Std}(X) \geq 0.$

(b) $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(c) Es gilt für $\alpha, \beta \in \mathbb{R}$

$$\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$$

(d) Sei $\mu := \mathbb{E}(X)$ $\sigma^2 := \text{Var}(X) > 0$. Dann gilt für

$$Y := \frac{X - \mu}{\sigma}$$

$$\mathbb{E}(Y) = 1, \text{Var}(Y) = 1 \text{ } Y \text{ heißt auch die } \textbf{Standardisierung von } X.$$

Beweis Übung

□

4.3.3 Schiefe

Sei X eine Zufallsvariable, für die $(X - \mathbb{E}(X))^3$ integrierbar ist. Dann heißt

$$\frac{\mathbb{E}([X - \mathbb{E}(X)]^3)}{\text{Var}(X)^{3/2}} = \frac{\mathbb{E}([X - \mathbb{E}(X)]^3)}{\text{Std}(X)^3}$$

Schiefe von X . Diese Kennzahl charakterisiert Abweichungen von der Symmetrie des Verteilungsgesetzes von X . Ist X symmetrisch verteilt zu einem Punkt $a \in \mathbb{R}$, dann ist die Schiefe gleich 0.

4.3.4 Höhere Momente

Sei $p \in \mathbb{N}$ und X eine Zufallsvariable für die X^p integrierbar ist. Dann heißt

$$\mathbb{E}(X^p) \text{ bzw. } \mathbb{E}(|X - \mathbb{E}(X)|^p)$$

das p -te bzw. das p -te zentrierte Moment von X . Es gilt wegen (4.11)

$$\mathbb{E}(X^p) = \int_{\mathbb{R}} t^p d\mathbb{P}_X(t).$$

Ist X diskret bzw. stetig verteilt mit Werten $x_i, i \in I \subset \mathbb{N}$ bzw. Dichte f , dann gilt also

$$\mathbb{E}(X^p) = \sum_{i \in I} x_i^p \mathbb{P}(X = x_i) \text{ bzw.}$$

$$\mathbb{E}(X^p) = \int_{\mathbb{R}} t^p f(t) dt.$$

4.3.5 Quantile

Sei $0 < p < 1$ und F die Verteilungsfunktion einer Zufallsvariablen X . Dann heißt die Zahl

$$x_p := \sup\{x \in \mathbb{R} \mid F(x) < p\} = \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$$

das **p -Quantil** von X . Ist F streng monoton wachsend, dann ist x_p eindeutig bestimmt durch die Gleichung

$$F(x_p) = p.$$

Das $p = 0.5$ -Quantil heißt **Median** von X . Nach der Definition genügt das Quantil x_p den folgenden Bedingungen:

$$\mathbb{P}(X < x_p) \leq p, \quad \mathbb{P}(X > x_p) \leq 1 - p.$$

4.4 Ungleichungen

Wir betrachten stets Zufallsvariable über einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Wir beginnen mit der Ungleichung von Markov.

Satz 4.4.1 (Markovsche-Ungleichung) *Sei $X \geq 0$ eine Zufallsvariable für die der Erwartungswert existiert. Dann gilt*

$$\mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(X)}{r}, \quad r > 0.$$

Beweis Sei $r > 0$. Es gilt

$$0 \leq 1_{\{X \geq r\}} \leq \frac{X}{r}$$

und somit

$$0 \leq \mathbb{E}(1_{\{X \geq r\}}) = \mathbb{P}(X \geq r) \leq \mathbb{E}\left(\frac{X}{r}\right) = \frac{\mathbb{E}(X)}{r}$$

□

Korollar 4.4.2 (Tschebyschevsche-Ungleichung) *Sei X eine Zufallsvariable mit Varianz $\text{Var}(X)$. Dann gilt*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}, \quad \epsilon > 0$$

Beweis Wendet man die Ungleichung von Markov auf $|X - \mathbb{E}(X)|^2$ an, folgt für $\epsilon > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq \epsilon^2) \leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{\epsilon^2} = \frac{\text{Var}(X)}{\epsilon^2}$$

□

Korollar 4.4.3 *Sei $p \geq 1$, X^p integrierbar. Dann gilt*

$$\mathbb{P}(|X|^p \geq r) \leq \frac{\mathbb{E}(|X|^p)}{r^p}, \quad r > 0.$$

Beweis Sei $r > 0$. Dann

$$\mathbb{P}(|X| \geq r) = \mathbb{P}(|X|^p \geq r^p) \leq \frac{\mathbb{E}(|X|^p)}{r^p}.$$

□

Satz 4.4.4 (Jensensche Ungleichung) Sei $X : \Omega \longrightarrow (a, b)$ eine integrierbare Zufallsvariable, $\varphi : (a, b) \longrightarrow \mathbb{R}$ konvex. Ist $\varphi \circ X$ integrierbar, dann gilt

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)). \quad (4.17)$$

Beweis Es gilt $\mathbb{E}(X) \in (a, b)$ $[X - a > 0, b - X > 0 \text{ auf } \Omega]$. Damit folgt wegen $\mathbb{P}(\Omega) = 1$

$$0 < \int_{\Omega} (X - a) d\mathbb{P} = \mathbb{E}(X) - a, \quad 0 < \int_{\Omega} (b - X) d\mathbb{P} = b - \mathbb{E}(X).$$

Mit der Ungleichung über den Verlauf von konvexen Funktionen (4.26), S. 60 gibt es ein $\beta \in \mathbb{R}$ mit

$$\forall s \in (a, b) : \varphi(s) \geq \varphi(\mathbb{E}(X)) + \beta(s - \mathbb{E}(X)).$$

Wegen $a < X < b$ folgt nun

$$\varphi(X) \geq \varphi(\mathbb{E}(X)) + \beta(X - \mathbb{E}(X)).$$

Da φ stetig ist, ist $\varphi \circ X$ messbar und durch Integration über Ω bezüglich $\mathbb{P}$ folgt

$$\mathbb{E}(\varphi \circ X) \geq \int_{\Omega} (\varphi(\mathbb{E}(X)) + \beta(X - \mathbb{E}(X))) d\mathbb{P} = \varphi(\mathbb{E}(X)) + 0.$$

□

Korollar 4.4.5 Sei X eine integrierbare Zufallsvariable mit X^p integrierbar. Dann ist X integrierbar und es gilt

$$\mathbb{E}(|X|)^p \leq \mathbb{E}(|X|^p).$$

Beweis Mit $\varphi(x) := x^p$ folgt die Behauptung mit (4.17). □

Bemerkung Allgemein kann man zeigen, dass für endliche Maße gilt

$$X^p \text{ integrierbar} \implies X \text{ integrierbar}$$

[B-MT], S. 88, 14.7 Korollar

4.5 Diskrete Verteilungen

Diskrete Verteilungen werden zur Modellierung von Schadenzenahlen benötigt. Ein diskretes Wahrscheinlichkeitsmaß ist durch die Wahrscheinlichkeitsfunktion (W-fkt.) vollständig beschrieben. Neben der Wahrscheinlichkeitsfunktion werden wir noch die relevanten Maßzahlen angeben und, falls vorhanden, die „anschauliche“ Bedeutung angeben.

Die Verteilungsfunktion diskreter Zufallsvariabler X sind Treppenfunktionen. Einfache Beispiele von diskreten Zufallsvariablen sind

- (a) Augenzahl beim Werfen eines Würfels.
- (b) Anzahl der Schäden in einer Versicherungssparte.

4.5.1 Die Binomialverteilung

Beschreibung

Ein Zufallsexperiment mit den möglichen Ergebnissen „Erfolg“ bzw. „Treffer“ ($=1$) und „Misserfolg“ ($=0$) wird n -mal wiederholt. Die Experimente werden unabhängig voneinander durchgeführt. Dabei soll „Erfolg“ jeweils mit der Wahrscheinlichkeit p eintreten. Beispielsweise könnte beim Roulette das Ereignis „Erfolg“ für „weiße Zahl“ oder „Zahl > 20 “ stehen. Wir fragen uns, wie groß die Wahrscheinlichkeit ist, dass das Ereignis „Erfolg“ bei den n Versuchen genau k -mal eintritt, wobei $k \in \{0, \dots, n\}$.

Die **Binomialverteilung** mit den Parametern $n \in \mathbb{N}$ $p \in (0, 1)$ ist folgenderweise definiert:

Bezeichnung	$X \sim B(n, p)$
W-fkt.	$\binom{n}{k} p^k (1-p)^{n-k}, k \in \{0, \dots, n\}$
Erwartungswert	np
Varianz	$np(1-p)$

Die Verteilung $B(1, p)$ heißt **Bernoulli Verteilung**.

4.5.2 Die negative Binomial-Verteilung

Beschreibung

Wie bei der Binomialverteilung gehen wir von einem Zufallsexperiment mit den möglichen Ergebnissen „Erfolg“ und „Misserfolg“ ($=0$) aus, die Erfolgswahrscheinlichkeit sei p . Das Experiment wird solange wiederholt, bis der

m -te Erfolg eintritt. Wie groß die Wahrscheinlichkeit ist, dass die „Anzahl Versuche“ gleich $k \in \{m, m+1, \dots\}$ ist?

Die **negative Binomialverteilung** mit den Parametern $m \in \mathbb{N}$, $p \in (0, 1)$ ist folgenderweise definiert:

Bezeichnung	$X \sim NB(m, p)$
W-fkt.	$\binom{k-1}{m-1} p^m (1-p)^{k-m}, k \in \{m, m+1, \dots\}$
Erwartungswert	$\frac{m}{p}$
Varianz	$\frac{m(1-p)}{p^2}$

Im Fall $m = 1$ liegt die **geometrische Verteilung** vor.

4.5.3 Die Poisson-Verteilung

Beschreibung

Nimmt man bei der Binomialverteilung ein sehr großes n , dann kann man die Wahrscheinlichkeit von k Treffern näherungsweise mit der Poisson-Verteilung berechnen. Die **Poisson-Verteilung** mit dem Parameter $\lambda > 0$ ist folgenderweise definiert:

Bezeichnung	$X \sim \text{Poi}(\lambda)$
W-fkt.	$e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}_0$
Erwartungswert	λ
Varianz	λ

Bemerkung

- Die Poisson-Verteilung eignet sich zur Beschreibung von Ereignissen folgenden Typs: In einer Telefonzentrale wird die Anzahl der innerhalb von 60 Sekunden eingehenden Anrufe ermittelt. Die „mittlere“ Anzahl ist dann der Parameter λ .

- Eingesetzt wurde die Poisson-Verteilung bei der Untersuchung der preussischen Armee über die Anzahl der Soldaten, die innerhalb eines Jahres an den Folgen eines Huftritts starben. Es waren $\lambda = 0,61$ Soldaten im Jahr. Beobachtet wurden in einem Jahr höchstens 4 Todesfälle. Mit der Poisson-Verteilung kann man allerdings auch die Wahrscheinlichkeit für 5, 6, ... Tote abschätzen. Dazu s. [LW], Beispiel 2.31, S. 44.
- In der Versicherungswirtschaft spielen die Binomial, negativ Binomial und die Poisson-Verteilung eine große Rolle bei der Kalkulation der Prämien. Die Schadenhäufigkeit (Anzahl großer Stürme in einem Jahr, Anzahl der Autounfälle pro 100 Versicherten, etc.) wird mit diesen Verteilungen modelliert.

Satz 4.5.1 (Eigenschaften der Poisson-Verteilung) *Die Poisson-Verteilung mit $\lambda = n \cdot p$ ist zur Approximation der Binomialverteilung geeignet. Man erhält eine brauchbare Näherung für $n \geq 50$, $p \leq 0,1$ und $n \cdot p \leq 10$.*

4.5.4 Die hypergeometrische Verteilung

Nehmen wir an, eine Urne enthalte Kugeln, n_1 weiße und n_2 schwarze Kugeln und wir ziehen $n \leq n_1 + n_2$ mal ohne Zurücklegen. Wie groß ist die Wahrscheinlichkeit, dass wir genau $k \in \{0, \dots, n\}$ weiße Kugeln ziehen?

Die **hypergeometrische Verteilung** mit den Parametern n_1, n_2 und $n \in \{1, \dots, n_1 + n_2\}$ ist folgenderweise definiert:

Bezeichnung	
W-fkt.	$\frac{\binom{n_1}{k} \cdot \binom{n_2}{n-k}}{\binom{n_1+n_2}{n}}, k \in \{0, \dots, n\}$
Erwartungswert	$n \frac{n_1}{n_1 + n_2}$
Varianz	$n \frac{n_1}{n_1 + n_2} \frac{n_2}{n_1 + n_2} \frac{n_1 + n_2 - n}{n_1 + n_2 - 1}$

Setzt man $p := \frac{n_1}{n_1 + n_2}$ dann ergibt sich wenn $n_1 + n_2$ wesentlich größer als n mit obigen Formeln für Erwartungswert und Varianz

$$\mathbb{E}(X) = np, \quad \text{Var}(X) \approx np(1 - p)$$

wie bei einer $B(n, p)$ Verteilung.

Ist die Anzahl der Kugeln $n_1 + n_2$ wesentlich größer als n , so kann die hypergeometrische Verteilung tatsächlich durch die $B\left(n, \frac{n_1}{n_1 + n_2}\right)$ -Verteilung approximiert werden. Hintergrund dafür ist, dass man dann „praktisch“ mit Zurücklegen die Kugeln aus der Urne zieht.

4.6 Stetige Verteilungen

Wir geben jeweils die Dichte f , die Verteilungsfunktion F (wenn möglich) und die relevanten Maßzahlen an.

4.6.1 Die Gleichverteilung

Beschreibung

Die Gleichverteilung ist das stetige Analogon der Laplace-Verteilung. Eine Gleichverteilung auf $[a, b]$ wird als Modell für die Wahl einer Zufallszahl aus $[a, b]$ verwendet.

Die **Gleichverteilung im Intervall** $[a, b]$ mit $a, b \in \mathbb{R}$, $a < b$ (auch Rechteck-Verteilung oder uniforme Verteilung) ist folgenderweise definiert:

Bezeichnung	$X \sim U(a, b)$
f	$1_{[a, b]}$
F	$F(x) = \begin{cases} 0 & \text{falls } x \leq a \\ \frac{x - a}{b - a} & \text{für } x \in [a, b] \\ 1 & \text{falls } x \geq b. \end{cases}$
Erwartungswert	$\frac{a + b}{2}$
Varianz	$\frac{(b - a)^2}{12}$

Diese einfachste stetige Verteilung spielt eine wichtige Rolle. Für Zufallsgeneratoren ist es entscheidend, dass sie auf dem Intervall $[0, 1]$ gleichverteilte Zahlen liefern.

4.6.2 Die Exponentialverteilung

Beschreibung

Eine Exponentialverteilung wird oft für die Lebensdauer von Verschleißteilen, Abfertigungszeiten von Kunden unterstellt.

Die **Exponentialverteilung mit Parameter** $\lambda > 0$ ist folgenderweise definiert:

Bezeichnung	$X \sim \text{Exp}(\lambda)$
f	$f(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x),$
F	$F(x) = (1 - e^{-\lambda x}) 1_{[0, \infty)}(x)$
Erwartungswert	$\frac{1}{\lambda}$
Varianz	$\frac{1}{\lambda^2}$

4.6.3 Die Normalverteilung

Dies ist die mit Abstand wichtigste Verteilungsfunktion. Sie wurde von Gauß eingesetzt und hat zahlreiche Anwendungen von denen wir einige kennen lernen werden.

Die **Normal-Verteilung mit Parametern** $\sigma > 0$, μ ist folgenderweise definiert:

Bezeichnung	$X \sim N(\mu, \sigma^2)$
Dichte	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), x \in \mathbb{R}$
F	nicht in geschlossener Form darstellbar
Erwartungswert	μ
Varianz	σ^2

Die Normalverteilung wird auch als **Gauß-Verteilung** bezeichnet und die Dichtefunktion als Gaußsche Glockenkurve. Für $\mu = 0$ und $\sigma = 1$ erhält man die **Standardnormalverteilung** $N(0, 1)$. Die Verteilungsfunktion der Standardnormalverteilung wird mit Φ bezeichnet, die Dichte mit φ .

Satz 4.6.1 (Eigenschaften der Normalverteilung) Sei $X \sim N(\mu, \sigma^2)$ mit $\sigma > 0$.

(a) Die Dichtefunktion ist symmetrisch zu μ , d.h.

$$f(\mu - x) = f(\mu + x).$$

Die Dichtefunktion besitzt ein globales Maximum an der Stelle μ und Wendepunkte an den Stellen $\mu - \sigma$, $\mu + \sigma$.

(b) Für alle $a, b \in \mathbb{R}$ gilt $a + bX \sim N(a + \mu, b^2\sigma^2)$ und $\frac{X - \mu}{\sigma} \sim N(0, 1)$.

(c) Für die Verteilungsfunktion F von X gilt:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (4.18)$$

$$F(\mu + t \cdot \sigma) = \Phi(t). \quad (4.19)$$

(d) Für die Standardnormalverteilung gilt für alle $x \in \mathbb{R}$:

$$\Phi(-x) = 1 - \Phi(x) \quad (4.20)$$

Die Standardnormalverteilung $N(0, 1)$ ist für $x \geq 0$ tabelliert. Wegen (4.20) ist das ausreichend um beliebige Werte der Verteilungsfunktion einer $N(\mu, \sigma)$ verteilten Zufallsvariablen zu berechnen.

Mit der Normalverteilung kann man auch Intervalle $[\mu - c\sigma, \mu + c\sigma]$ angeben, in denen mit einer vorgegebenen Wahrscheinlichkeit die Ergebnisse liegen werden. Es gilt allgemein für eine Zufallsvariable X mit $X \sim N(\mu, \sigma^2)$

$$\mathbb{P}(\mu - 1,96 \cdot \sigma \leq X \leq \mu + 1,96 \cdot \sigma) = 95\% \quad (4.21)$$

$$\mathbb{P}(\mu - 2,58 \cdot \sigma \leq X \leq \mu + 2,58 \cdot \sigma) = 99\% \quad (4.22)$$

$$\mathbb{P}(\mu - 3,29 \cdot \sigma \leq X \leq \mu + 3,29 \cdot \sigma) = 99,9\%. \quad (4.23)$$

4.6.4 Die Gamma-Verteilung

Beschreibung

Die Gamma-Verteilung ist eine Verallgemeinerung der Exponentialverteilung.

Die **Gamma-Verteilung mit Parameter** $a, \lambda > 0$ ist folgenderweise definiert:

Bezeichnung	$X \sim \Gamma(a, \lambda)$
f	$f(x) = \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)}, x \geq 0$
F Erwartungswert	nicht in geschlossener Form darstellbar $\frac{a}{\lambda}$
Varianz	$\frac{a}{\lambda^2}$

Die Definition der Gamma-Funktion findet man am Ende dieses Kapitels auf Seite 61.

Für $a = 1$ ergibt sich die Exponential-Verteilung mit Parameter λ .

Weitere Verteilungsfunktionen, die in der Statistik eingesetzt werden (t , χ^2 , F -Verteilung) werden wir dort vorstellen.

4.7 Erzeugende Funktionen

Mit erzeugenden Funktionen sind manche Rechnungen leichter durchzuführen.

4.7.1 Die wahrscheinlichkeitserzeugende Funktion

Es sei $X : \Omega \longrightarrow \mathbb{N}_0$ eine diskrete Zufallsvariable. Dann heißt

$$EF_X(t) := \mathbb{E}(t^X)$$

die **wahrscheinlichkeitserzeugende Funktion** (Englisch: probability generating function p.g.f.) für diejenigen $t \in \mathbb{R}$ für die das Integral existiert. Es gilt

$$\mathbb{E}(t^X) = \sum_{n=0}^{\infty} t^n \mathbb{P}(X = n).$$

Die Reihe ist absolut konvergent für $|t| \leq 1$ wegen $\sum_{n=0}^{\infty} \mathbb{P}(X = n) = 1$. Somit ist ihr Konvergenzradius größer oder gleich 1.

Satz 4.7.1 Die Funktion $EF : (-1, 1) \longrightarrow \mathbb{R}$ ist unendlich oft differenzierbar und es gilt

$$\forall n \in \mathbb{N}_0 : EF^{(n)}(0) = \frac{\mathbb{P}(X = n)}{n!}$$

Beweis EF ist eine Potenzreihe, die auf $(-1, 1)$ absolut konvergiert. Damit ist im Entwicklungspunkt 0 EF gleich seiner Taylor-Reihe. $\square$

Korollar 4.7.2 Sind X, Y Zufallsvariablen mit $EF_X = EF_Y$ auf $[-1, 1]$ dann folgt für die Verteilungsfunktionen $F_X = F_Y$.

Beweis Folgt aus Satz 4.7.1. $\square$

Beispiel 4.7.3 Man bestimme die Wahrscheinlichkeitserzeugende Funktion für die Binomial-, Poisson- und negativ Binomialverteilung.

Lösung

B(n,p)

$$\begin{aligned} \sum_{k=0}^n t^k \mathbb{P}(X = k) &= \sum_{k=0}^n t^k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (tp)^k (1-p)^{n-k} = \\ &= (tp + (1-p))^n. \end{aligned}$$

Poi(λ)

$$\begin{aligned} \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k) &= \sum_{k=0}^{\infty} t^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda)^k}{k!} = \\ &= e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)} \end{aligned}$$

NB(m,p) Übung.

$\square$

4.7.2 Momentenerzeugende Funktion

Für eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ sei

$$MEF_X(t) := \mathbb{E}(e^{tX}) \quad (4.24)$$

die momentenerzeugende Funktion (Englisch: moment generating function m.g.f.) für diejenigen $t \in \mathbb{R}$ für die das Integral existiert. Es gilt stets $MEF_X(0) = 1$.

Satz 4.7.4 Angenommen es gibt $\varepsilon > 0$ so, dass $MEF_X(t)$ für $t \in (-\varepsilon, \varepsilon)$ definiert. Dann

(a) existieren alle Momente von X .

(b) MEF ist unendlich oft differenzierbar

(c) $\mathbb{E}(X^n) = MEF^{(n)}(0)$.

Begründung Der Beweis verwendet wesentlich den Satz von der dominierten Konvergenz. Man zeigt, dass dann in (4.24) Differentiation nach t und Integration vertauscht werden können. Damit folgt für $t \in (-\varepsilon, \varepsilon)$

$$MEF'_X(t) = \frac{d}{dt} \int_{\mathbb{R}} e^{tx} d\mathbb{P}_X(x) = \int_{\mathbb{R}} \frac{d}{dt} (e^{tx}) d\mathbb{P}_X(x) = \int_{\mathbb{R}} x (e^{tx}) d\mathbb{P}_X(x).$$

Setzt man $t = 0$ ein folgt $MEF'_X(0) = \mathbb{E}(X)$. Differenziert man weiter, folgt (c). $\square$

Beispiel 4.7.5 Man bestimme die momentenerzeugende Funktion für die Binomial-, Poisson-, negativ Binomial- und Exponentialverteilung.

Lösung

B(n,p)

$$\begin{aligned} \sum_{k=0}^n e^{tk} \mathbb{P}(X = k) &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^t p)^k (1-p)^{n-k} = \\ &= (e^t p + (1-p))^n. \end{aligned}$$

Poi(λ)

$$\begin{aligned} \sum_{k=0}^{\infty} e^{tk} \mathbb{P}(X = k) &= \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

NB(m,p) Übung.

Exp(λ)

$$E(e^{tX}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}.$$

$MEF_{\text{Exp}(\lambda)}(t)$ ist also für $t \in (-\infty, \lambda)$ definiert.

$\square$

Hilfreich ist das folgende Ergebnis:

Satz 4.7.6 Sind X, Y Zufallsvariablen mit $MEF_X = MEF_Y$ in einer Umgebung der 0, dann folgt für die Verteilungsfunktionen $F_X = F_Y$.

4.7.3 Charakteristische Funktion

Es sei X eine reellwertige Zufallsvariable. Dann heißt

$$CF_X(t) := \mathbb{E}(e^{itX})$$

die **charakteristische Funktion** von X .

Die Definition der charakteristischen Funktion verwendet den Erwartungswert einer komplexwertigen Zufallsvariablen. Sei für $Z = U + iV$, wobei U und V reellwertige Zufallsvariablen sind,

$$\mathbb{E}(Z) := \mathbb{E}(U) + i\mathbb{E}(V).$$

Die charakteristische Funktion ist für alle $t \in \mathbb{R}$ definiert, denn

$$\forall t, x \in \mathbb{R} : |e^{itx}| = 1.$$

und ist $1_{\mathbb{R}}$ integrierbar.

Satz 4.7.7 *Sei X eine Zufallsvariable. Ist das k -te Moment von X endlich, so ist CF_X k -mal differenzierbar und*

$$\frac{d^k}{dt^k} CF_X(0) = i^k \mathbb{E}(X^k).$$

Für den Beweis benötigen wir folgendes

Lemma 4.7.8 *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : (a, b) \times \Omega \longrightarrow \mathbb{R}$ messbar mit folgenden Eigenschaften:*

- $f(\cdot, \omega)$ ist für alle $\omega \in \Omega$ differenzierbar und es gibt $h : \Omega \longrightarrow \mathbb{R}$ integrierbar mit

$$\forall t \in (a, b) : |f_t(t, \cdot)| \leq h(\omega).$$

- Für alle $t \in (a, b)$ ist $f(t, \cdot)$ integrierbar.

Dann ist die Funktion

$$F : (a, b) \longrightarrow \mathbb{R}, \quad F(t) := \int_{\Omega} f(t, \omega) d\mu$$

differenzierbar und

$$F'(t) = \int_{\Omega} f_t(t, \omega) d\mu$$

Beweis Sei $t_0 \in (a, b)$ und eine Folge $(t_n)_{n \in \mathbb{N}}$ mit $t_n \rightarrow t_0$, $t_n \neq 0$. Die Funktionen $g_n : \Omega \rightarrow \mathbb{R}$,

$$g_n(\omega) := \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0}$$

sind messbar, $g_n \rightarrow f_t(t_0, \cdot)$ auf Ω und mit dem Mittelwertsatz folgt

$$|g_n(\omega)| = |f_t(\xi_n, \omega)| \leq h(\omega).$$

Mit dem Satz von der dominierten Konvergenz (Satz 2.6.14, S. 21) gilt

$$\lim_{n \rightarrow \infty} \int g_n d\mu = \int f_t(t_0, \cdot) d\mu$$

und die Behauptung folgt wegen

$$\frac{F(t_n) - F(t_0)}{t_n - t_0} = \int g_n d\mu$$

□

Beweis von Satz 4.7.7: Sei $u, v : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $u(t, x) := \cos(tx)$, $v(t, x) := \sin(tx)$. Dann sind u, v nach t differenzierbar,

$$\begin{aligned} |u(t, x)| &\leq 1 & |v(t, x)| &\leq 1 \\ |u_t(t, x)| &= |-x \sin(tx)| \leq |x| & |v_t(t, x)| &= |x \cos(tx)| \leq |x|. \end{aligned}$$

Existiert das erste Moment von X , dann ist $\mathbb{R} \ni x \mapsto |x|$ $\mathbb{P}_X$ -integrierbar. Somit ist CF laut Lemma 4.7.8 differenzierbar und man kann Integration und Differentiation vertauschen, also

$$CF'(t) = \int ixe^{itx} d\mathbb{P}_X(x).$$

Analog zeigt man die Behauptung für die höheren Momente. □

Beispiel 4.7.9 Man bestimme die charakteristische Funktion für Binomial-, Poisson- und Exponentialverteilung.

Lösung

B(n,p)

$$\begin{aligned} \sum_{k=0}^n e^{itk} \mathbb{P}(X = k) &= \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^t p)^k (1-p)^{n-k} = \\ &= (e^{it} p + (1-p))^n. \end{aligned}$$

Poi(λ)

$$\begin{aligned} \sum_{k=0}^{\infty} e^{itk} \mathbb{P}(X = k) &= \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = \\ &= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)} \end{aligned}$$

Exp(λ)

$$E(e^{itX}) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - it}.$$

□

Weitere Eigenschaften von erzeugenden Funktionen werden im Zusammenhang mit unabhängigen Zufallsvariablen behandelt.

4.8 Mehrdimensionale Zufallsvariablen

Was wir bisher über Zufallsvariablen wissen, lässt sich auf mehrdimensionale Zufallsvariablen übertragen. Wir betrachten einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 4.8.1 (a) Eine Abbildung $X = (X_1, \dots, X_n) : \Omega \longrightarrow \mathbb{R}^n$ heißt ***n*-dimensionale Zufallsvariable** wenn X $\mathcal{A} - \mathcal{B}^n$ messbar ist.

(b) $X = (X_1, \dots, X_n)$ sei eine *n*-dimensionale Zufallsvariable. Die Funktion

$$\begin{aligned} F : \mathbb{R}^n &\longrightarrow [0, 1] \\ (x_1, \dots, x_n) &\longmapsto \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

heißt die ***gemeinsame Verteilungsfunktion*** der Zufallsvariablen $X_1, \dots, X_n$. Das Bildmaß $\mathbb{P}_X$ auf $(\mathbb{R}^n, \mathcal{B}^n)$ wird für $B \in \mathcal{B}^n$ definiert:

$$\begin{aligned} \mathbb{P}_X(B) &= \mathbb{P}(\{\Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\}) \\ &= \mathbb{P}((X_1, \dots, X_n) \in B). \end{aligned}$$

$\mathbb{P}_X$ heißt ***gemeinsame Verteilung von $X_1, \dots, X_n$***

(c) Eine *n*-dimensionale Zufallsvariable heißt ***diskret***, wenn sie nur endlich viele oder abzählbar viele Werte annimmt. In diesem Fall heißt

$$f(x_1, \dots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

die ***Wahrscheinlichkeitsfunktion***.

- (d) Eine n -dimensionale Zufallsvariable heißt **stetig (verteilt)**, wenn $\mathbb{P}_X$ eine λ^n Dichte $f : \mathbb{R}^n \longrightarrow [0, \infty)$ besitzt, also

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_n \dots dy_1$$

Dann gilt auch

$$\mathbb{P}_X(B) = \int_B f(y_1, \dots, y_n) dy_n \dots dy_1 := \int_B f d\lambda^n.$$

Ist eine Zufallsvariable $X = (X_1, \dots, X_n) : \Omega \longrightarrow \mathbb{R}^n$ gegeben, dann werden damit folgende Maße induziert:

- Für jedes X_i das Wahrscheinlichkeitsmaß $\mathbb{P}_{X_i}$ auf $(\mathbb{R}, \mathcal{B}^1)$. Es wird mit der Verteilungsfunktion F_i von X_i und Fortsetzung von

$$\mathbb{P}_{X_i}((a, b]) := F_i(b) - F_i(a), \quad a, b \in \mathbb{R}, a \leq b$$

auf $\mathcal{B}^1$ konstruiert (siehe auch Satz 4.2.2, S. 35).

Die $\mathbb{P}_{X_i}$ bzw. F_{X_i} heißen **Randverteilungen von X** .

- Das Produktmaß $\bigotimes_{i=1}^n \mathbb{P}_{X_i} := \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ auf $(\mathbb{R}^n, \mathcal{B}^n)$ als Fortsetzung von

$$\bigotimes_{i=1}^n \mathbb{P}_{X_i} \left(\prod_{i=1}^n (a_i, b_i] \right) := \prod_{i=1}^n \mathbb{P}_{X_i}(a_i, b_i] = \prod_{i=1}^n F_i(b_i) - F_i(a_i), \quad a_i \leq b_i$$

auf $\mathcal{B}^n$ (siehe Unterabschnitt 2.8.1, S. 22)

- Das Bildmaß $\mathbb{P}_X$ auf $(\mathbb{R}^n, \mathcal{B}^n)$.

Es gilt stets

$$F_i(x) = \mathbb{P}(X_i \leq x) = \mathbb{P}(X_i \leq x, \cap_{j \neq i} \{X_j \in \mathbb{R}\}) = \mathbb{P}_X(X_i \leq x, \cap_{j \neq i} \{X_j \in \mathbb{R}\}).$$

Unter bestimmten Voraussetzungen an X gilt auch $\mathbb{P}_X = \bigotimes_{i=1}^n \mathbb{P}_{X_i}$ (s. nächstes Kapitel).

4.8.1 Kovarianz und Korrelation

Definition 4.8.2 Seien X und Y zwei Zufallsvariablen, für die die Varianzen existieren. Die Größe

$$\text{Cov}(X, Y) := E\left((X - E(X)) \cdot (Y - E(Y))\right)$$

heißt **Kovarianz** von X und Y . Wenn $\text{Var}(X) \neq 0 \neq \text{Var}(Y)$ heißt

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

Korrelationskoeffizient.

Satz 4.8.3 (Eigenschaften von Kovarianz und Korrelation) Seien X und Y Zufallsvariablen wie in der Definition 4.8.2.

- (a) Die Kovarianz ist bilinear und symmetrisch.
- (b) $\rho(X, Y) \in [-1, 1]$ misst den Zusammenhang zwischen X und Y . Falls $\rho(X, Y) = 1$ heißen X, Y **vollständig korreliert**, bei $\rho(X, Y) = -1$ **entgegengesetzt korreliert** und bei $\rho(X, Y) = 0$ **unkorreliert**.
- (c)

$$\begin{aligned} \text{Cov}(X, Y) &= E(X \cdot Y) - E(X) \cdot E(Y) \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ \text{Var}(X) &= \text{Cov}(X, X). \end{aligned} \quad (4.25)$$

Sind X, Y unkorreliert, dann gilt $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

- (d) $\rho(X, Y) = \pm 1$ genau dann, wenn reelle Zahlen a, b existieren mit $X = aY + b$ f.s.

Beweis Hier verwenden wir die Eigenschaften des Hilbertraums $L^2(\Omega)$

(a) folgt aus der Eigenschaft des Skalarprodukts auf $L^2(\Omega)$, denn

$$\text{Cov}(X, Y) = \langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle.$$

(b) Wegen (2.16) folgt

$$\begin{aligned} |\text{Cov}(X, Y)| &= |\langle X - \mathbb{E}(X), Y - \mathbb{E}(Y) \rangle| \leq \|X - \mathbb{E}(X)\|_2 \|Y - \mathbb{E}(Y)\|_2 \\ &= \sqrt{\text{Var}(X) \text{Var}(Y)}. \end{aligned}$$

(c)

$$\begin{aligned}
& \mathbb{E}((X - E(X))(Y - E(Y))) \\
&= \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y)) \\
&= \mathbb{E}(XY) - \mathbb{E}(X\mathbb{E}(Y)) - \mathbb{E}(Y\mathbb{E}(X)) + \mathbb{E}(XY) \\
&= \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Die übrigen Gleichungen folgen analog.

(d) Gilt $Y = aX + b$ f.s. dann

$$\begin{aligned}
\text{Cov}(aX + b, X) &= a\text{Cov}(X, X) + \text{Cov}(b, X) = a \cdot \text{Var}(X) \\
\text{Var}(Y) &= \text{Var}(aX + b) = \text{Var}(aX) = a^2\text{Var}(X)
\end{aligned}$$

also

$$\frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{a\text{Var}(X)}{|a|\text{Var}(X)} = \frac{a}{|a|}.$$

Die Umkehrung beweisen wir später im Rahmen der linearen Regression. $\square$

Warnung

Der Korrelationskoeffizient

- ist ein relatives Maß
- schwer zu interpretieren (je näher an ± 1 , desto stärker sind die Datenpunkte um eine Gerade gehäuft, über die Steigung der Geraden sagt ρ nichts aus)
- zeigt nur lineare Assoziation, keine Kausalbeziehung.

4.8.2 Die n -dimensionale Normalverteilung

Definition 4.8.4 Es seien $X_1, \dots, X_n$ Zufallsvariablen, $X := (X_1, \dots, X_n)^T$. Dann heißen (falls existent)

$$\mathbb{E}(X) := (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))^T$$

und

$$\mathbb{V}(X) := (\text{Cov}(X_i, X_j))_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$$

Erwartungswert bzw. **Kovarianzmatrix** von X .

Die Kovarianzmatrix $\mathbb{V}(X)$ von X ist symmetrisch und positiv semidefinit.

Definition 4.8.5 Sei $\Sigma \in \mathbb{R}^{n \times n}$ eine symmetrische, positiv definite Matrix und $\mu \in \mathbb{R}^n$. Eine Zufallsvariable, $X := (X_1, \dots, X_n)^T : \Omega \longrightarrow \mathbb{R}^n$ heißt ***n-dimensional normalverteilt***, wenn die gemeinsame Verteilung der X_i die Dichte

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} \exp \left(-\frac{1}{2} \langle x - \mu, \Sigma^{-1}(x - \mu) \rangle \right)$$

besitzt.

Bezeichnung $X \sim N_n(\mu, \Sigma)$

Wir zitieren zunächst zwei grundlegende Ergebnisse:

Satz 4.8.6 Sei $X \sim N_n(\mu, \Sigma)$ für μ, Σ wie in der Definition 4.8.5. Dann gilt $\mathbb{E}(X) = \mu$ und $\mathbb{V}(X) = \Sigma$.

Satz 4.8.7 Sei $X : \Omega \longrightarrow \mathbb{R}^n$ eine Zufallsvariable. X ist genau dann *n-dimensional normalverteilt*, wenn für alle $a \in \mathbb{R}^n, a \neq 0$ die reellwertige Zufallsvariable $\langle a, X \rangle$ normalverteilt ist.

Linearkombinationen normalverteilter (mehrdimensionaler) Zufallsvariablen bleiben (mehrdimensional) normalverteilt.

Satz 4.8.8 Sei $X \sim N_n(\mu, \Sigma)$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Dann gilt

$$AX + b \sim N_m(A\mu + b, A\Sigma A^T)$$

4.9 Hilfsaussagen

4.9.1 Konvexe Funktionen

Definition 4.9.1 Sei $-\infty \leq a < b \leq \infty$. Eine Funktion $f : (a, b) \longrightarrow \mathbb{R}$ heißt ***konvex***, wenn für alle $x, y \in I$, $x < y$ und alle $\lambda \in [0, 1]$ gilt

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Eine Funktion ist genau dann konvex, wenn der Graph von f unter der Verbindungsstrecke von $(x, f(x))$ nach $(y, f(y))$ liegt.

Satz 4.9.2 Sei $f : (a, b) \longrightarrow \mathbb{R}$ konvex. Dann ist f stetig, die links- und rechtsseitig Ableitungen f'_-, f'_+ existieren und $f'_- \leq f'_+$. Für $x \in (a, b)$ und $\beta \in [f'_-(x), f'_+(x)]$ gilt

$$\forall y \in (a, b) : f(y) \geq f(x) + \beta(y - x). \quad (4.26)$$

(4.26) besagt, dass die Geraden im Punkt $(x, f(x))$ unterhalb des Graphen von f verlaufen, wenn ihre Steigung zwischen der links- und rechtsseitigen Ableitung liegt.

4.9.2 Die Γ -Funktion

Für $x > 0$ setzt man

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt.$$

Es handelt sich um die Fortsetzung von $n!$. Es gilt nämlich:

(a) $\forall x > 0 : \Gamma(x+1) = x\Gamma(x)!$

(b) $\forall n \in \mathbb{N} : \Gamma(n+1) = n!$

Es besteht folgender Zusammenhang zwischen der Beta-Funktion

$$B(\alpha, \beta) := \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt, \quad \alpha, \beta > 0$$

und der Γ -Funktion:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

5 Unabhängigkeit und bedingte Wahrscheinlichkeiten

$(\Omega, \mathcal{A}, \mathbb{P})$ sei stets ein Wahrscheinlichkeitsraum.

5.1 Bedingte Wahrscheinlichkeit

Nehmen wir an, dass wir zwei Würfel werfen. Es gibt 36 mögliche Ergebnisse, jedes tritt mit einer Wahrscheinlichkeit von $1/36$ ein. Wir wollen weiter annehmen, dass der erste Wurf eine 4 ergibt. Wie groß ist nun die Wahrscheinlichkeit, dass die Augensumme 10 ist?

Um diese Wahrscheinlichkeit zu berechnen gehen wir wie folgt vor: Nachdem bekannt ist, dass die erste Zahl eine 4 ist, sind nur noch die Ergebnisse $(4,1)$, $(4,2)$, $(4,3)$, $(4,5)$, $(4,5)$ und $(4,6)$ möglich. Jedes dieser Ergebnisse tritt mit einer Wahrscheinlichkeit von $1/6$ ein. Da $(4,6)$ als einziges Ergebnis die Augensumme 10 liefert, ist die gesuchte Wahrscheinlichkeit $1/6$.

Obiges Ereignis kann man formulieren als „Augensumme 10 unter der Bedingung, dass die erste Zahl eine 4 ist“, die dazugehörige Wahrscheinlichkeit

$$\mathbb{P}(\text{Augensumme} = 10 \mid \text{Erste Zahl ist } 4).$$

Zur Motivation der Definition der bedingten Wahrscheinlichkeiten argumentieren wir erneut mit den relativen Häufigkeiten.

Wir wiederholen ein Zufallsexperiment (zweimaliger Wurf eines Würfels) n -mal unter den gleichen Bedingungen und untersuchen das Eintreten der Ereignisse A (Augensumme = 10) und B (erste Zahl = 4). Das Ereignis A bzw. B trete n_A (ca. $n \cdot 3/36$) bzw. n_B -mal (ca. $n \cdot 1/6$) auf, das Ereignis $A \cap B$ trete $n_{A \cap B}$ (ca. $n \cdot 1/36$) auf. Die Größe

$$\frac{n_{A \cap B}}{n_B}$$

ist eine relative Häufigkeit: Betrachtet werden nur die Ergebnisse bei denen B (erste Zahl 4) eintritt, und davon diejenigen Ergebnisse, bei denen auch A (Augensumme = 10) eingetreten ist. Mit anderen Worten: Wie oft ist A eingetreten unter der Bedingung, dass B schon eingetreten ist (relative Häufigkeit von A unter der Bedingung B). Dies motiviert die folgende Definition:

Definition 5.1.1 (bedingte Wahrscheinlichkeit) Seien $A, B \in \mathcal{A}$, $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (5.1)$$

die *bedingte Wahrscheinlichkeit* von A unter der Bedingung B .

In unserem Anfangsbeispiel haben wir also

$$\begin{aligned} A &= \{(4, 6), (5, 5), (6, 4)\} \\ B &= \{(4, 1), (4, 2), (4, 3), (4, 5), (4, 6)\} \\ A \cap B &= \{(4, 6)\} \\ \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/36}{6/36} = \frac{1}{6}. \end{aligned}$$

Beispiel 5.1.2 *Wie groß ist die Wahrscheinlichkeit, dass wir eine 2 oder eine 4 gewürfelt haben, falls das Würfelergebnis gerade ist?*

Lösung Wendet man (5.1) an, ergibt sich

$$\begin{aligned} \mathbb{P}(\text{Ergebnis gerade}) &= \mathbb{P}(\{2, 4, 6\}) = 0,5 \\ \mathbb{P}(\text{Ergebnis gerade und Ergebnis 2 oder 4}) &= \mathbb{P}(\{2, 4\}) = \frac{2}{6} \\ \mathbb{P}(\text{Ergebnis 2 oder 4} | \text{Ergebnis gerade}) &= \frac{2/6}{1/2} = \frac{2}{3}. \end{aligned}$$

Man kann aber auch folgendermaßen argumentieren: Die Ergebnismenge ist $\Omega = \{2, 4, 6\}$, die Ergebnisse, für die wir uns interessieren sind $\{2, 4\}$. Da Laplace Wahrscheinlichkeiten vorliegen (Definition 3.2.3) können wir die übliche Formel anwenden und erhalten

$$\frac{|\{2, 4\}|}{|\{2, 4, 6\}|} = \frac{2}{3}.$$

□

Auch für bedingte Wahrscheinlichkeiten gelten die Axiome und die daraus abgeleiteten Regeln.

Wir wollen nun eine oft nützliche Technik zur Berechnung der Wahrscheinlichkeit des Durchschnitts von Ereignissen (bzw. von mehrstufigen Zufallsexperimenten) kennen lernen.

Beispiel 5.1.3 *Unterscheidet man jeweils ob der Absatz eines Produktes gegenüber der Vorperiode steigt oder nicht steigt, so sind für die kommenden drei Zeitperioden acht Verläufe denkbar. Die Marketingabteilung schätzt folgende Wahrscheinlichkeiten (+ steht für Absatzsteigerung, - für Absatzverringerung:*

	Bedingte Wahrscheinlichkeiten							
1. Jahr	+				-			
	1/2				1/2			
2. Jahr	+		-		+		-	
	2/3		1/3		1/3		2/3	
3. Jahr	+	-	+	-	+	-	+	-
	1/4	3/4	1/2	1/2	1/3	2/3	1/5	4/5

Wie berechnet man die Wahrscheinlichkeit des Ereignisses B „in jeder Periode steigt der Absatz“.

Lösung $\mathbb{P}(B) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4}$ □

Grundlage der Lösung ist die Formel

$$\begin{aligned} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_m) \\ = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \mathbb{P}(A_m|A_1 \cap \dots \cap A_{m-1}) \cdot \dots \end{aligned} \quad (5.2)$$

Diese Formel besagt, dass man zur Berechnung eines Ereignisses die (den) dazugehörige(n) Pfad(e) im Wahrscheinlichkeitsbaum verfolgt und die jeweiligen bedingten Wahrscheinlichkeiten miteinander multipliziert.

5.2 Unabhängige Ereignisse

Nun kommen wir zur Definition der Unabhängigkeit. Bei unserem vorangegangenen „Würfelbeispiel“ haben wir gesehen, dass die Wahrscheinlichkeit unseres Ereignisses A , d.h. eine 2 oder eine 4 zu werfen, zwei Drittel ist, falls eine gerade Zahl (Ereignis B) geworfen wird. Falls wir nicht das bedingte Ereignis betrachten, sondern einfach die Wahrscheinlichkeit des Ereignisses A ausrechnen wollten, so ergibt sich ein Drittel. Die Wahrscheinlichkeit des Ereignisses A ist also nicht unabhängig vom Ereignis B , A ist unter der Bedingung B wahrscheinlicher als ohne die Bedingung B :

$$\frac{1}{3} = \mathbb{P}(A) < \mathbb{P}(A|B) = \frac{2}{3}.$$

Definition 5.2.1 (a) Zwei Ereignisse A und B heißen **unabhängig**, falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

(b) Eine **Familie** $(A_i)_{i \in I}$ von Ereignissen heißt **unabhängig**, wenn für jede endliche Teilfamilie $(A_j)_{j \in J}$ (also $J \subset I$ endlich) gilt:

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Falls die Ereignisse A und B unabhängig sind mit $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$ gilt:

$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A) \\ \mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(B)\end{aligned}$$

Wir können also praktisch überprüfen, ob 2 Ereignisse unabhängig sind, indem wir die Beziehung in der Definition der Unabhängigkeit nachrechnen.

Beispiel 5.2.2 *Wir werfen eine Münze zweimal. Ist die Wahrscheinlichkeit, dass wir genau einmal „Zahl“ werfen davon abhängig ob der erste Wurf „Zahl“ war?*

Lösung Sei A das Ereignis „1. Wurf liefert Zahl“ und B das Ereignis „bei zwei Würfeln genau einmal Zahl gewürfelt“. Es gilt also

$$\begin{aligned}\Omega &= \{(Z, W); (W, Z); (W, W); (Z, Z)\} \\ A &= \{(Z, W); (Z, Z)\} \\ B &= \{(Z, W); (W, Z)\} \\ A \cap B &= \{(Z, W)\}\end{aligned}$$

Somit gilt: $\mathbb{P}(A) = 0,5$, $\mathbb{P}(B) = 0,5$, $\mathbb{P}(A \cap B) = 0,25$. Wegen $\mathbb{P}(A) \cdot \mathbb{P}(B) = 0,25 = \mathbb{P}(A \cap B)$ sind die beiden Ereignisse voneinander unabhängig. $\square$

Beispiel 5.2.3 *Betrachte beim zweifachen Münzwurf die Ereignisse*

$$\begin{aligned}A_1 &= \{1. \text{ Wurf Wappen}\} = \{(W, W), (W, Z)\} \\ A_2 &= \{2. \text{ Wurf Wappen}\} = \{(W, W), (Z, W)\} \\ A_3 &= \{1. \text{ und } 2. \text{ Wurf gleich}\} = \{(W, W), (Z, Z)\}\end{aligned}$$

Wegen $\mathbb{P}(A_i \cap A_j) = 1/4 = \mathbb{P}(A_i)\mathbb{P}(A_j)$ für $i \neq j$ sind die Ereignisse paarweise unabhängig, aber wegen

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\{(W, W)\}) = 1/4 \neq 1/8 = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

sind A_1 , A_2 , A_3 nicht unabhängig.

5.3 Die Formel von Bayes

Beispiel 5.3.1 *In der Bevölkerung ist von einer Virusinfektion bekannt (z.B. HIV), dass 10.000 von 100.000.000 infiziert sind. Es wurde ein Test mit folgenden Merkmalen entwickelt:*

Person	Test positiv	Test negativ
infiziert	99 %	1 %
gesund	2 %	98 %

Das bedeutet, dass bei einem Prozent der infizierten Patienten der Test negativ ausfällt, also das falsche Ergebnis liefert und bei zwei Prozent der Gesunden der Test positiv ausfällt.

Eine zufällig ausgewählte Person habe ein positives Testergebnis. Wie groß ist die Wahrscheinlichkeit, dass die Testperson tatsächlich infiziert ist? Schätzen Sie zunächst das Ergebnis nach Gefühl (gehen Sie qualitativ vor) und lösen Sie anschließend die Aufgabe mit einem Wahrscheinlichkeitsbaum.

Wir interessieren uns für die Wahrscheinlichkeit

$$\mathbb{P}(\text{Person ist infiziert} | \text{Test positiv})$$

Bekannt sind allerdings nur

$$\begin{aligned}\mathbb{P}(\text{Test positiv} | \text{Patient ist infiziert}) &= 0,99 \\ \mathbb{P}(\text{Person ist infiziert}) &= 0,0001.\end{aligned}$$

Gibt es eine Möglichkeit aus den gegebenen Informationen das gewünschte Ergebnis zu erhalten? Die Lösung dieses Problems liefert die Formel von Bayes.

Satz 5.3.2 *Seien $B_i \in \mathcal{A}$, $i \in I \subset \mathbb{N}$ disjunkt mit $\bigcup_{i \in I} B_i = \Omega$ und $\mathbb{P}(B_i) > 0$ für alle $i \in I$. Sei $A \in \mathcal{A}$. Es gilt:*

Satz von der totalen Wahrscheinlichkeit

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i). \quad (5.3)$$

Formel von Bayes *Gilt zusätzlich $\mathbb{P}(A) > 0$, dann folgt für alle $i \in I$*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

Beweis

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{i \in I} (A \cap B_i)\right) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}(A \cap B_i) \frac{\mathbb{P}(B_i)}{\mathbb{P}(B_i)} \\
&= \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i), \\
\mathbb{P}(B_i|A) &= \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} \cdot \frac{\mathbb{P}(B_i)}{\mathbb{P}(B_i)} = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)}.
\end{aligned}$$

□

Kehren wir zu unserem Beispiel zurück. Dann wählen wir:

B_1 : “Person ist infiziert“

B_2 : “Person ist gesund“

A : “Test ist positiv“

Wir erhalten mit der Formel von Bayes

$$\begin{aligned}
&\mathbb{P}(\text{Person ist infiziert} | \text{Test positiv}) \\
&= \mathbb{P}(B_1|A) \\
&= \frac{\mathbb{P}(A|B_1) \cdot \mathbb{P}(B_1)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(\text{Test positiv} | \text{Person ist infiziert}) \cdot \mathbb{P}(\text{Person ist infiziert})}{\mathbb{P}(\text{Test positiv})} \\
&= \frac{0,99 \cdot 0,0001}{\mathbb{P}(\text{Test positiv})}.
\end{aligned}$$

Die noch unbekannte Wahrscheinlichkeit $\mathbb{P}(\text{Test positiv})$ erhalten wir mit (5.3):

$$\begin{aligned}
\mathbb{P}(\text{Test positiv}) &= \mathbb{P}(\text{Test positiv} | \text{Patient infiziert}) \cdot \mathbb{P}(\text{Patient infiziert}) \\
&\quad + \mathbb{P}(\text{Test positiv} | \text{Patient gesund}) \cdot \mathbb{P}(\text{Patient gesund}) \\
&= 0,99 \cdot 0,0001 + 0,02 \cdot 0,9999 = 0,020097.
\end{aligned}$$

Setzen wir obige Rechnung fort, erhalten wir

$$\mathbb{P}(\text{Person ist infiziert} | \text{Test positiv}) = \frac{0,99 \cdot 0,0001}{0,020097} \approx 0,49\%.$$

5.4 Unabhängige σ -Algebren

Definition 5.4.1 Eine Familie von σ -Algebren $(\mathcal{A}_i)_{i \in I}$, $\mathcal{A}_i \in \mathcal{A}$ heißt **unabhängig**, wenn für jedes $J \subset I$ endlich gilt

$$\forall A_j \in \mathcal{A}_j, j \in J : \mathbb{P} \left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j)$$

Der nächste Satz stellt sicher, dass es ausreicht die Unabhängigkeit von Erzeugern von σ -Algebren zu prüfen.

Satz 5.4.2 Sei $(\mathcal{A}_i)_{i \in I}$, $\mathcal{A}_i \in \mathcal{A}$ eine Familie von σ -Algebren und $\mathcal{E}_i$ seien $\cap$ -stabiler Erzeuger von $\mathcal{A}_i$ für $i \in I$, also $\sigma(\mathcal{E}_i) = \mathcal{A}_i$. Dann gilt:

$$(\mathcal{A}_i)_{i \in I} \text{ ist unabhängig} \iff (\mathcal{E}_i)_{i \in I} \text{ ist unabhängig}$$

Beweis „ $\implies$ “ trivial.

„ $\impliedby$ “ [B-WT], S. 45, 6.4. Korollar. □

5.5 Unabhängige Zufallsvariablen

Unabhängigkeit zweier Zufallsvariablen X, Y bedeutet, dass die Kenntnis der Realisation einer Zufallsvariablen X keine Aussage darüber ermöglicht, welche Werte die andere Zufallsvariable Y annimmt. Fragen nach Unabhängigkeit von Zufallsvariablen kommen in der Praxis sehr oft vor:

- Sind Aktienkurse und Katastrophenereignisse unabhängig voneinander?
- Sind die Renditen von Anleihen und Aktien abhängig?
- Sind in der Versicherung die Konjunktur und die Feuerschäden abhängig?
- Sind der Aufwand für Werbung und der Umsatz eines Produkts abhängig?

5.5.1 Grundlagen

Die nachfolgenden Ergebnisse kann man allgemein für messbare Abbildungen $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ formulieren und wortgleich beweisen.

Definition 5.5.1 Für eine Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ sei

$$\sigma(X) := \sigma(X^{-1}(\mathcal{B}^1)) = X^{-1}(\mathcal{B}^1)$$

die von X erzeugte σ -Algebra .

Definition 5.5.2 Eine Familie von Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$, $i \in I$ heißt **unabhängig**, wenn $(\sigma(X_i))_{i \in I}$ unabhängig sind.

Nun kommen wir zum grundlegenden Ergebnis:

Satz 5.5.3 Seien $X_1, \dots, X_n$ Zufallsvariablen. $X_1, \dots, X_n$ sind genau dann unabhängig, wenn für alle $-\infty \leq a_i \leq b_i < \infty$, $i = 1, \dots, n$ gilt:

$$\mathbb{P} \left(\bigcap_{i=1}^n \{X_i \in (a_i, b_i]\} \right) = \prod_{i=1}^n \mathbb{P}(X_i \in (a_i, b_i]) \quad (5.4)$$

In diesem Fall gilt

$$\mathbb{P}_X = \bigotimes_{i=1}^n \mathbb{P}_{X_i} \quad (5.5)$$

wobei $\bigotimes_{i=1}^n \mathbb{P}_{X_i}$ das Produktmaß der $\mathbb{P}_{X_i}$ ist.

Bemerkung 5.5.4 Im allgemeinen Fall wird im obigen Satz die Menge der Quader durch einen $\cap$ -stabilen Erzeuger $\mathcal{E}'$ von $\mathcal{A}'$ ersetzt.

Beweis „ $\Rightarrow$ “ Laut Definition sind die σ -Algebren $\sigma(X_1), \dots, \sigma(X_n)$ unabhängig, also auch die Mengensysteme $X^{-1}(\mathcal{Q}^n)$.

„ $\Leftarrow$ “ Wegen Korollar 2.5.3, S. 11 gilt $\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{Q}^n))$. Laut Satz 5.4.2 reicht es zu zeigen, dass die $\cap$ -stabilen Erzeuger von $\sigma(X_i)$ unabhängig sind. Für $j = 1, \dots, k$, $k \leq n$ und Intervalle $(a_j, b_j]$ mit $a_j \leq b_j$ gilt

$$\begin{aligned} & \mathbb{P}(X_j \in (a_j, b_j], j = 1, \dots, k) \\ &= \mathbb{P} \left(\left[\bigcap_{j=1}^k X_j^{-1}(a_j, b_j] \right] \cap \left[\bigcap_{j=k+1}^n \Omega \right] \right) \stackrel{(5.4)}{=} \\ &= \prod_{j=1}^k \mathbb{P}(X_j^{-1}(a_j, b_j]) \cdot \prod_{j=k+1}^n \mathbb{P}(\Omega) \\ &= \prod_{j=1}^k \mathbb{P}(X_j^{-1}(a_j, b_j]) = \prod_{j=1}^k \mathbb{P}_{X_j}((a_j, b_j]). \end{aligned}$$

In analoger Weise argumentiert man für alle endlichen Teilmengen $J \subset \{1, \dots, n\}$.

Aus der Gleichung (5.4) folgt mit dem Fortsetzungssatz Satz 2.4.1, S. 9 $\square$

Korollar 5.5.5 Seien X_1, X_2 Zufallsvariablen mit gemeinsamer Verteilungsfunktion F und Randverteilungen F_1, F_2 .

(a) X_1, X_2 sind genau dann unabhängig, wenn

$$\forall (x_1, x_2) \in \mathbb{R}^2 : F(x_1, x_2) = F_1(x_1)F_2(x_2) \quad (5.6)$$

gilt.

(b) Sind X_1, X_2 stetig verteilt mit Dichten f_1, f_2 und unabhängig, dann ist die Zufallsvariable (X_1, X_2) stetig verteilt mit Dichte $f_1 \cdot f_2$.

(c) Sind $X_1, X_2 : \Omega \rightarrow \mathbb{N}$ diskret verteilt mit Wahrscheinlichkeitsfunktionen f_1, f_2 und unabhängig, dann ist die Zufallsvariable (X_1, X_2) diskret verteilt mit Wahrscheinlichkeitsfunktion $f_1 \cdot f_2$.

Beweis (a) Seien X_1, X_2 unabhängig. Wählt man in (5.4) $a_i = -\infty$, dann folgt (5.6). Gilt umgekehrt (5.6), dann folgt durch wiederholte Anwendung von Satz 2.3.1 (a), S. 9

$$\begin{aligned} & \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) \\ &= \mathbb{P}(X_1 \leq b_1, X_2 \leq b_2) - \mathbb{P}(X_1 \leq b_1, X_2 \leq a_2) \\ & \quad - \mathbb{P}(X_1 \leq a_1, X_2 \leq b_2) + \mathbb{P}(X_1 \leq a_1, X_2 \leq a_2). \end{aligned}$$

Wendet man darauf (5.6) an, folgt (5.4).

(b) Seien $x_1, x_2 \in \mathbb{R}$. Dann gilt

$$\begin{aligned} & \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_1(t_1) f_2(t_2) dt_1 dt_2 \\ &= \int_{-\infty}^{x_1} f_1(t_1) \int_{-\infty}^{x_2} f_2(t_2) dt_2 dt_1 = F_1(x_1) F_2(x_2) \stackrel{(a)}{=} F(x_1, x_2) \end{aligned}$$

(c) Klar. □

5.5.2 Produkte von Zufallsvariablen

Satz 5.5.6 Seien X_1, X_2 integrierbare, unabhängige Zufallsvariablen. Es gilt

$$\mathbb{E}(X_1 \cdot X_2) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2).$$

Beweis Es sei $\mathbb{P}_{12}$ die gemeinsame Verteilung von X_1, X_2 und $\mathbb{P}_i$ die Verteilung von X_i . Dann gilt

$$\begin{aligned} \mathbb{E}(X_1 X_2) &= \int_{\Omega} X_1 X_2 d\mathbb{P} = \int_{\mathbb{R}^2} x_1 x_2 d\mathbb{P}_{12} \stackrel{(5.5)}{=} \int_{\mathbb{R}^2} xy d\mathbb{P}_1 \otimes \mathbb{P}_2 \stackrel{(\text{Fubini, S. 23})}{=} \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x_1 x_2 d\mathbb{P}_1(x_1) \right) d\mathbb{P}_2(x_2) = \int_{\mathbb{R}} x_2 \underbrace{\int_{\mathbb{R}} x_1 d\mathbb{P}_1(x_1)}_{\mathbb{E}(X_1)} d\mathbb{P}_2(x_2) \\ &= \mathbb{E}(X_1) \mathbb{E}(X_2). \end{aligned}$$

□

Korollar 5.5.7 Sind X_1, X_2 unabhängige, integrierbare Zufallsvariablen, dann sind sie unkorreliert.

Beweis Es gilt

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$$

und die Behauptung folgt mit Satz 5.5.6. $\square$

Bemerkung 5.5.8 Die Umkehrung von Korollar 5.5.7 gilt im Allgemeinen nicht.

Lösung Sei $\Omega := \{1, 2, 3, 4\}$, $\mathbb{P}(1) = \mathbb{P}(2) = 2/5$, $\mathbb{P}(3) = \mathbb{P}(4) = 1/10$, und die Zufallsvariablen X, Y mit

ω	$X(\omega)$	$Y(\omega)$
1	1	-1
2	-1	1
3	2	2
4	2	-2

X, Y sind korreliert jedoch nicht abhängig: Es gilt $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\text{Cov}(X, Y) = \mathbb{E}(XY) = 0$ aber

$$\mathbb{P}(X = 1, Y = -1) = \frac{2}{5}, \quad \mathbb{P}(X = 1)\mathbb{P}(Y = -1) = \frac{4}{25}.$$

$\square$

5.5.3 Summen von Zufallsvariablen

Lemma 5.5.9 Die Zufallsvariablen X_i , $i = 1, \dots, n$ seien unabhängig und die Varianzen existieren. Dann gilt:

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i). \quad (5.7)$$

Beweis Die X_i sind paarweise unkorreliert. Die Behauptung folgt induktiv mit (4.25), S. 58. $\square$

Satz 5.5.10 Seien X_1, X_2 unabhängige Zufallsvariablen.

- (a) Sind X_1, X_2 stetig verteilt mit Dichten f_1, f_2 , dann ist die Zufallsvariable $X_1 + X_2$ stetig verteilt mit der Dichte

$$f_1 * f_2(x) := \int_{\mathbb{R}} f_1(x-y) f_2(y) dy, \quad x \in \mathbb{R}. \quad (5.8)$$

Die Funktion $f_1 * f_2$ heißt **Faltung von f_1 und f_2** . Die Faltung ist kommutativ, also gilt $f_1 * f_2 = f_2 * f_1$.

- (b) Sind $X_1, X_2 : \Omega \longrightarrow \mathbb{N}_0$ diskret verteilt, dann ist die Zufallsvariable $X_1 + X_2$ diskret verteilt mit der Wahrscheinlichkeitsfunktion

$$\mathbb{P}(X_1 + X_2 = k) = \sum_{i=0}^k \mathbb{P}(X_1 = i) \mathbb{P}_2(X_2 = k - i), \quad k \in \mathbb{N}_0. \quad (5.9)$$

Beweis (a) Die gemeinsame Verteilung $\mathbb{P}_{12}$ von (X_1, X_2) besitzt laut Korollar 5.5.5 (b) die Dichte $f_1 \cdot f_2$. Für $s \in \mathbb{R}$ sei $B := \{(x, y) : x + y \leq s\}$. Dann gilt

$$\begin{aligned} \mathbb{P}(X_1 + X_2 \leq s) &= \mathbb{P}_{12}(B) = \int_B d\mathbb{P}_{12} \stackrel{(5.5)}{=} \int_B d(\mathbb{P}_1 \otimes \mathbb{P}_2) \\ &\stackrel{(2.18)}{=} \int_{\mathbb{R}} \mathbb{P}_1(B_y) d\mathbb{P}_2(y). \end{aligned}$$

Nun gilt für $x \in \mathbb{R}$: $B_y = \{x | x + y \leq s\} = (-\infty, s - y]$. Es folgt weiter

$$\begin{aligned} \mathbb{P}(X_1 + X_2 \leq s) &= \int_{\mathbb{R}} \int_{-\infty}^{s-y} d\mathbb{P}_1(x) d\mathbb{P}_2(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{s-y} f_1(x) f_2(y) dx dy \\ &= \int_{-\infty}^{\infty} f_2(y) \left(\int_{-\infty}^{s-y} f_1(x) dx \right) dy. \end{aligned}$$

Nun gilt mit der Transformationsformel für das Lebesgueintegral (Substitution $t = x - y$)

$$\int_{-\infty}^s f_1(x - y) dy = \int_{-\infty}^{s-y} f_1(t) dt.$$

Setzt man dies in die vorherige Rechnung ein, dann erhalten wir

$$\begin{aligned} \mathbb{P}(X_1 + Y_2 \leq s) &= \int_{-\infty}^{\infty} \int_{-\infty}^s f_1(x - y) f_2(y) dx dy \\ &\stackrel{\text{Fubini}}{=} \int_{-\infty}^s \int_{-\infty}^{\infty} f_1(x - y) f_2(y) dy dx \\ &= \int_{\mathbb{R}} f_1 * f_2(x) dx. \end{aligned}$$

Die Kommutativität der Faltung ist offensichtlich.

(b) Dies folgt sofort aus

$$\{X_1 + X_2 = k\} = \bigcup_{i=0}^k \{X_1 = i, X_2 = k - i\}$$

wobei die Mengen auf der rechten Seite paarweise disjunkt sind und Korollar 5.5.5 (c). $\square$

Bemerkung 5.5.11 Die Sätze über Produkte und Summen unabhängiger Zufallsvariablen können induktiv auf endlich viele Zufallsvariable übertragen werden.

Beispiel 5.5.12 Seien X_1, X_2 unabhängig.

(a) $X_i \sim B(1, p)$, $i = 1, 2$. Dann gilt $X_1 + X_2 \sim B(2, p)$.

(b) $X_i \sim U[0, 1]$, $i = 1, 2$. Bestimme die Dichte von $X_1 + X_2$ (Dreiecksverteilung).

Lösung (a) Laut (5.9) gilt

$$\begin{aligned} \mathbb{P}(X_1 + X_2 = 0) &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 0) = (1 - p)(1 - p) = (1 - p)^2 \\ \mathbb{P}(X_1 + X_2 = 1) &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 1) + \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 0) \\ &= 2p(1 - p) \\ \mathbb{P}(X_1 + X_2 = 2) &= \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 2) + \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) \\ &\quad + \mathbb{P}(X_1 = 2)\mathbb{P}(X_2 = 0) = (1 - p)^2 \end{aligned}$$

(b) Laut (5.8) ist die Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch

$$\int_{\mathbb{R}} 1_{[0,1]}(y) 1_{[0,1]}(x - y) dy = \int_0^1 1_{[0,1]}(x - y) dy = \int_{x-1}^x 1_{[0,1]}(t) dt$$

und es ergibt sich

$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ x & \text{für } x \in (0, 1] \\ 1 - x & \text{für } x \in (1, 2] \\ 0 & \text{für } x > 2 \end{cases}.$$

$\square$

Bemerkung 5.5.13 Bei der Binomialverteilung ist also die Summe zweier $B(1, p)$ verteilter unabhängiger Zufallsvariablen wieder binomial verteilt. Diese Eigenschaft, dass der Verteilungstyp beibehalten wird, haben einige der vorgestellten Verteilungen. Darauf gehen wir im übernächsten Unterabschnitt 5.5.5, S. 75 ein.

5.5.4 Erzeugende Funktionen

Für diesen Unterabschnitt benötigen wir eine wichtige Eigenschaft der erzeugenden Funktionen. Das nächste Lemma dient zur Vorbereitung.

Lemma 5.5.14 *Seien X, Y unabhängige Zufallsvariablen, $f, g : \mathbb{R} \rightarrow \mathbb{R}$ messbar. Dann sind auch $f(X)$ und $g(Y)$ unabhängig.*

Beweis Seien $A, B \in \mathcal{B}^1$. Dann gilt

$$\begin{aligned} \mathbb{P}(f \circ X \in A, g \circ Y \in B) &= \mathbb{P}(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= \mathbb{P}(X \in f^{-1}(A))\mathbb{P}(Y \in g^{-1}(B)) \\ &= \mathbb{P}(f \circ X \in A)\mathbb{P}(g \circ Y \in B). \end{aligned}$$

Die Behauptung folgt aus Satz 5.5.3, S. 69. $\square$

Satz 5.5.15 *Seien X, Y unabhängige Zufallsvariable. Dann gilt falls existent*

$$\begin{aligned} MEF_{X+Y}(t) &= MEF_X(t) \cdot MEF_Y(t) \\ CF_{X+Y}(t) &= CF_X(t) \cdot CF_Y(t) \end{aligned}$$

Sind X, Y diskret, dann gilt, falls existent auch

$$EF_{X+Y}(t) = EF_X(t) \cdot EF_Y(t)$$

Beweis Da X, Y unabhängig sind, sind laut obigem Lemma auch folgende Paare

$$e^{tX}, e^{tY}$$

$$e^{itX}, e^{itY}$$

$$t^X, t^Y$$

jeweils unabhängig. Somit folgt mit Satz 5.5.6, S. 70 und der Definition von CF , S. 54 beispielsweise

$$CF_{X+Y}(t) = \mathbb{E}(e^{itX+itY}) = \mathbb{E}(e^{itX}e^{itY}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}) = CF_X(t)CF_Y(t).$$

Die anderen Behauptungen folgen analog. $\square$

5.5.5 Beispiele Faltung

Mit den erzeugenden Funktionen können wir nun leicht die Aufgabe aus Bemerkung 5.5.13 lösen.

Satz 5.5.16 *Seien X_1, X_2 unabhängige Zufallsvariable.*

- (a) $X_i \sim B(n_i, p) \implies X_1 + X_2 \sim B(n_1 + n_2, p)$
- (b) $X_i \sim \text{Poi}(\lambda_i) \implies X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$
- (c) $X_i \sim NB(m_i, p) \implies X_1 + X_2 \sim NB(m_1 + m_2, p)$
- (d) $X_i \sim N(\mu_i, \sigma_i^2) \implies X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- (e) $X_i \sim \Gamma(\alpha_i, \lambda) \implies X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$

Beweis Mit Satz 5.5.15 sieht man aus der folgenden Tabelle

Verteilung	$EF(t)$	$MEF(t)$
$B(n, p)$	$(pt + (1 - p))^n$	
$\text{Poi}(\lambda)$	$\exp(\lambda(t - 1))$	
$NB(m, p)$	$(1 - p)^m (1 - pt)^{-m}$	
$N(\mu, \sigma)$		$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$
$\Gamma(\alpha, \lambda)$		$\left(\frac{\lambda}{\lambda - t}\right)^\alpha, t < \lambda$

dass die Potenzen der jeweilige erzeugende Funktion vom gleichen Typ sind. Die Behauptungen folgen aus den Eindeutigkeitssätzen Satz 4.7.2 und Satz 4.7.6. $\square$

5.5.6 Eigenschaften der mehrdimensionalen Normalverteilung

Wir stellen noch einige wichtige Eigenschaften zusammen. Zunächst kehren wir auf Bemerkung 5.5.8, S. 71 zurück.

Satz 5.5.17 *Sei $X \sim N_n(\mu, \Sigma)$. Die Komponenten von X sind genau dann unabhängig, wenn sie unkorreliert sind, d.h. wenn Σ eine Diagonalmatrix ist.*

Der nächste Satz behandelt die „Dekorrelation“.

Satz 5.5.18 Sei $X \sim N_n(\mu, \Sigma)$ und $A \in \mathbb{R}^{n \times n}$ invertierbar mit

$$A \cdot A^T = \Sigma. \quad [A \text{ Wurzel aus } \Sigma]$$

Sei

$$Y := A^{-1}(X - \mu) \quad [Y \text{ heißt Standardisierte von } X].$$

Dann gilt für die Komponenten Y_i , $i = 1, \dots, n$ von Y :

$$Y_i \sim N(0, 1) \text{ und die } Y_i \text{ sind unabhängig.}$$

6 Bedingte Verteilungen und Momente

Die hier eingeführten Begriffe sind für stochastische Prozesse (Finanzmathematik) entscheidend. Sie finden auch in der Credibility Theorie eine wichtige Anwendung.

Zunächst erklären wir das Konzept an Beispielen mit diskreten Zufallsvariablen, übertragen dann das Vorgehen auf allgemeine Zufallsvariablen.

Wir erinnern an folgende Ergebnisse für einen W-Raum $(\Omega, \mathbb{P}, \mathcal{A})$ und ein $B \subset \Omega$ mit $\mathbb{P}(B) > 0$:

- $(\Omega, \mathcal{A}, \mathbb{P}_B)$ mit $\mathbb{P}_B(A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, $A \in \mathcal{A}$ ist ein W-Raum.
- Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $X1_B \geq 0$ bzw. $X1_B$ integrierbar. Dann gilt

$$\int_{\Omega} X d\mathbb{P}_B = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} \quad (6.1)$$

bzw.

$$\mathbb{E}_B(X) := \mathbb{E}(X|B) := \mathbb{E}_{\mathbb{P}_B}(X) = \int_{\Omega} X d\mathbb{P}_B = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} \quad (6.2)$$

6.1 Diskrete Zufallsvariablen

Beispiel 6.1.1 Es seien X_1, X_2 unabhängige, binomialverteilte Zufallsvariablen, $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$.

- Bestimme den bedingten Erwartungswert von X_1 gegeben $X_1 + X_2$, in Zeichen $\mathbb{E}(X_1|X_1 + X_2)$.
- Zeige $\mathbb{E}(\mathbb{E}(X_1|X_1 + X_2)) = \mathbb{E}(X_1)$.

Lösung (a) Wegen Satz 5.5.16 (a) ist $X_1 + X_2 \sim B(n_1 + n_2, p)$. Laut (6.2) gilt

$$\begin{aligned} \mathbb{E}(X_1|X_1 + X_2 = n) &= \mathbb{E}_{\{X_1+X_2=n\}}(X_1) = \sum_{k=0}^n k \mathbb{P}_{\{X_1+X_2=n\}}(X_1 = k) \\ &= \sum_{k=0}^n k \mathbb{P}(X_1 = k|X_1 + X_2 = n). \end{aligned}$$

Nun:

$$\begin{aligned}
& \mathbb{P}(X_1 = k | X_1 + X_2 = n) \\
&= \frac{\mathbb{P}(X_1 = k, X_2 = n - k)}{\mathbb{P}(X_1 + X_2 = n)} = \frac{\mathbb{P}(X_1 = k) \mathbb{P}(X_2 = n - k)}{\mathbb{P}(X_1 + X_2 = n)} \\
&= \frac{\binom{n_1}{k} p^k (1-p)^{n_1-k} \binom{n_2}{n-k} p^{n-k} (1-p)^{n_2-n+k}}{\binom{n_1+n_2}{n} p^n (1-p)^{n_1+n_2-n}} \\
&= \frac{\binom{n_1}{k} \binom{n_2}{n-k}}{\binom{n_1+n_2}{n}}
\end{aligned}$$

Somit liegt eine hypergeometrische Verteilung mit Parametern n_1 , n_2 und n . Dann ist

$$\mathbb{E}(X_1 | X_1 + X_2 = n) = n \cdot \frac{n_1}{n_1 + n_2}$$

also

$$\mathbb{E}(X_1 | X_1 + X_2) = (X_1 + X_2) \cdot \frac{n_1}{n_1 + n_2}.$$

$$(b) \mathbb{E}(\mathbb{E}(X_1 | X_1 + X_2)) = (n_1 p + n_2 p) \frac{n_1}{n_1 + n_2} = n_1 p. \quad \square$$

Bemerkung 6.1.2 Setzt man im obigen Beispiel $Y := X_1 + X_2$, dann ist $\mathbb{E}(X|Y)$ eine Zufallsvariable und nicht eine Zahl. Der Erwartungswert von $\mathbb{E}(X|Y)$ ist dann gleich dem nicht bedingten Erwartungswert von X , also

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X).$$

6.2 Definition im allgemeinen Fall

Satz 6.2.1 Sei $(\Omega, \mathbb{P}, \mathcal{A})$ ein W-Raum, $X : \Omega \rightarrow \mathbb{R}$ integrierbar. $\mathcal{C} \subset \mathcal{A}$ eine σ -Algebra. Dann gibt es eine $\mathcal{C}$ -messbare Zufallsvariable $X_0 : \Omega \rightarrow \mathbb{R}$ mit

$$\forall C \in \mathcal{C} : \int_C X d\mathbb{P} = \int_C X_0 d\mathbb{P}. \quad (6.3)$$

X_0 ist bis auf eine Nullmenge eindeutig bestimmt.

Beweis [B-WT], S. 117 f. Es wird der Satz von Radon Nykodym auf die Maße $(X\mathbb{P})|_{\mathcal{C}}$ und $\mathbb{P}|_{\mathcal{C}}$ angewendet. $\square$

X_0 in (6.3) ist nicht eindeutig bestimmt, sondern nur f.s., d.h. jedes messbare $\tilde{X}_0$ mit $\tilde{X}_0 = X_0$ f.s. erfüllt auch (6.3).

Definition 6.2.2 Unter den Begebenheiten von Satz 6.2.1 heißt $X_0 : \Omega \longrightarrow \mathbb{R}$ *bedingte Erwartung von X unter der Bedingung $\mathcal{C}$* und wird mit

$$\mathbb{E}(X|\mathcal{C}) \text{ bzw. } \mathbb{E}^{\mathcal{C}}(X)$$

bezeichnet. Ist $Y : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable, dann bezeichnen wir mit

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y))$$

die *bedingte Erwartung von X gegeben Y* .

Für $\mathbb{E}(X|Y)$ wird (6.3) wegen $\sigma(Y) = Y^{-1}(\mathcal{B}^1)$ zu

$$\forall B \in \mathcal{B}^1 : \quad \int_{Y^{-1}(B)} X d\mathbb{P} = \int_{Y^{-1}(B)} X_0 d\mathbb{P}. \quad (6.4)$$

Da $\mathbb{E}(X|\mathcal{C})$ nur $\mathbb{P}$ -f.s. eindeutig bestimmt ist, sprechen wir von den verschiedenen $\mathbb{P}$ f.s. gleichen, **Versionen** der bedingten Erwartung. Mit $\mathbb{E}(X|\mathcal{C})$ ist also stets eine Zufallsvariable gemeint, die eine Version der bedingten Erwartung ist.

Wir wollen nun zeigen, dass die hier gegebene Definition mit unserem Vorgehen am Kapitelanfang zusammenpasst.

Beispiel 6.2.3 Sei $\Omega = \bigcup_{i=1}^{\infty} B_i$, wobei die $B_i \in \mathcal{A}$ paarweise disjunkt seien. Für eine Zufallsvariable X gilt für $\mathcal{C} := \sigma(B_i : i \in \mathbb{N})$

$$\mathbb{E}(X|\mathcal{C}) = \sum_{i=1}^{\infty} \mathbb{E}_{B_i}(X) 1_{B_i}. \quad (6.5)$$

Beweis Das Mengensystem bestehend aus den B_i und der leeren Menge ist ein $\cap$ -stabiler Erzeuger von $\mathcal{C}$ mit $\bigcup_{i=1}^{\infty} B_i = \Omega$. Wegen Korollar 2.4.2, S. 10 reicht es (6.3) für alle B_i zu überprüfen also

$$\forall i \in \mathbb{N} : \quad \int_{B_i} X d\mathbb{P} = \int_{B_i} \sum_{j=1}^{\infty} \mathbb{E}_{B_j}(X) 1_{B_j} d\mathbb{P}.$$

Sei $i \in \mathbb{N}$. Dann gilt

$$\begin{aligned} \int_{B_i} \sum_{j=1}^{\infty} \mathbb{E}_{B_j}(X) 1_{B_j} d\mathbb{P} &\stackrel{B_j \text{ dis.}}{=} \int_{B_i} \mathbb{E}_{B_i}(X) 1_{B_i} d\mathbb{P} = \mathbb{E}_{B_i}(X) \int_{B_i} d\mathbb{P} \\ &= \mathbb{E}_{B_i}(X) \cdot \mathbb{P}(B_i) \stackrel{(6.2)}{=} \int_{B_i} X d\mathbb{P}. \end{aligned}$$

□

Dieses Beispiel können wir sofort auf diskrete Zufallsvariablen übertragen.

Beispiel 6.2.4 Seien $X, Y : \Omega \longrightarrow \mathbb{N}$ diskrete Zufallsvariablen. Es sei

$$B_j := \{Y = j\}, \quad j \in \mathbb{N}.$$

Dann gilt

$$\mathbb{E}(X|Y)|_{B_j} = \mathbb{E}_{B_j}(X) \stackrel{(6.2)}{=} \mathbb{E}(X|Y = j) = \sum_{i=1}^{\infty} i\mathbb{P}(X = i|Y = j)$$

d.h. die Zufallsvariable $\mathbb{E}(X|Y)$ nimmt mit Wahrscheinlichkeit $\mathbb{P}(Y = j)$ den Wert $\mathbb{E}(X|Y = j)$ an.

Beweis Es seien

$$A_i := \{X = i\}, \quad B_j := \{Y = j\}, \quad i, j \in \mathbb{N}.$$

Wir verwenden das vorherige Beispiel 6.2.3. Es gilt mit $B_j := \{Y = j\}$ und $\mathcal{C} := \sigma(\{B_i | i \in \mathbb{N}\})$

$$\mathbb{E}_{B_j}(X) = \mathbb{E}_{B_j} \left(\sum_{i=1}^{\infty} i1_{A_i} \right) = \int \sum_{i=1}^{\infty} i1_{A_i} d\mathbb{P}_{B_j}$$

Wegen $0 \leq \sum_{i=1}^n i1_{A_i} \uparrow \sum_{i=1}^{\infty} i1_{A_i}$ kann man mit dem Satz von der monotonen Konvergenz die Summe und Integral vertauschen, also

$$\begin{aligned} \mathbb{E}_{B_j}(X) &= \sum_{i=1}^{\infty} i \int 1_{A_i} d\mathbb{P}_{B_j} = \sum_{i=1}^{\infty} i \frac{1}{\mathbb{P}(B_j)} \int_{B_j} 1_{A_i} d\mathbb{P} \\ &= \sum_{i=1}^{\infty} i\mathbb{P}(X = i|Y = j) \end{aligned}$$

und (6.5), S. 79

$$\mathbb{E}(X|\mathcal{C}) = \sum_{j=1}^{\infty} \mathbb{E}_{B_j}(X)1_{B_j} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} i\mathbb{P}(X = i|Y = j)1_{B_j}.$$

Nun gilt für $\omega \in B_k$

$$\mathbb{E}(X|\mathcal{C})(\omega) = \sum_{i=1}^{\infty} i\mathbb{P}(X = i|Y = k)1_{B_k}(\omega) = \sum_{i=1}^{\infty} i\mathbb{P}(X = i|Y = k).$$

□

6.3 Eigenschaften

Die folgenden Regeln werden immer wieder angewendet: Es gilt $\mathbb{P}$ -f.s.:

$$\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X), \mathbb{E}(X|\mathcal{A}) = X \quad (6.6)$$

$$\mathbb{E}(\alpha X + \beta Y|\mathcal{C}) = \mathbb{E}(\alpha X|\mathcal{C}) + \mathbb{E}(\beta Y|\mathcal{C}) \quad (6.7)$$

$$X \leq Y \implies \mathbb{E}(X|\mathcal{C}) \leq \mathbb{E}(Y|\mathcal{C}) \quad (6.8)$$

$$Y \text{ } \mathcal{C}\text{-messbar} \implies \mathbb{E}(XY|\mathcal{C}) = Y\mathbb{E}(X|\mathcal{C}) \quad (6.9)$$

$$\mathcal{C}_1, \mathcal{C}_2 \text{ } \sigma\text{-Algebren, } \mathcal{C}_1 \subset \mathcal{C}_2 \implies \mathbb{E}(\mathbb{E}(X|\mathcal{C}_2)|\mathcal{C}_1) = \mathbb{E}(X|\mathcal{C}_1) \quad (6.10)$$

$$\mathbb{E}(\mathbb{E}(X|\mathcal{C})) = \mathbb{E}(X) \quad (6.11)$$

$$\sigma(X), \mathcal{C} \text{ unabhängig} \implies \mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X). \quad (6.12)$$

Ist $\mathcal{C}$ durch eine messbare Abbildung $Z : \Omega \longrightarrow \mathbb{R}$ erzeugt, also

$$\mathcal{C} = \sigma(Z),$$

dann gelten die obigen Regeln (6.7)-(6.12) entsprechend:

$$\mathbb{E}(\alpha X + \beta Y|Z) = \mathbb{E}(\alpha X|Z) + \mathbb{E}(\beta Y|Z) \quad (6.13)$$

$$X \leq Y \implies \mathbb{E}(X|Z) \leq \mathbb{E}(Y|Z) \quad (6.14)$$

$$f : \mathbb{R} \longrightarrow \mathbb{R} \text{ messbar} \implies \mathbb{E}(Xf(Z)|Z) = f(Z)\mathbb{E}(X|Z) \quad (6.15)$$

$$\mathbb{E}(\mathbb{E}(X|Y, Z)|Y) = \mathbb{E}(X|Y) \quad (6.16)$$

$$X, Y \text{ unabhängig} \implies \mathbb{E}(X|Y) = \mathbb{E}(X). \quad (6.17)$$

Die bedingte Varianz bzw. Kovarianz von X bzw. X, Y gegeben $\mathcal{C}$ wird wie folgt definiert:

Definition 6.3.1 *Seien X, Y quadratintegrierbar und $\mathcal{C} \subset \mathcal{A}$ eine σ -Algebra. Dann heißt*

$$\begin{aligned} \text{Var}(X|\mathcal{C}) &:= \mathbb{E}([X - \mathbb{E}(X|\mathcal{C})]^2|\mathcal{C}) \text{ bzw.} \\ \text{Cov}(X, Y|\mathcal{C}) &:= \mathbb{E}([X - \mathbb{E}(X|\mathcal{C})][Y - \mathbb{E}(Y|\mathcal{C})]|\mathcal{C}) \end{aligned}$$

bedingte Varianz bzw. **Kovarianz** von X bzw. Y unter der Bedingung $\mathcal{C}$.

Lemma 6.3.2 *Unter den Bedingung von Definition 6.3.1 gilt*

$$\text{Cov}(X, Y|\mathcal{C}) = \mathbb{E}(XY|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C}) \quad (6.18)$$

$$\text{Cov}(X, Y) = \mathbb{E}(\text{Cov}(X, Y|\mathcal{C})) + \text{Cov}(\mathbb{E}(X|\mathcal{C}), \mathbb{E}(Y|\mathcal{C})) \quad (6.19)$$

Beweis Wir verwenden die oben zusammen gestellten Regeln für das Rechnen mit bedingten Erwartungen.

$$\begin{aligned}
\text{Cov}(X, Y|\mathcal{C}) &\stackrel{(6.7)}{=} \mathbb{E}(XY|\mathcal{C}) - \mathbb{E}(X\mathbb{E}(Y|\mathcal{C})|\mathcal{C}) - \mathbb{E}(Y\mathbb{E}(X|\mathcal{C})|\mathcal{C}) \\
&\quad + \mathbb{E}(\mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})|\mathcal{C}) \\
&\stackrel{(6.9)}{=} \mathbb{E}(XY|\mathcal{C}) - \mathbb{E}(Y|\mathcal{C})\mathbb{E}(X|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C}) \\
&\quad + \mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})\underbrace{\mathbb{E}(1|\mathcal{C})}_{=1} \\
&= \mathbb{E}(XY|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})
\end{aligned}$$

Nun folgt

$$\begin{aligned}
\mathbb{E}(\text{Cov}(X, Y|\mathcal{C})) &= \mathbb{E}[\mathbb{E}(XY|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})] \\
&\stackrel{(6.7)}{=} \mathbb{E}(\mathbb{E}(XY|\mathcal{C})) - \mathbb{E}(\mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})) \\
&\stackrel{(6.11)}{=} \mathbb{E}(XY) - \mathbb{E}(\mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})) \\
\text{Cov}(\mathbb{E}(X|\mathcal{C}), \mathbb{E}(Y|\mathcal{C})) &= \mathbb{E}(\mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})) - \mathbb{E}(\mathbb{E}(X|\mathcal{C})) \cdot \mathbb{E}(\mathbb{E}(Y|\mathcal{C})) \\
&\stackrel{(6.11)}{=} \mathbb{E}(\mathbb{E}(X|\mathcal{C})\mathbb{E}(Y|\mathcal{C})) - \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

Addiert man $\mathbb{E}(\text{Cov}(X, Y|\mathcal{C}))$ und $\text{Cov}(\mathbb{E}(X|\mathcal{C}), \mathbb{E}(Y|\mathcal{C}))$ folgt das Gewünschte. $\square$

6.4 Darstellung

Ist $Y : \Omega \longrightarrow \mathbb{N}$ diskret, kann man wegen Beispiel 6.2.3 mühelos die die Werte bzw. Wahrscheinlichkeitsfunktion von $\mathbb{E}(X|Y)$ angeben, nämlich $\mathbb{E}(X|Y = j)$ bzw. $\mathbb{P}(Y = j)$. In der Folge wird das Analogon im allgemeinen Fall vorgestellt.

In der Folge sei $X : \Omega \longrightarrow \mathbb{R}$ integrierbar, $Y : \Omega \longrightarrow \mathbb{R}$ messbar.

Lemma 6.4.1 (Faktorisierungslemma) *Sei $Y : \Omega \longrightarrow \Omega'$ und $\mathcal{A}'$ eine σ -Algebra auf Ω' . $X : \Omega \longrightarrow \mathbb{R}$ ist bezüglich $\mathcal{A} := Y^{-1}(\mathcal{A}')$ genau dann messbar, wenn es ein $\mathcal{A}' - \mathcal{B}^1$ -messbares $g : \Omega' \longrightarrow \mathbb{R}$ gibt mit*

$$X = g \circ Y.$$

Beweis [B-MT], S. 71, 11.7 Faktorisierungslemma. $\square$

Ist $Y : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable, dann ist also $X : \Omega \longrightarrow \mathbb{R}$ genau dann $\sigma(Y)$ messbar, wenn es ein messbares $g : \mathbb{R} \longrightarrow \mathbb{R}$ gibt mit

$$X = g \circ Y$$

also

$$\begin{array}{ccc} \Omega & \xrightarrow{Y} & (\mathbb{R}, \mathcal{B}^1) \\ & \searrow X & \swarrow g \\ & \mathbb{R} & \end{array} \quad (6.20)$$

Lemma 6.4.2 Sei $X : \Omega \longrightarrow \mathbb{R}$ integrierbar, $Y : \Omega \longrightarrow \mathbb{R}$ messbar und $\mathbb{E}(X|Y)$ eine Version der bedingten Erwartung. Dann gibt es ein $g : \mathbb{R} \longrightarrow \mathbb{R}$ messbar mit

$$\mathbb{E}(X|Y) = g(Y) \quad \mathbb{P}_Y - f.s. \quad (6.21)$$

Beweis Mit dem Faktorisierungslemma angewendet auf $\mathbb{E}(X|Y)$ und Y ergibt sich sofort die Behauptung.

$$\begin{array}{ccc} \Omega & \xrightarrow{Y} & (\mathbb{R}, \mathcal{B}^1) \\ & \searrow \mathbb{E}(X|Y) & \swarrow g \\ & \mathbb{R} & \end{array}$$

$\square$

Lemma 6.4.2 verallgemeinert die Bemerkung 6.1.2 auf alle Zufallsvariablen. $\mathbb{E}(X|Y)$ ist eine Zufallsvariable. Die Werte, die sie annimmt hängen von Y ab, also $\mathbb{E}(X|Y)(\omega) = g(Y(\omega))$ f.s.

Bevor wir konkrete Beispiele berechnen können, benötigen wir noch ein Eindeutigkeitsergebnis:

Satz 6.4.3 Sei $g : \mathbb{R} \longrightarrow \mathbb{R}$ messbar.

(a) Erfüllt g (6.21) dann ist g $\mathbb{P}_Y$ -integrierbar und es gilt

$$\forall B \in \mathcal{B}^1 : \int_B g d\mathbb{P}_Y = \int_{Y^{-1}(B)} X d\mathbb{P}. \quad (6.22)$$

g ist hierdurch $\mathbb{P}_Y$ -f.s. bestimmt.

(b) Erfüllt g (6.22) und ist $\mathbb{P}_Y$ -integrierbar, dann ist $g \circ Y$ eine Version von $\mathbb{E}(X|Y)$.

Beweis (a) Auch als Anwendung der Integration beweisen wir (6.22): Sei $B \in \mathcal{B}^1$

$$\begin{aligned} \int_{Y^{-1}(B)} X \, d\mathbb{P} &\stackrel{(6.4)}{=} \int_{Y^{-1}(B)} \mathbb{E}(X|Y) \, d\mathbb{P} \stackrel{(6.21)}{=} \int_{Y^{-1}(B)} g \circ Y \, d\mathbb{P} \\ &= \int_{\Omega} 1_{Y^{-1}(B)} \cdot g \circ Y \, d\mathbb{P} \end{aligned}$$

Nun gilt

$$1_{Y^{-1}(B)}(\omega) = 1 \iff \omega \in Y^{-1}(B) \iff Y(\omega) \in B \iff 1_B(Y(\omega)) = 1$$

und somit folgt weiter

$$\int_{Y^{-1}(B)} X \, d\mathbb{P} = \int_{\Omega} 1_B(Y) \cdot g \circ Y \, d\mathbb{P} \stackrel{(2.19)}{=} \int_{\mathbb{R}} 1_B \cdot g \, d\mathbb{P}_Y = \int_B g \, d\mathbb{P}_Y$$

Die $\mathbb{P}_Y$ -f.s. Eindeutigkeit von g folgt so: Erfüllt $\tilde{g}$ auch (6.21), dann folgt wie oben

$$\forall B \in \mathcal{B}^1 : \int_B \tilde{g} \, d\mathbb{P}_Y = \int_{Y^{-1}(B)} X \, d\mathbb{P} = \int_B g \, d\mathbb{P}_Y$$

und $g = \tilde{g}$ $\mathbb{P}_Y$ -f.s. wegen Lemma 6.9.1 (b), S. 94.

(b) Wie in (a) folgt für alle $B \in \mathcal{B}^1$

$$\int_{Y^{-1}(B)} g \circ Y \, d\mathbb{P} = \int_B g \, d\mathbb{P}_Y \stackrel{(6.22)}{=} \int_{Y^{-1}(B)} X \, d\mathbb{P},$$

d.h. $g \circ Y$ erfüllt (6.4) somit das Gewünschte. $\square$

Zusammenfassung

- $\mathbb{E}(X|Y)$ ist eine $\sigma(Y)$ -messbare Zufallsvariable
- Es gibt $g : \mathbb{R} \rightarrow \mathbb{R}$ messbar mit $\mathbb{E}(X|Y) = g(Y)$ $\mathbb{P}_Y$ -f.s.
- g ist $\mathbb{P}_Y$ -f.s. eindeutig bestimmt.
- $\mathbb{E}(X|Y)$ ist für $\mathbb{P}_Y$ -fast alle $y \in \mathbb{R}$ konstant auf den Mengen $\{Y = y\}$.
(Beachte: Für $B \in \mathcal{B}^1$, $B \cap Y(\Omega) = \emptyset$ ist $\mathbb{P}_Y(B) = 0$, also interessieren nur $y \in Y(\Omega)$).

Dies führt zu folgender Definition bzw. Schreibweise:

Definition 6.4.4 Sei X integrierbar, $g : \mathbb{R} \rightarrow \mathbb{R}$ sei eine messbare Abbildung, die (6.22) erfüllt. Dann heißt $g(y)$ für jedes $y \in \mathbb{R}$ **die bedingte Erwartung von X gegeben $Y = y$** und wir setzen

$$\mathbb{E}(X|Y = y) := g(y).$$

Es bleibt zu klären, dass die diese Definition für diskrete Y konsistent ist mit der früher verwendeten, d.h. für $Y(\Omega) = \{y_i | i \in I \subset \mathbb{N}\}$ muss gelten

$$\forall i \in I : \quad \mathbb{E}(X|Y = y_i) = \mathbb{E}_{Y=y_i}(X).$$

Dies ist aber nach kurzer Rechnung völlig klar.

Bei diskretem Y gilt

$$\mathbb{E}(X|Y = y_i) = \frac{1}{\mathbb{P}(Y = y_i)} \int_{\{Y=y_i\}} X d\mathbb{P}.$$

Bei stetig verteiltem Y gilt $\mathbb{P}(Y = y) = 0$. Der Versuch also $\mathbb{E}(X|Y = y)$ mittels

$$\mathbb{E}(X|Y = y) = \frac{1}{\mathbb{P}(Y = y)} \int_{\{Y=y\}} X d\mathbb{P}$$

zu bestimmen, würde in diesem Fall scheitern. Dennoch kann man „wie üblich“ rechnen. Im Falle einer gemeinsamen Verteilung von X, Y mit Dichte f kann man $\mathbb{E}(X|Y = y)$ direkt berechnen:

Satz 6.4.5 Es seien X, Y Zufallsvariable, (X, Y) sei stetig verteilt mit Dichte $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. X sei integrierbar und gelte

$$\forall y \in Y(\Omega) : f_Y(y) := \int_{\mathbb{R}} f(x, y) dx > 0.$$

Dann gilt

$$\mathbb{E}(X|Y = y) = \frac{1}{f_Y(y)} \int_{\mathbb{R}} x f(x, y) dx \quad \mathbb{P}_Y \text{ f.s.} \quad (6.23)$$

und

$$\mathbb{E}(X|Y) = \frac{1}{f_Y(Y)} \int_{\mathbb{R}} x f(x, Y) dx \quad \mathbb{P} \text{ f.s.}$$

Beweis [B-WT], S. 130, 15.11 Satz. □

Die Funktion

$$f(x|y) := \frac{f(x, y)}{f_Y(y)}, x \in \mathbb{R}, y \in \mathbb{R}, f_Y(y) > 0$$

heißt **bedingte Dichte**. Für $y \in \mathbb{R}$ mit $f_Y(y) = 0$ kann man $f(x|y)$ beliebig definieren. Damit wird (6.23) zu

$$\mathbb{E}(X|Y = y) = \int_{\mathbb{R}} xf(x|y) dx \quad \mathbb{P}_Y \text{ f.s.}$$

Die bedingte Dichte erinnert an die Definition der elementaren bedingten Wahrscheinlichkeit. Man beachte weiter die Analogie zur Berechnung des nicht bedingten Erwartungswerts bei gegebener Dichte. Zur Berechnung von $\mathbb{E}(g(X)|Y = y)$ siehe (6.27), S. 90.

Beispiel 6.4.6 Die gemeinsame Dichte von (X, Y) sei gegeben durch

$$f(x, y) = \begin{cases} 6xy(2 - x - y) & x, y \in (0, 1) \\ 0 & \text{sonst.} \end{cases}$$

Bestimmen Sie die bedingte Dichte von X gegeben $Y = y$.

Lösung $f(x|y) = \frac{5 - 4y}{8 - 6y}$

□

Beispiel 6.4.7 (X, Y) sei zwei-dimensional Normalverteilt, $X, Y \sim N(0, \sigma^2)$, $\sigma > 0$ und $\rho(X, Y) = \rho \in (-1, 1)$. Die gemeinsame Verteilung von X, Y sei dann durch die Dichte

$$f(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right]$$

gegeben. Zeigen Sie:

$$f(x|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} (x - \rho y)^2 \right].$$

und $\mathbb{E}(X|Y) = \rho Y$ $\mathbb{P}$ -f.s.

Interpretation $f(\cdot|y)$ ist die Dichte einer $N(\rho y, \sigma^2(1-\rho^2))$ verteilten Zufallsvariablen. Gegeben $Y = y$ ist $X \sim N(\rho y, \sigma^2(1-\rho^2))$. Dabei ist hier der Parameter Y wiederum normalverteilt.

Beispiel 6.4.8 Bei gegebenem $\Theta = \vartheta$ sei $X \sim N(\vartheta, \sigma^2)$ -verteilt, d.h.

$$f(x|\vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \vartheta)^2}{2\sigma^2} \right]$$

Ferner sei $\Theta \sim N(\mu, \tau^2)$ verteilt (d.h. der Parameter ϑ der Zufallsvariablen X ist selbst eine Zufallsvariable)

(a) Bestimmen Sie die gemeinsame Dichte von (X, Θ) .

(b) Bestimmen Sie $\mathbb{E}(X|\Theta = \vartheta)$.

(c) Bestimmen Sie $\mathbb{E}(X)$ und $\text{Var}(X)$.

Lösung $\mathbb{E}(X|\Theta = \vartheta) = \vartheta$, $\mathbb{E}(X) = \mu$, $\text{Var}(X) = \sigma^2 + \tau^2$

□

6.5 Berechnungen durch Bedingen

Die beiden vorherigen Abschnitte kann man zur Berechnung von Erwartungswerten und Wahrscheinlichkeiten heranziehen.

6.5.1 Erwartungswerte

Lemma 6.5.1 Sei X integrierbar, Y messbar. Dann gilt

$$\mathbb{E}(X) = \int_{\mathbb{R}} \mathbb{E}(X|Y = y) d\mathbb{P}_Y.$$

Ist Y diskret verteilt, bzw. stetig verteilt mit Dichte f_Y gilt

$$\mathbb{E}(X) = \sum_{i \in I} \mathbb{E}(X|Y = y_i) \mathbb{P}(Y = y_i) \text{ bzw.} \quad (6.24)$$

$$\mathbb{E}(X) = \int_{\mathbb{R}} \mathbb{E}(X|Y = y) f_Y(y) dy. \quad (6.25)$$

Beweis Aus $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$ folgt durch Einsetzen

$$\mathbb{E}(X) = \int_{\Omega} \mathbb{E}(X|Y) d\mathbb{P}$$

und daraus mit Satz 6.4.3 die Behauptung. □

Beispiel 6.5.2 Ein Student hat die Auswahl zwischen zwei Büchern, Mathematik und Wirtschaft. Die Anzahl der Druckfehler sei jeweils Poisson-verteilt mit Erwartungswert 5 bzw. 2. Die Wahrscheinlichkeit eines der beiden Bücher zu wählen ist jeweils 0,5. Bestimme die erwartete Anzahl der Fehler, die der Student finden wird.

Lösung Sei X die Anzahl der gefundenen Fehler, Y das gewählte Buch, $Y = M$ bzw. $Y = W$

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y)) \\ &\stackrel{(6.24)}{=} \mathbb{E}(X|Y = M) \mathbb{P}(Y = M) + \mathbb{E}(X|Y = W) \mathbb{P}(Y = W) \\ &= 5 \cdot 0,5 + 2 \cdot 0,5 = 3,5. \end{aligned}$$

□

Beispiel 6.5.3 *Ein Bergmann befindet in einem Raum in dem drei Gängen beginnen. Der erste gang führt nach zwei Stunden zur Oberfläche, der zweite bzw. dritte gang führt nach drei bzw. fünf Stunden wieder zurück zum Raum. Wir nehmen an, dass die Auswahlwahrscheinlichkeit jeder Türe zu jedem Zeitpunkt jeweils $1/3$ beträgt, d.h. der Bergmann vergisst welche Türe er gewählt hatte. Welches ist die erwartete Zeit, die der Bergmann zur Oberfläche benötigt?*

Lösung Sei $X : \Omega \longrightarrow \{1, 2, 3\}$ die Zufallsvariable, die angibt für welchen Weg sich der Bergmann entschieden hat, T sei die Anzahl der Stunden.

$$\begin{aligned} \mathbb{E}(T) &= \mathbb{E}(\mathbb{E}(T|X)) = \sum_{i=1}^3 \mathbb{E}(T|X=i) \mathbb{P}(X=i) \\ &= \frac{1}{3} (2 + (3 + \mathbb{E}(T)) + (5 + \mathbb{E}(T))) \\ &= \frac{10}{3} + \frac{2}{3} \mathbb{E}(T) \\ \implies \mathbb{E}(T) &= 10. \end{aligned}$$

□

6.5.2 Wahrscheinlichkeiten

Für $A \in \mathcal{A}$ gilt laut Definition

$$\mathbb{P}(A) = \mathbb{E}(1_A).$$

Damit definieren wir

$$\mathbb{P}(A|Y) := \mathbb{E}(1_A|Y).$$

Lemma 6.5.4 *Sei $A \in \mathcal{A}$, Y messbar. Dann gilt*

$$\mathbb{P}(A) = \int_{\mathbb{R}} \mathbb{P}(A|Y=y) d\mathbb{P}_Y.$$

Ist Y diskret verteilt, bzw. stetig verteilt mit Dichte f_Y gilt

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i \in I} \mathbb{P}(A|Y=y_i) \mathbb{P}(Y=y_i) \text{ bzw.} \\ \mathbb{P}(A) &= \int_{\mathbb{R}} \mathbb{P}(A|Y=y) f_Y(y) dy. \end{aligned}$$

Beweis Analog zum Beweis von Lemma 6.5.1 indem man dort X durch 1_A ersetzt. $\square$

Beispiel 6.5.5 Jeder Kunde, der den Top-Moden Laden betritt, kauft mit Wahrscheinlichkeit $p \in (0, 1)$ einen Anzug. Die Anzahl der Kunden, die den Laden betreten sei Poisson verteilt mit Erwartungswert λ .

- (a) Wie groß ist die Wahrscheinlichkeit, dass Top-Moden keinen Anzug verkauft.
- (b) Wie groß ist die Wahrscheinlichkeit, dass Top-Moden genau k Anzüge verkauft.

Lösung (a) Sei N die Anzahl Kunden die den Laden betreten, E das Ereignis keinen Anzug zu verkaufen.

$$\begin{aligned}\mathbb{P}(E) &= \sum_{n \geq 0} \mathbb{P}(E|N = n) \mathbb{P}(N = n) = \sum_{n \geq 0} (1-p)^n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \sum_{n \geq 0} e^{-\lambda} \frac{((1-p)\lambda)^n}{n!} = e^{-\lambda} e^{\lambda(1-p)} = e^{-\lambda p}\end{aligned}$$

(b) Sei N die Anzahl Kunden die den Laden betreten, E das Ereignis k Anzüge zu verkaufen.

$$\begin{aligned}\mathbb{P}(E) &= \sum_{n \geq k} \mathbb{P}(E|N = n) \mathbb{P}(N = n) = \sum_{n \geq k} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \lambda^k p^k \frac{1}{k!} \sum_{n \geq k} \frac{((1-p)\lambda)^{n-k}}{(n-k)!} = e^{-\lambda} \lambda^k p^k \frac{1}{k!} e^{(1-p)\lambda} \\ &= \frac{(p\lambda)^k}{k!} e^{-p\lambda} = \text{Poi}(k, p\lambda).\end{aligned}$$

$\square$

6.6 Bedingte Wahrscheinlichkeiten

Beispiel 6.6.1 Sei $X_i \sim B(n_i, p)$, $i = 1, 2$. Dann gilt laut Beispiel 6.1.1

$$\mathbb{E}(X_1|X_1 + X_2) = (X_1 + X_2) \frac{n_1}{n_1 + n_2}.$$

Für festes $n \in \{0, \dots, n_1 + n_2\}$ ist für $A \subset \{0, \dots, n_1 + n_2\}$ durch

$$\mathbb{P}(X_1 \in A | X_1 + X_2 = n)$$

ein Wahrscheinlichkeitsmaß definiert, denn

$$\mathbb{P}(X_1 = k | X_1 + X_2 = n) = \begin{cases} \frac{\binom{n_1}{k} \binom{n_2}{n-k}}{\binom{n_1+n_2}{n}} & k \leq n \\ 0 & \text{sonst.} \end{cases}$$

Dieses Vorgehen kann man verallgemeinern:

Definition 6.6.2 Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{C} \subset \mathcal{A}$ eine σ -Algebra und $A \in \mathcal{A}$. Dann heißt

$$\mathbb{P}^{\mathcal{C}}(A) := \mathbb{E}^{\mathcal{C}}(1_A) := \mathbb{P}(A|\mathcal{C}) = \mathbb{E}(1_A|\mathcal{C})$$

die **bedingte Wahrscheinlichkeit von A gegeben $\mathcal{C}$** . Für eine Zufallsvariable $Y : \Omega \rightarrow \mathbb{R}$ schreiben wir für $A \in \mathcal{A}$

$$\mathbb{P}(A|Y) := \mathbb{P}(A|\sigma(Y)).$$

$\mathbb{P}(A|\mathcal{C})$ ist also eine $\mathcal{C}$ messbare Funktion mit

$$\forall C \in \mathcal{C} : \int_C \mathbb{P}(A|\mathcal{C}) d\mathbb{P} = \int_C 1_A d\mathbb{P} = \mathbb{P}(A \cap C)$$

laut Definition der bedingten Erwartung von 1_A . Man kann zeigen, dass für eine Zufallsvariable $Y : \Omega \rightarrow \mathbb{R}$

$$\begin{aligned} \mathbb{P}(\cdot|Y=y) : \quad \mathcal{A} &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}(A|Y=y) \end{aligned}$$

für alle $y \in Y(\Omega)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$ ist. (Im Allgemeinen ist $\mathbb{P}(\cdot|\mathcal{C})(\omega)$, kein Wahrscheinlichkeitsmaß.)

Satz 6.6.3 Seien X, Y Zufallsvariablen, (X, Y) besitze die gemeinsame Dichte f , X sei integrierbar. Dann gilt:

$$\mathbb{P}(X \in B | Y = y) = \int_B f(x|y), \quad \forall B \in \mathcal{B}^1 \quad (6.26)$$

Ist $g : \mathbb{R} \rightarrow \mathbb{R}$ messbar und $g(X)$ integrierbar, dann gilt

$$\mathbb{E}(g(X)|Y=y) = \int_{\mathbb{R}} g(x)f(x|y)dx. \quad (6.27)$$

Sind X, Y unabhängig, $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ messbar und $h(X, Y)$ integrierbar, dann gilt

$$\mathbb{E}(h(X, Y)|Y=y) = \mathbb{E}(h(X, y)). \quad (6.28)$$

Beispiel 6.6.4 Wir betrachten die Situation von Beispiel 6.4.8, S. 86. Man bestimme

$$(a) \mathbb{P}(X \leq x | \Theta = \vartheta)$$

$$(b) \mathbb{E}(X^2 | \Theta = \vartheta)$$

Lösung Es gilt mit (6.26)

$$\mathbb{P}(X \leq x | \Theta = \vartheta) = \int_{-\infty}^x f(x | \vartheta) dx.$$

Da $f(\cdot | \vartheta)$ Dichte von $N(\vartheta, \sigma^2)$ ist, folgt

$$\mathbb{P}(X \leq x | \Theta = \vartheta) = \Phi\left(\frac{x - \vartheta}{\sigma}\right)$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$ ist. $\mathbb{E}(X^2 | \Theta = \vartheta)$ bestimmt sich ebenso aus den Eigenschaften der Normalverteilung. Es gilt wegen (6.27)

$$\mathbb{E}(X^2 | \Theta = \vartheta) = \int_{\mathbb{R}} x^2 f(x | \vartheta) dx$$

und somit

$$\mathbb{E}(X^2 | \Theta = \vartheta) = \sigma^2 + \mu^2.$$

□

Beispiel 6.6.5 Seien X_1, X_2 unabhängig, stetig verteilt mit Dichten f_1, f_2 . Man bestimme $\mathbb{P}(X_1 + X_2 \leq x)$.

Lösung Sei $h : \mathbb{R}^2 \rightarrow \mathbb{R}$, $h(x_1, x_2) := 1_{(-\infty, x]}(x_1 + x_2)$. Dann gilt mit Lemma 6.5.4, S. 88.

$$\begin{aligned} & \mathbb{P}(X_1 + X_2 \leq x) \\ &= \int_{\mathbb{R}} \mathbb{P}(X_1 + X_2 \leq x | X_2 = y) f_2(y) dy = \int_{\mathbb{R}} \mathbb{E}(h(X_1, X_2) | X_2 = y) f_2(y) dy \\ &\stackrel{(6.28)}{=} \int_{\mathbb{R}} \mathbb{E}(h(X_1, y)) f_2(y) dy = \int_{\mathbb{R}} \mathbb{P}(X_1 + y \leq x) f_2(y) dy \\ &= \int_{\mathbb{R}} \mathbb{P}(X_1 \leq x - y) f_2(y) dy = \int_{\mathbb{R}} \left(\int_{-\infty}^{x-y} f_1(t) dt \right) f_2(y) dy \\ &= \int_{\mathbb{R}} \left(\int_{-\infty}^x f_1(t - y) dt \right) f_2(y) dy = \int_{-\infty}^x \int_{\mathbb{R}} f_1(t - y) f_2(y) dy dt \\ &= \int_{-\infty}^x f_1 * f_2(t) dt \end{aligned}$$

Somit haben wir auf anderem Weg das schon bekannte Ergebnis erhalten, dass die Faltung $f_1 * f_2$ die Dichte von $X_1 + X_2$ ist. $\square$

6.7 Bedingte Unabhängigkeit

Nun folgt die Verallgemeinerung des Begriffs der Unabhängigkeit.

Definition 6.7.1 Seien X_1, X_2, Y Zufallsvariablen. Dann heißen X_1, X_2 **bedingt unabhängig gegeben** Y , wenn $X_1, X_2 \mathbb{P}(\cdot|Y = y)$ -unabhängig für alle $y \in Y(\Omega)$.

Beispiel 6.7.2 Betrachtet man ein Versicherungskollektiv, dann geht man davon aus, dass die Risiken unabhängig voneinander sind. Ob Schadenfälle eintreten hängt jedoch auch von wechselnden äußeren Bedingungen ab: Sterblichkeitsverbesserungen in der Lebensversicherung, Klimaerwärmung und Zunahme der Tropenstürme in der Sturmversicherung, Glatteis in der Autoversicherung etc. Die Risiken sind also unabhängig von einander, bis auf die äußeren Bedingungen.

Beispiel 6.7.3 (Fortsetzung Bsp. 6.4.8, S. 86) Seien $X_1, X_2 \sim N(\vartheta, \sigma^2)$ bedingt unabhängig gegeben $\Theta = \vartheta$ und $\Theta \sim N(\mu, \tau^2)$. Man bestimme die Kovarianz $\text{Cov}(X_1, X_2)$. Hinweis: (6.19), S. 81

Lösung $\text{Cov}(X_1, X_2) = \text{Var}(\Theta) = \tau^2$. $\square$

6.8 Gestutzte Verteilungen

In unterschiedlichen Versicherungszweigen begegnet man Selbstbehalten: Krankenversicherung, Kasko, Rückversicherung, ... Diese Selbstbehalte müssen bei der Tarifierung berücksichtigt werden.

Beispiel 6.8.1 Sei $C > 0$ und X eine Zufallsvariable mit der Verteilungsfunktion

$$F(x) = \begin{cases} 1 - \left(\frac{c}{x}\right)^2 & x \geq c \\ 0 & \text{sonst.} \end{cases}$$

Sei $a > c$. Definiere $Y := \min(X, a)$.

(a) Bestimme $\mathbb{P}(X = a)$, $\mathbb{P}(Y = a)$.

(b) Bestimme $\mathbb{E}(X)$, $\mathbb{E}(Y)$.

(c) Besitzen X bzw. Y Dichten?

(d) Man interpretiere die Ergebnisse.

Lösung (a) $\mathbb{P}(X = a) = 0$, $\mathbb{P}(Y = a) = \left(\frac{c}{a}\right)^2$.

(b) $\mathbb{E}(X) = 2c$, $\mathbb{E}(Y) = 2c - \frac{c^2}{a}$

(c) X besitzt eine Dichte, Y nicht, wegen $\mathbb{P}(Y = a) > 0$.

(d) Selbstbehalt = a . Entlastung beträgt c^2/a . $\square$

Sei $X : \Omega \rightarrow [0, \infty)$ eine integrierbare Zufallsvariable und $a \in \mathbb{R}$ mit $\mathbb{P}(X \geq a) > 0$. Die Verteilungsfunktion von X sei F . Dann gilt

$$\mathbb{E}(X - a | X > a) = \frac{\int_a^\infty (x - a) d\mathbb{P}_X(x)}{\mathbb{P}(X > a)} = \frac{\int_a^\infty (x - a) dF(x)}{1 - F(a)},$$

Dies folgt aus (6.2), S. 77 und (4.11), 39.

Interpretation

X Schaden Ist ein Schaden größer als a , dann gibt $\mathbb{E}(X - a | X > a)$ den mittleren Überschaden an.

X Lebensdauer Funktioniert ein System noch zum Zeitpunkt a , dann wird es im Mittel noch $\mathbb{E}(X - a | X > a)$ -lang weiterfunktionieren.

Beispiel 6.8.2 Sei $X \sim \exp(\lambda)$, $\lambda > 0$. Man bestimme $\mathbb{E}(X - a | x > a)$ für $a > 0$.

Lösung Die Dichte von X ist gegeben durch $\lambda e^{-\lambda x}$ für $x \geq 0$. Dann

$$\mathbb{E}(X - a | X > a) = \frac{\lambda \int_a^\infty (x - a) e^{-\lambda x} dx}{e^{-\lambda a}} = \lambda \int_a^\infty (x - a) e^{-\lambda(x-a)} dx = \frac{1}{\lambda}.$$

Dies entspricht der Eigenschaft der Gedächtnislosigkeit der Exponentialverteilung, die verbliebene Lebensdauer hängt nicht von a ab. $\square$

Betrachten wir nun

$$Y := \min(X, a),$$

dann gilt

$$\mathbb{E}(Y) = \int_0^a x d\mathbb{P}_X + a \cdot \mathbb{P}(X > a) = \int_0^a x d\mathbb{P}_X + a \cdot (1 - F(a))$$

denn mit (4.11)

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}(\min(X, a)) = \int \min(X, a) d\mathbb{P} = \int_{\mathbb{R}} \min(x, a) d\mathbb{P}(x) \\ &= \int_0^a x d\mathbb{P}_X + a \int_a^\infty d\mathbb{P}_X = \int_0^a x d\mathbb{P}_X + a(1 - F(a)). \end{aligned}$$

Interpretation

X Schaden Y ist der Schaden, der entsteht, wenn höchstens der Betrag a zu bezahlen ist.

X Lebensdauer Y ist die Lebensdauer, wenn man der Betrachtungszeitraum auf $[0, a]$ abgekürzt wird.

Beispiel 6.8.3 Sei $X \exp(\lambda)$, $\lambda > 0$. Man bestimme $\mathbb{E}(Y)$ für $a > 0$.

Lösung

$$\mathbb{E}(Y) = \frac{1}{\lambda} (1 - e^{-\lambda a}).$$

□

6.9 Hilfssätze

Lemma 6.9.1 Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, f, g seien integrierbar. Dann gilt:

- $\int_A f d\mu \leq \int_A g d\mu$ für alle $A \in \mathcal{A} \implies f \leq g$ f.ü.
- $\int_A f d\mu = \int_A g d\mu$ für alle $A \in \mathcal{A} \implies f = g$ f.ü.

Beweis (a) Sei $M := \{f > g\}$ und $M_n := \{f > g + \frac{1}{n}\}$ für $n \in \mathbb{N}$. Dann gilt $M, M_n \in \mathcal{A}$ und

$$\int_{M_n} f d\mu \geq \int_{M_n} g + \frac{1}{n} d\mu = \int_{M_n} g d\mu + \frac{1}{n} \mu(M_n) \geq \int_{M_n} f d\mu + \frac{1}{n} \mu(M_n).$$

Somit folgt $\mu(M_n) = 0$ für alle n . Wegen $M_n \uparrow M$ erhalten wir aus der Stetigkeit von unten (s. Satz 2.3.1, S. 9) $\mu(M) = 0$.

(b) folgt sofort aus (a). □

7 Grenzwertsätze

$(\Omega, \mathcal{A}, \mathbb{P})$ sei stets ein Wahrscheinlichkeitsraum.

In diesem Abschnitt werden zentrale Ergebnisse über das Verhalten der Summe bzw. des arithmetischen Mittels

$$\sum_{i=1}^n X_i \text{ bzw. } \bar{X}_{(n)} = \frac{1}{n} \sum_{i=1}^n X_i$$

vorgestellt, wenn n groß wird. Dabei sind die Zufallsvariablen X_i unabhängig und identisch verteilt (d.h. sie haben alle die gleiche Verteilungsfunktion).

Schreibweise: iid steht für unabhängig und identisch verteilt (aus dem Englischen independent and identically distributed).

7.1 Gesetze der großen Zahlen

Lemma 7.1.1 Seien $X_1, \dots, X_n$ iid mit $\mathbb{E}(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $\sigma > 0$. Es gilt

$$\mathbb{E}(\bar{X}_{(n)}) = \mu \quad (7.1)$$

$$\text{Var}(\bar{X}_{(n)}) = \frac{\sigma^2}{n}. \quad (7.2)$$

Beweis Es ergibt sich mit (4.14), S. 40

$$\begin{aligned} \mathbb{E}(\bar{X}_{(n)}) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n X_i\right) \stackrel{(4.14)}{=} \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{=\mu} \\ &= \frac{1}{n} \cdot \underbrace{(\mu + \dots + \mu)}_{n \text{ mal}} = \frac{1}{n} \cdot (n\mu) = \mu \end{aligned}$$

und mit (5.7), S. 71

$$\begin{aligned} \text{Var}(\bar{X}_{(n)}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \stackrel{(5.7)}{=} \frac{1}{n^2} \sum_{i=1}^n \underbrace{\text{Var}(X_i)}_{=\sigma^2} \\ &= \frac{1}{n^2} \cdot \underbrace{(\sigma^2 + \dots + \sigma^2)}_{n \text{ mal}} = \frac{1}{n^2} \cdot (n\sigma^2) = \frac{\sigma^2}{n}. \end{aligned}$$

□

An den Gleichungen (7.1) und (7.2) erkennt man, dass die Varianz von $\bar{X}_{(n)}$ mit wachsendem n abnimmt, zu erwarten ist also eine Konvergenz von $(\bar{X}_{(n)})$ gegen den Erwartungswert. Was bedeutet hier konvergent?

Satz 7.1.2 (Schwachtes Gesetz der großen Zahl) *Es sei $X_1, X_2, \dots$ iid mit Erwartungswert μ und Varianz σ^2 . Dann gilt*

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_{(n)} - \mu| \geq \varepsilon) = 0$$

Beweis Wegen (7.1) und (7.2) und der Tschebyschev Ungleichung (S. 43) folgt für $\varepsilon > 0$

$$\mathbb{P}(|\bar{X}_{(n)} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0 \quad (n \longrightarrow \infty).$$

□

Satz 7.1.3 (Starkes Gesetz der großen Zahl) *Sei $X_i, i \in \mathbb{N}$ eine Folge von iid Zufallsvariablen mit Erwartungswert μ . Dann gilt*

$$\mathbb{P}(\lim_{n \rightarrow \infty} (\bar{X}_{(n)} - \mu) = 0) = 1,$$

d.h. $\bar{X}_{(n)} \rightarrow \mu$ f.s.

Beweis [B-WT], S. 86. □

Anschaulich bedeuten die Sätze von der großen Zahl, dass wenn man oft genug ein Zufallsexperiment durchführt und das arithmetische Mittel aller Versuchsausgänge bildet, dem Erwartungswert μ der zugrunde liegenden Verteilung beliebig nahe kommt.

Von Bedeutung ist das zum Beispiel in der Versicherung: Ist der Versicherungsbestand groß genug und homogen, dann kann man durch Bildung des Mittelwerts der Schäden ziemlich sicher sein, dass die so berechnete Größe nahe am Erwartungswert liegt.

Zur Interpretation: Die Bezeichnung „schwach,, bzw. „stark“ bezieht sich auf die „Stärke“ der Konvergenz. Hierzu werden zwei Konvergenzbegriffe definiert.

Definition 7.1.4 *Sei $(Y_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen.*

(a) $(Y_n)_{n \in \mathbb{N}}$ konvergiert **stochastisch** gegen eine Zufallsvariable Y , wenn

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) = 0.$$

(b) $(Y_n)_{n \in \mathbb{N}}$ konvergiert **fast sicher** gegen eine Zufallsvariable Y , wenn

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1.$$

Fast sichere Konvergenz bedeutet, dass für fast alle ω die Folge $Y_n(\omega)$ gegen $Y(\omega)$ konvergiert, stochastische Konvergenz, dass Y_n mit wachsendem n mit großer Wahrscheinlichkeit in der Nähe von Y liegt. Über eine Konvergenz von $(Y_n(\omega))$ wird nichts gesagt.

Satz 7.1.5 *Fast sichere Konvergenz impliziert stochastische Konvergenz, aber nicht umgekehrt.*

Beweis [B-MT], S. 131, 20.5 Satz und S. 133, Bsp. 3. $\square$

Über die Konvergenzgeschwindigkeit lassen sich auch Aussagen treffen, siehe [B-WT], S. 93-96.

7.2 Der Zentrale Grenzwertsatz

Aus dem Gesetz der großen Zahlen wissen wir, dass für große n die Abweichungen des Mittelwerts $\bar{X}_{(n)}$ von μ mit großer Wahrscheinlichkeit sehr klein sind. Es ist bemerkenswert, dass diese Abweichungen wiederum näherungsweise normalverteilt sind mit

$$\mathbb{E}(\bar{X}_{(n)} - \mu) = 0, \text{Var}(\bar{X}_{(n)} - \mu) = \frac{\sigma^2}{n}.$$

Annähernd gilt also

$$\bar{X}_{(n)} - \mu \sim N\left(0, \frac{\sigma^2}{n}\right).$$

Um die allgemeinste Form des zentralen Grenzwertsatzes zu formulieren, benötigen wir die

Lindeberg-Bedingung: Sei $X_1, X_2, \dots$ eine Folge von unabhängigen Zufallsvariablen mit Erwartungswerten μ_n und Varianz σ_n^2 und

$$s_n := \sqrt{\sum_{i=1}^n \sigma_i^2}.$$

(s_n ist die Standardabweichung von $\sum_{i=1}^n X_i$). Für $\varepsilon > 0$ sei

$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x-\mu_i| > \varepsilon s_n} (x - \mu_i)^2 d\mathbb{P}_{X_i}(x). \quad (7.3)$$

Die Folge $X_1, X_2, \dots$ erfüllt die Lindeberg-Bedingung, wenn

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} L_n(\varepsilon) = 0. \quad (7.4)$$

Satz 7.2.1 (Zentraler Grenzwertsatz) Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit Erwartungswerten $\mathbb{E}(X_n) =: \mu_n$ und Varianz $\text{Var}(X_n) =: \sigma_n^2$. Genügt $(X_n)_{n \in \mathbb{N}}$ der Lindebergbedingung (7.4), dann gilt für jedes $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \leq x \right) = \Phi(x).$$

Φ ist die Verteilungsfunktion der Standardnormalverteilung also von $N(0, 1)$.

In dieser Form besagt der zentrale Grenzwertsatz, dass die Standardisierung von $S_n := \sum_{i=1}^n X_i$ für große N approximativ normalverteilt ist. Die Lindeberg Bedingung stellt sicher, dass keine der Variablen einen beherrschenden Einfluss auf die Verteilung der Summe S_n ausübt. Dies kann man an der äquivalenten Feller Bedingung sehen:

$$(X_n)_{n \in \mathbb{N}} \text{ erfüllt die Lindeberg Bedingung} \iff \lim_{n \rightarrow \infty} \left(\max_{1 \leq j \leq n} \frac{\sigma_j}{s_n} \right) = 0.$$

Korollar 7.2.2 Die Folge $(X_n)_{n \in \mathbb{N}}$ erfülle einer der folgenden Bedingungen:

- (a) X_n iid mit $\mathbb{E}(X_n) = \mu$, $\text{Var}(X_n) := \sigma^2$.
- (b) X_n erfüllt Lyapunov Bedingung: Es gibt $\delta > 0$ mit

$$\lim_{n \rightarrow \infty} s_n^{-2-\delta} \sum_{i=1}^n \mathbb{E}(|X_i - \mu_i|^{2+\delta}) = 0.$$

Dann erfüllt $(X_n)_{n \in \mathbb{N}}$ die Lindebergbedingung, es gilt also

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \leq x \right) = \Phi(x).$$

Beweis (a) X_i iid: Es gilt $s_n^2 = n\sigma^2$ und somit wenn $\mathbb{P}_{X_i} =: \mathbb{P}_X$

$$\begin{aligned} 0 \leq L_n(\varepsilon) &= \frac{1}{n\sigma^2} \sum_{i=1}^n \int_{\{|x-\mu| > \varepsilon\sigma\sqrt{n}\}} (x - \mu)^2 d\mathbb{P}_X(x) \\ &= \frac{1}{\sigma^2} \int_{\{|x-\mu| > \varepsilon\sigma\sqrt{n}\}} (x - \mu)^2 d\mathbb{P}_X(x). \end{aligned}$$

Die Behauptung folgt mit dem Satz von der monotonen Konvergenz angewendet auf $f_n(x) := (x - \mu)^2 \cdot 1_{\{|x - \mu| > \varepsilon \sigma \sqrt{n}\}}$.

(b) Lyapunov Bedingung: Es gilt

$$|x - \mu_i| \geq \varepsilon s_n \implies |x - \mu_i|^{2+\delta} \geq |x - \mu_i|^2 (\varepsilon s_n)^\delta \implies |x - \mu_i|^2 \leq \frac{|x - \mu_i|^{2+\delta}}{(\varepsilon s_n)^\delta}$$

und somit

$$\begin{aligned} L_n(\varepsilon) &= \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x - \mu_i| > \varepsilon s_n} (x - \mu_i)^2 d\mathbb{P}_{X_i}(x) \\ &\leq \frac{1}{(\varepsilon s_n)^\delta s_n^2} \sum_{i=1}^n \int_{|x - \mu_i| > \varepsilon s_n} |x - \mu_i|^{2+\delta} d\mathbb{P}_{X_i}(x) \\ &\leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{i=1}^n \int_{\mathbb{R}} |x - \mu_i|^{2+\delta} d\mathbb{P}_{X_i}(x) \longrightarrow 0. \end{aligned}$$

□

Bemerkung Der Zentrale Grenzwertsatz liefert bezüglich der Binomialverteilung eine gute Approximation für $n \cdot p \geq 5$ und $n \cdot (1 - p) \geq 5$. Eine etwas genauere „korrigierte“ Formel bezüglich der Binomialverteilung ist übrigens

$$\mathbb{P}(k \leq X \leq l) \approx \Phi\left(\frac{l - np + 0,5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - np - 0,5}{\sqrt{np(1-p)}}\right).$$

Die hier bereitgestellten Ergebnisse findet man mit Beweisen in [B-WT], S. 237 ff.

7.3 Der Zentralsatz der Statistik

Das nun folgende Ergebnis ist auch als Satz von Glivenko Cantelli bekannt und bildet die Grundlage der Statistik.

7.3.1 Empirische Verteilungsfunktion

Definition 7.3.1 Für eine Messreihe $x_1, \dots, x_n$ heißt $F_n(\cdot; x_1, \dots, x_n) : \mathbb{R} \longrightarrow [0, 1]$ definiert durch

$$\begin{aligned} F_n(z; x_1, \dots, x_n) &:= \frac{1}{n} \cdot \text{Anzahl der } x_i \leq z \\ &:= \frac{1}{n} |\{i | x_i \leq z, 1 \leq i \leq n\}| \end{aligned}$$

die **empirische Verteilungsfunktion** der Messreihe $x_1, \dots, x_n$.

Sind $X_1, \dots, X_n$ Zufallsvariablen, dann ist $F_n(z; X_1, \dots, X_n)$ eine Zufallsvariable. Wir interessieren uns nun für die empirische Verteilungsfunktion, wenn die $X_1, \dots, X_n$ iid. sind.

7.3.2 Satz von Glivenko Cantelli

Satz 7.3.2 (Glivenko-Cantelli) Seien $X_1, \dots, X_n$ iid. mit Verteilungsfunktion F . $F_n : \mathbb{R} \rightarrow [0, 1]$ sei die empirische Verteilungsfunktion von $X_1, \dots, X_n$ und

$$D_n := D_n(X_1, \dots, X_n) := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Dann gilt

$$\mathbb{P}(\lim_{n \rightarrow \infty} D_n = 0) = 1$$

d.h. $D_n \rightarrow 0$ $\mathbb{P}_F$ f.s.

Beweis [P], S. 156, Proposition. □

Eine genauere Aussage bewies Kolmogorov unter der zusätzlichen Voraussetzung der Stetigkeit von F :

Satz 7.3.3 (Kolmogorov) Seien $X_1, \dots, X_n$ iid. mit stetiger Verteilungsfunktion F . $F_n : \mathbb{R} \rightarrow [0, 1]$ sei die empirische Verteilungsfunktion von $X_1, \dots, X_n$ und

$$D_n(X_1, \dots, X_n) = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n} D_n(X_1, \dots, X_n) \leq y) = K(y), \quad y \in \mathbb{R}$$

wobei $K : \mathbb{R} \rightarrow [0, 1]$ gegeben ist durch

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

K heißt **Kolmogorowsche Verteilungsfunktion**

Beweis [GS], S. 152. □

Teil III

Statistik

Stichprobe, Grundgesamtheit

Der zentrale Begriff der Statistik ist der einer **Stichprobe**, je nach Situation auch Messreihe genannt. Besteht die Stichprobe aus n Zahlen, so geben wir diese in der Form

$$x = (x_1, \dots, x_n)$$

an.

Die schließende Statistik, auch Inferenzstatistik genannt, interpretiert diese Stichprobe als das Ergebnis eines durchgeführten Zufalls-Experimentes. Die Werte $x_1, \dots, x_n$ sind nach dieser Vorstellung zufällig entstanden. Sie könnten bei einer anderen Durchführung des Zufallsexperiments anders ausfallen. Eine solche Durchführung stellt man sich vor als das zufällige Ziehen von Werten aus einer **Grundgesamtheit**. Von der Grundgesamtheit macht sich der Statistiker eine Vorstellung in Form eines Modells, das angibt, wie die Werte in ihr verteilt sind. Solche Verteilungsannahmen beinhalten gewisse Parameter ϑ , deren (wahre) Werte unbekannt sind, über die gerade die Stichprobe einen Rückschluss (Inferenz) geben soll. In den nächsten Kapiteln werden Methoden vorgestellt um

- Parameter ϑ zu schätzen (Kapitel 8),
- Konfidenzbereiche für ϑ zu konstruieren (Kapitel 9) und
- Hypothesen über ϑ zu testen. (Kapitel 10)

Einteilen der Verfahren

Bei den **Ein-Stichproben Verfahren** gehen wir von n (unabhängig voneinander erhobenen) Werten

$$x_1, \dots, x_n$$

derselben Messgröße X aus (statt Messgröße sagt man auch: **Merkmal, Variable**).

Sind diese Werte in zwei verschiedenen Gruppen (Grundgesamtheiten) erhoben worden, d. h. liegen zwei Messreihen

$$x_1, \dots, x_{n_1}, \quad y_1, \dots, y_{n_2}$$

derselben Messgröße X vor, so befinden wir uns in der **Zwei-Stichproben** Situation.

Schließlich wird der Fall behandelt, dass eine Messreihe von Wertepaaren (x, y) zweier verschiedener Messgrößen X und Y vorliegt. Eine solche **bivariate Stichprobe** lässt sich in der Form

$$(x_1, y_1), \dots, (x_n, y_n)$$

schreiben.

Skalen

Variablen (Messgrößen) können unterschiedlicher Skalennatur sein, und danach richten sich auch die statistischen Verfahren, die zur Anwendung kommen. An Skalen unterscheiden wir:

Nominalskala Eine solche Skala besteht aus endlich vielen, wechselseitig sich ausschließenden Kategorien (Ausprägungen). Zu je zwei Werten a und b einer nominal-skalierten Variablen lässt sich nur feststellen, ob sie der gleichen Kategorie angehören oder nicht, sinnlos ist ein Größenvergleich oder ihre Differenz. Sind nur zwei Kategorien möglich, nennt man solche Variablen auch **dichotom**.

Beispiel: Merkmal: Geschlecht, Skala: männlich weiblich.

Ordinalskala Eine solche Skala besteht aus Werten, die in einer Ordnungsrelation zueinander stehen: Für je zwei verschiedene Werte a und b lässt sich angeben, ob $a < b$ oder $b < a$ gilt. Eine Differenz $a - b$ ist jedoch nicht sinnvoll.

Beispiel: Skala der Schulnoten: 1, 2, 3, 4, 5, 6.

Metrische Skala Dies ist eine physikalische Mess-Skala. Für je zwei verschiedene Werte a und b lässt sich angeben, ob $a < b$ oder $b < a$ gilt und die Differenz berechnen.

Beispiel: Temperaturskala 0 - 100 Grad Celsius.

Nominal- und ordinalskalierte Variablen nennt man auch **kategoriale** oder **diskrete Variable**.

8 Punktschätzer

Wir betrachten stets eine Messreihe $x_1, \dots, x_n$, die Realisierungen der iid Zufallsvariablen $X_1, \dots, X_n$ sind. Ihre Verteilungsfunktion F_ϑ hänge von $\vartheta \in \Theta \subset \mathbb{R}^d$ ab. Ziel ist es aus der Messreihe den Parameter ϑ zu schätzen.

8.1 Begriff

Wir haben einige Verteilungsfunktionen kennen gelernt: Binomialverteilung, Poissonverteilung, Exponentialverteilung, Normalverteilung. Bei der Definition wurden jeweils Parameter verwendet, die diese Verteilung vollständig beschreiben:

Verteilung	Anzahl Parameter	Parameter
Binomial	1	p
Negativ Binomial	2	m, p
Poisson	1	λ
Gleich-Verteilung	2	a, b
Exponential	1	λ
Normal	2	μ, σ
Gamma	2	a, λ

Je nach Fragestellung interessiert man sich für den Verteilungstyp und die dazugehörigen Parameter bzw. Erwartungswert, Varianz, Schiefe etc. In diesem Abschnitt gehen wir davon aus, dass der Verteilungstyp bekannt ist. Die Parameter müssen aus der gegebenen Messreihe berechnet werden. Für ϑ benötigen wir also eine Funktion $T : \mathbb{R}^n \longrightarrow \Theta \subset \mathbb{R}^d$. Diese wird jedes mal verwendet um den Parameter ϑ aus der Messreihe (Stichprobe) zu schätzen. In der Folge sei $X := (X_1, \dots, X_n)$, also die mehrdimensionale Zufallsvariable mit den Komponenten X_i .

Definition 8.1.1 Ein **Schätzer** (auch **Punktschätzer**) für ϑ ist eine messbare Abbildung $T : (\mathbb{R}^n, \mathcal{B}^n) \longrightarrow (\mathbb{R}^d, \mathcal{B}^d)$ die nicht von ϑ abhängt. Eine Realisation

$$\hat{\vartheta} := T(x_1, \dots, x_n)$$

heißt **Schätzung** (auch **Schätzwert**).

Eine Schätzung $\hat{\vartheta}$ kann als Realisierung der entsprechenden Zufallsvariablen $T(X)$ aufgefasst werden. $T(X)$ nennen wir auch Schätzer (genauer wäre

Schätzvariable). Seine Verteilung und die existierenden Maßzahlen hängen von F_ϑ ab. Um diese Abhängigkeit anzudeuten schreiben wir

$$\mathbb{E}_\vartheta(\cdot), \text{Var}_\vartheta(\cdot), \dots$$

8.2 Eigenschaften von Schätzern

Der Einfachheit halber betrachten wir den Fall $d = 1$.

Definition 8.2.1 Sei T ein Schätzer für ϑ .

- (a) T bzw. $T(X)$ heißt **erwartungstreu** für einen Parameter ϑ , wenn für alle $\vartheta \in \Theta$

$$\mathbb{E}_\vartheta(T(X)) = \vartheta$$

gilt.

- (b) Die Größe

$$\mathbb{E}_\vartheta(T(X)) - \vartheta$$

heißt **Bias** (auch **Verzerrung**) von T .

- (c) Der **mittlere quadratische Fehler** (mean squared error) von T wird definiert durch

$$\text{MSE}(T(X)) := \mathbb{E}_\vartheta([T(X) - \vartheta]^2)$$

Satz 8.2.2 Sei T ein Schätzer für ϑ . Dann gilt

$$\text{MSE}(T(X)) = \text{Var}_\vartheta(T(X)) + [\mathbb{E}_\vartheta(T(X)) - \vartheta]^2. \quad (8.1)$$

Beweis

$$\begin{aligned} & \text{Var}_\vartheta(T(X)) + [\mathbb{E}_\vartheta(T(X)) - \vartheta]^2 \\ &= \mathbb{E}_\vartheta(T(X)^2) - \mathbb{E}_\vartheta(T(X))^2 + \mathbb{E}_\vartheta(T(X))^2 - 2\mathbb{E}_\vartheta(T(X))\vartheta + \vartheta^2 \\ &= \mathbb{E}_\vartheta(T(X)^2 - 2\vartheta T(X) + \vartheta^2) = \mathbb{E}_\vartheta[(T(X) - \vartheta)^2]. \end{aligned}$$

□

Definition 8.2.3 Sei $(T_n)_{n \in \mathbb{N}}$ eine Folge von Schätzern für ϑ .

- (a) Die Folge heißt **asymptotisch erwartungstreu**, wenn für alle $\vartheta \in \Theta$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{E}_\vartheta(T_n(X)) = \vartheta$$

(b) Die Folge heißt **konsistent**, wenn gilt:

$$T_n(X) \xrightarrow{n \rightarrow \infty} \vartheta \text{ stochastisch.}$$

(c) Die Folge heißt **asymptotisch normalverteilt**, wenn für alle $\vartheta \in \Theta$

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\vartheta} \left(\frac{T_n(X) - \mathbb{E}_{\vartheta}(T_n(X))}{\sqrt{\text{Var}_{\vartheta}(T_n(X))}} \leq x \right) = \Phi(x), \quad x \in \mathbb{R}.$$

(die Standardisierungen von $T_n(X)$ erfüllen den Zentralen Grenzwertsatz.)

Satz 8.2.4 Sei $(T_n)_{n \in \mathbb{N}}$ eine asymptotisch erwartungstreue Folge von Schätzern für ϑ mit

$$\lim_{n \rightarrow \infty} \text{Var}_{\vartheta}(T_n(X)) = 0.$$

Dann ist $(T_n)_{n \in \mathbb{N}}$ konsistent.

Beweis Sei $\varepsilon > 0$. Wie beim Beweis der Tschebyschev'schen Ungleichung gilt

$$\begin{aligned} \mathbb{P}_{\vartheta}(|T_n(X) - \vartheta| \geq \varepsilon) &\leq \frac{E_{\vartheta}(|T_n(X) - \vartheta|^2)}{\varepsilon^2} \\ &\stackrel{(8.1)}{=} \frac{1}{\varepsilon^2} (\text{Var}_{\vartheta}(T_n(X)) + [\mathbb{E}_{\vartheta}(T_n(X)) - \vartheta]^2) \longrightarrow 0 \end{aligned}$$

□

8.3 Schätzen von Erwartungswert und Varianz

Satz 8.3.1 Gegeben seien iid Zufallsvariablen $X_1, \dots, X_n$ mit $\mathbb{E}(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $\sigma > 0$ und $\mu_4 := \mathbb{E}((X - \mu)^4)$.

$$\begin{aligned} \bar{X}_n &:= \frac{1}{n} \sum_{i=1}^n X_i && (\text{Mittelwert}) \\ S_n^2 &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 && (\text{empirische Varianz}) \end{aligned}$$

(a) $\bar{X}_n$ ist erwartungstreu für den Erwartungswert μ .

(b) S_n^2 ist erwartungstreu für die Varianz σ^2 .

(c) Es gilt

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \quad (8.2)$$

$$\text{Var}(S_n^2) = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) \quad (8.3)$$

(d) $\bar{X}_n$ und S_n^2 sind konsistent.

Beweis (a)-(c) [P], S. 19.

(d) folgt aus (c) mit Satz 8.2.4. $\square$

Bemerkung 8.3.2 (a) Der Schätzer für die Varianz enthält hier einen Schätzer, nämlich den Schätzer $\bar{X}_n$ für den Erwartungswert.

(b) $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ist asymptotisch erwartungstreu, jedoch nicht erwartungstreu.

(c) Für bekannten Erwartungswert μ ist $T(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ ein erwartungstreuer Schätzer für die Varianz σ^2 :

$$E(T(X)) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right) = \frac{1}{n} \sum_{i=1}^n \underbrace{E((X_i - \mu)^2)}_{=\sigma^2} = \sigma^2$$

(d) Mit Hilfe der $X_1, \dots, X_n$ kann man auch die Varianz von $\bar{X}_n$ und S_n^2 schätzen, indem man in (8.2) und (8.3) σ^2 und μ_4 durch ihre Schätzer S_n^2 und

$$\hat{\mu}_4 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^4$$

ersetzt, also

$$\begin{aligned} \widehat{\text{Var}(\bar{X}_n)} &= \frac{S_n^2}{n} \\ \widehat{\text{Var}(S_n^2)} &= \frac{1}{n} \left(\hat{\mu}_4 - \frac{n-3}{n-1} S_n^4 \right). \end{aligned}$$

Aufgabe 8.3.3 Eine Firma produziert in zwei Werken. Die Firmenleitung will die Akzeptanz einer neuen Arbeitszeitregelung in der Belegschaft abschätzen und zwar in beiden Werken einzeln und in der Gesamtbelegschaft. Dazu wird eine Befragung durchgeführt.

In Werk 1 arbeiten 2200 Beschäftigte, in Werk 2 arbeiten 600. Die Befragung von Stichproben ergibt folgende Werte:

	Stichprobenumfang	Zustimmung	Schätzwerte
Werk 1	$n_1 = 70$	42	$\hat{p}_1 =$
Werk 2	$n_2 = 30$	25	$\hat{p}_2 =$

- (a) Man gebe Schätzwerte $\hat{p}_1, \hat{p}_2$ für die Zustimmungsrate p_1 bzw. p_2 in den Werken 1 und 2 an.
- (b) Gegeben seien folgende Schätzer für die Zustimmungsrate p der Gesamtbelegschaft

$$\begin{aligned}\hat{p} &= \frac{1}{n_1 + n_2}(n_1\hat{p}_1 + n_2\hat{p}_2) \\ \hat{p}' &= \frac{1}{2}(\hat{p}_1 + \hat{p}_2) \\ \hat{p}'' &= \frac{1}{2800}(2200\hat{p}_1 + 600\hat{p}_2).\end{aligned}$$

Berechnen Sie die Schätzwerte für die vorliegende Stichprobe.

- (c) Welchen der drei Schätzwerte bevorzugen Sie? Welche der vorgeschlagenen Schätzer sind erwartungstreu?

8.4 Momenten-Schätzer

Die „einfachste“ Art der Parameterschätzung ist die Momentenschätzung. Für $r \in \mathbb{N}$ ist das r -te Moment einer Zufallsvariable X

$$m_r = \mathbb{E}_\vartheta(X^r) = \int X^r d\mathbb{P}.$$

Die Momente hängen von den Parametern der Verteilungsfunktion ab. Der Schätzer für m_r ist

$$\hat{m}_r := \frac{1}{n} \sum_{i=1}^n x_i^r$$

und heißt **Momentenschätzer**. Momentenschätzer sind erwartungstreu, konsistent und wegen des Zentralen Grenzwertsatzes asymptotisch normalverteilt.

Will man nun k Parameter $\vartheta_1, \dots, \vartheta_k$ der Verteilungsfunktion F schätzen, so reicht es das Gleichungssystem

$$\begin{aligned}\hat{m}_1 &= \mathbb{E}_{\vartheta_1, \dots, \vartheta_k}(X) \\ \hat{m}_2 &= \mathbb{E}_{\vartheta_1, \dots, \vartheta_k}(X^2) \\ \dots &= \dots \\ \hat{m}_k &= \mathbb{E}_{\vartheta_1, \dots, \vartheta_k}(X^k)\end{aligned}$$

nach $\vartheta_1, \dots, \vartheta_k$ aufzulösen.

Beispiel 8.4.1 Man bestimme durch die Momentenmethode Schätzer für die Parameter $a, b \in \mathbb{R}$, $a < b$ einer Gleichverteilung für eine Stichprobe vom Umfang n .

Lösung Es gilt

$$\begin{aligned}m_1 &= \frac{a+b}{2} \\ m_2 &= \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}.\end{aligned}$$

Löst man dieses Gleichungssystem nach a, b auf, ergibt sich

$$\begin{aligned}a &= m_1 - \sqrt{3(m_2 - m_1^2)} \\ b &= m_1 + \sqrt{3(m_2 - m_1^2)}.\end{aligned}$$

Nun können wir mit

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left(\sum_{i=1}^n x_i^2 \right) - \bar{x}^2 = \hat{m}_2 - \hat{m}_1^2$$

die Schätzer definieren:

$$\begin{aligned}\hat{a} &= \hat{m}_1 - \sqrt{3(\hat{m}_2 - \hat{m}_1^2)} = \bar{x} - \sqrt{3s_n^2} \\ \hat{b} &= \hat{m}_1 + \sqrt{3(\hat{m}_2 - \hat{m}_1^2)} = \bar{x} + \sqrt{3s_n^2}.\end{aligned}$$

□

Aufgabe 8.4.2 Jede Wartezeit vor einer Ampel ist gleich wahrscheinlich. Wir halten folgende Wartezeiten in Sekunden fest: 15, 20, 25, 30, 31, 33. Man schätze mit der Momenten-Methode die Mindest- und Höchstwartezeit vor dieser Ampel.

8.5 Das Maximum-Likelihood-Prinzip

Bei diesem Prinzip wird der zu schätzende Parameter so geschätzt, dass die Wahrscheinlichkeit der beobachteten Stichprobe maximiert wird. Es geht also die Wahrscheinlichkeits- bzw. Dichtefunktion f_{ϑ} mit ein, die bis auf einen (oder mehrere) Parameter als bekannt vorausgesetzt wird.

Aufgabe 8.5.1 *Ein Produzent entnimmt zufällig aus der Produktion 100 Sicherungen und stellt fest, dass 5 defekt sind. Der Produzent will den Ausschussanteil p an der Produktion feststellen.*

- (a) *Man bestätige folgende Wahrscheinlichkeiten. Man kann von einer $B(100, p)$ -Verteilung ausgehen.*

p	$\mathbb{P}(5 \text{ Sicherungen defekt})$
0,02	0,035
0,03	0,101
0,04	0,160
0,05	0,180
0,06	0,164
0,07	0,128

- (b) *Welcher Wert für p ist der „beste“?*

Für die Zufallsvariablen $X_1, \dots, X_n$ möge wieder die Verteilungsfunktion F_{ϑ} in Frage kommen. Sind die X_i stetig verteilt bzw. diskret, bezeichnen wir mit f_{ϑ} die Dichte bzw. die Wahrscheinlichkeitsfunktion von X_i . Da die $X_1, \dots, X_n$ iid sind, ist

$$(x_1, \dots, x_n) \longmapsto \prod_{i=1}^n f_{\vartheta}(x_i)$$

die gemeinsame Dichte bzw. Wahrscheinlichkeitsfunktion von $X = (X_1, \dots, X_n)$.

Definition 8.5.2 (Maximum-Likelihood-Prinzip) *Sei $x_1, \dots, x_n$ eine Realisation von $X = (X_1, \dots, X_n)$.*

- (a) $L := L(\vartheta; x_1, \dots, x_n) : \Theta \longrightarrow \mathbb{R}$,

$$L(\vartheta) = L(\vartheta; x_1, \dots, x_n) := \prod_{i=1}^n f_{\vartheta}(x_i)$$

heißt **Likelihood(funktion) zur Realisierung** $(x_1, \dots, x_n)$.

(b) $\hat{\vartheta}$ heißt **Maximum Likelihood Schätzer (ML Schätzer)** für ϑ , wenn er die Likelihood maximiert, also

$$L(\hat{\vartheta}) = \max_{\vartheta \in \Theta} L(\vartheta).$$

Bei der konkreten Bestimmung von ML Schätzern, wird oft die Log-Likelihood eingesetzt:

$$l(\vartheta) := l(\vartheta; x_1, \dots, x_n) := \ln L(\vartheta) = \sum_{i=1}^n \ln f_{\vartheta}(x_i).$$

Oft sind ML-Schätzer asymptotisch erwartungstreu, konsistent und asymptotisch normalverteilt.

Beispiel 8.5.3 Eine Dartscheibe wurde bei 20 Würfeln 15 mal getroffen. Welches ist ein ML-Schätzer für die Trefferwahrscheinlichkeit p ? (Wir erwarten $\hat{p} = 15/20 = 0,75$.)

Lösung Wir gehen von einer $B(1, p)$ Verteilung mit $n = 20$ aus, also lautet die Likelihood

$$L(p) := p^{15}(1-p)^5, \quad p \in (0, 1)$$

Wir suchen nun das Maximum von L . Anstelle von L maximieren wir hier $l := \ln L$,

$$l(p) = 15 \ln p + 5 \ln(1-p).$$

Leiten wir nach p ab, folgt

$$\begin{aligned} l'(p) &= \frac{15}{p} - \frac{5}{1-p} = \frac{15 - 15p - 5p}{p(1-p)} = \frac{15 - 20p}{p(1-p)} \\ l''(p) &= -\frac{15}{p^2} - \frac{5}{(1-p)^2} < 0. \end{aligned}$$

Setzt man die erste Ableitung $= 0$, folgt $\hat{p} = \frac{15}{20}$, und wegen

$$l''(15/20) < 0$$

folgt das Gewünschte. □

Lemma 8.5.4 Seien $X_1, \dots, X_n$ iid mit einer $B(m, p)$ Verteilung und $x_1, \dots, x_n$ eine Realisierung. Dann ist die relative Häufigkeit von Treffern

$$\hat{p} := \frac{\sum_{i=1}^n x_i}{nm}$$

ML-Schätzer für p .

Beweis Für die Log Likelihood l gilt

$$\begin{aligned} l(p) &= \sum_{i=1}^n \left[\ln \binom{m}{x_i} + x_i \ln p + (m - x_i) \ln(1 - p) \right] \\ l'(p) &= \sum_{i=1}^n \left[\frac{x_i}{p} - \frac{m - x_i}{1 - p} \right] \\ l''(p) &= - \sum_{i=1}^n \left[\frac{x_i}{p^2} - \frac{m - x_i}{(1 - p)^2} \right] < 0 \end{aligned}$$

Die Behauptung folgt nun durch Nullsetzen der zweiten Gleichung. $\square$

Beispiel 8.5.5 Die Lebensdauer von Glühbirnen einer Firma sei exponentialverteilt mit unbekanntem λ . Es wurden Lebensdauern von 5 mal 1 Jahr, 3 mal 2 Jahren und 2 mal 3 Jahren in einer Stichprobe festgestellt. Wir suchen einen ML-Schätzer für λ . Wir können annehmen, dass die Glühbirnen unabhängig voneinander sind.

Lösung Aufgrund der Unabhängigkeit der Lebensdauer der einzelnen Glühbirnen gilt für die Likelihood

$$L(\lambda; x_1, \dots, x_n) = \lambda e^{-\lambda x_1} \cdot \dots \cdot \lambda e^{-\lambda x_n} = \lambda^n \exp \left(-\lambda \sum_{i=1}^n x_i \right).$$

Analog zu vorher ergibt sich

$$\begin{aligned} l(\lambda) &:= \ln L(\lambda) = \ln(\lambda^n) - \lambda \sum_{i=1}^n x_i = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i \\ l'(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^n x_i \\ l''(\lambda) &= -\frac{n}{\lambda^2} < 0 \end{aligned}$$

und somit

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{10}{17}.$$

$\square$

Bemerkung Der Kehrwert der Stichprobenmittels

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}.$$

ist also der ML-Schätzer für den Parameter λ der Exponentialverteilung.

Beispiel 8.5.6 Sind $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ iid, erhält man den ML-Schätzer $\bar{X}_n$ für den Erwartungswert und

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

für die Standardabweichung.

Beweis [BM], S. 247, Beispiel 3.10. □

Satz 8.5.7 Ist $g : \mathbb{R}^d \rightarrow \mathbb{R}^c$ messbar und ist $\hat{\vartheta} \in \mathbb{R}^d$ ein ML-Schätzer für ϑ , so ist $g(\hat{\vartheta})$ ein ML-Schätzer für $h(\vartheta)$.

Beweis [P], S. 25, Satz. □

Beispiel 8.5.8 (zu Satz 8.5.7) • Ist $\hat{\sigma}$ ein ML-Schätzer für die Standardabweichung, so ist $\hat{\sigma}^2$ ein ML-Schätzer für die Varianz: $g(x) := x^2$ ist messbar.

- Im vorangegangenen Beispiel 8.5.5 (Glühbirnen) berechnen sich die Schätzer für Erwartungswert und Varianz der Glühbirnenlebensdauer zu $\frac{17}{10}$ für den Erwartungswert ($g(x) = \frac{1}{x}$) und für die Varianz $1,7^2$ ($h(x) = \frac{1}{x^2}$), da der Erwartungswert bzw. die Varianz bei der Exponentialverteilung $\frac{1}{\lambda}$ bzw. $\frac{1}{\lambda^2}$ betragen.

8.6 Zusammenfassung

Wir stellen hier noch einmal die Schätzer dieses Kapitels und ihre Eigenschaften zusammen.

Wir gehen weiter davon aus, dass jeweils eine Stichprobe $x_1, \dots, x_n$ vom Umfang n vorliegt, $\bar{x}$ sei der Mittelwert der Stichprobe.

Vert.	Param.	Schätzer	Eigenschaften	Bem.
$B(1, p)$	p	$\frac{k}{n}$	erwartungstreu ML	k Anzahl der Treffer
$\text{Poi}(\lambda)$	λ	$\frac{1}{n} \sum_{i=1}^n x_i$	erwartungstreu ML	$\mu = \lambda$
$\text{Exp}(\lambda)$	λ	$n / \sum_{i=1}^n x_i$	ML	
$N(\mu, \sigma)$	μ	$\frac{1}{n} \sum_{i=1}^n x_i$	erwartungstreu ML	
$N(\mu, \sigma)$	σ^2	$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	erwartungstreu	μ unbekannt
$N(\mu, \sigma)$	σ^2	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	ML	μ unbekannt
$N(\mu, \sigma)$	σ	$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$	ML	μ unbekannt
alle	μ	$\frac{1}{n} \sum_{i=1}^n x_i$	erwartungstreu	
alle	σ^2	$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	erwartungstreu	μ unbekannt

9 Intervallschätzer

Da die im letzten Kapitel betrachteten Schätzwerte $T(x_1, \dots, x_n)$ in der Regel vom zu schätzenden Parameter ϑ abweichen, stellt sich in natürlicher Weise die Frage nach Schranken, zwischen denen der unbekannte Parameter mit einer gewissen Sicherheit liegt. Wir sind genauer an einem Verfahren interessiert,

- mit dem sich aus den Beobachtungen $x_1, \dots, x_n$ ein (möglichst kleines) Intervall

$$I(x_1, \dots, x_n) = [U(x_1, \dots, x_n), O(x_1, \dots, x_n)]$$

berechnen lässt

- das zu einem vorgegebenen $\alpha \in]0, 1[$ mit einer Wahrscheinlichkeit von $1 - \alpha$ ein Intervall liefert, das den wahren Parameter ϑ enthält.

Wir gehen in diesem Kapitel **generell** von $X_1, \dots, X_n$ iid mit einer Verteilungsfunktion F_ϑ aus. Der „wahre“ Parameterwert ϑ ist unbekannt, es gilt jedoch $\vartheta \in \Theta \subset \mathbb{R}^d$. Weiter sei $\mathbb{E}(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$ und eine Realisation $x_1, \dots, x_n$ vom Umfang n . Wir setzen $X = (X_1, \dots, X_n)$.

Die Endpunkte eines „zufälligen Intervalls $I(X)$ “ werden durch Zufallsvariablen $U(X) \leq O(X)$ beschrieben also

$$I(X) = [U(X), O(X)].$$

O und U sollen so bestimmt werden, dass

$$\forall \vartheta \in \Theta : \quad \mathbb{P}_\vartheta(U(X) \leq \vartheta \leq O(X)) = 1 - \alpha$$

für $\alpha \in]0, 1[$ gilt. Dieses Intervall $I(X)$ heißt **Konfidenzintervall zum Niveau $1 - \alpha$** für ϑ . Typische Werte von α sind 5 %, 1 %,

9.1 Testverteilungen

In diesem Abschnitt werden einige spezielle Verteilungen zusammengestellt, die bei der Konstruktion von Konfidenzintervallen und beim Testen von Hypothesen wichtig sind. Mit $N(0, 1)$ und u_γ werden die Standardnormalverteilung und ihr γ -Quantil ($0 < \gamma < 1$) bezeichnet.

9.1.1 χ^2 -Verteilung

Definition 9.1.1 Eine Zufallsvariable mit der Dichte

$$f(x) = \begin{cases} 0 & \text{falls } x \leq 0, \\ K_m x^{\frac{m-2}{2}} e^{-\frac{x}{2}} & \text{falls } x > 0, \end{cases}$$

wobei $m \in \mathbb{N}$ und K_m die Konstante

$$K_m := \frac{1}{2^{m/2} \Gamma\left(\frac{m}{2}\right)} \quad (\Gamma \text{ die Gammafunktion})$$

ist, heißt χ^2 -**verteilt** (chi-Quadrat verteilt) mit m Freiheitsgraden (FG) oder kurz χ_m^2 -verteilt.

Die χ_m^2 Verteilung ist ein Spezialfall der Gamma-Verteilung, es gilt $\chi_m^2 = \Gamma\left(\frac{m}{2}, \frac{1}{2}\right)$

Den folgenden wichtigen Zusammenhang mit normalverteilten Zufallsvariablen kann man auch als Definition der χ_m^2 -Verteilung verwenden.

Satz 9.1.2 Sind $Z_1, \dots, Z_m$ iid, $N(0, 1)$ -verteilt, dann gilt

$$\sum_{i=1}^n Z_i^2 \sim \chi_m^2.$$

Satz 9.1.3 Sind die Zufallsvariablen X_1, X_2 unabhängig, $X_i \sim \chi_{m_i}^2$, $i = 1, 2$, dann ist $X_1 + X_2 \sim \chi_{m_1+m_2}^2$.

Für $X \sim \chi_m^2$ gilt:

Momente

$$\mathbb{E}(X) = m, \quad \text{Var}(X) = 2m$$

Sonderfälle

$m = 1$: Ist $Z \sim N(0, 1)$ dann $X \sim Z^2$ (d.h. X und Z besitzen die gleiche Verteilung)

$m = 2$: $X \sim \text{Exp}\left(\frac{1}{2}\right)$

Quantile

Das γ -Quantil der χ_m^2 -Verteilung bezeichnen wir $\chi_{m,\gamma}^2$,

$$\mathbb{P}(X \leq \chi_{m,\gamma}^2) = \gamma, \quad \gamma \in (0, 1).$$

9.1.2 t -Verteilung

Definition 9.1.4 Eine Zufallsvariable mit der Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(x) := K'_m \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad K'_m := \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\Gamma\left(\frac{m}{2}\right)}, m \in \mathbb{N}$$

heißt **t -verteilt** (auch Student-verteilt) mit m Freiheitsgraden, oder kurz t_m verteilt.

Wichtig ist der folgende Zusammenhang mit normal- und χ^2 verteilten Zufallsvariablen:

Satz 9.1.5 Sind $Z \sim N(0, 1)$ und $Y \sim \chi_m^2$ unabhängig, dann gilt:

$$\frac{Z}{\sqrt{\frac{Y}{m}}} \sim t_n.$$

Für $X \sim t_m$ gilt:

Momente

$$\mathbb{E}(X) = 0 \text{ für } m \geq 2, \text{ Var}(X) = \frac{m}{m-2} \text{ für } m \geq 3$$

Sonderfälle

- $m = 1$: Cauchy-Verteilung, ohne Erwartungswert und Varianz
- $m \rightarrow \infty$: $f_m \rightarrow \varphi$ ($m \rightarrow \infty$) punktweise, wobei f_m bzw. φ die Dichten der t_m bzw. $N(0, 1)$ Verteilungen sind.

Quantile Das γ -Quantil der t_m -Verteilung bezeichnen wir $t_{m,\gamma}$,

$$\mathbb{P}(X \leq t_{m,\gamma}) = \gamma, \quad \gamma \in (0, 1).$$

Für $\gamma > \frac{1}{2}$ fällt $t_{m,\gamma}$ in m , $\lim_{m \rightarrow \infty} t_{m,\gamma} = u_\gamma$. Wegen der Symmetrie der Dichte gilt für jedes $m \in \mathbb{N}$

$$t_{m,\gamma} = -t_{m,1-\gamma} \text{ und } \mathbb{P}(|X| \leq t_{m,1-\frac{\alpha}{2}}) = 1 - \alpha.$$

9.1.3 F-Verteilung

Definition 9.1.6 Eine Zufallsvariable mit der Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(x) = \begin{cases} 0 & \text{falls } x \leq 0, \\ K''_{m,n} x^{\frac{m-2}{2}} (mx+n)^{-\frac{m+n}{2}} & \text{falls } x > 0, \end{cases}$$

wobei $m \in \mathbb{N}$ und K''_m die Konstante

$$K''_{m,n} := \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} m^{m/2} n^{n/2}$$

ist, heißt **F-verteilt** (auch Fisher-verteilt) mit m und n Freiheitsgraden, oder kurz $F_{m,n}$ verteilt.

Wichtig ist der folgende Zusammenhang mit χ^2 verteilten Zufallsvariablen:

Satz 9.1.7 Sind $X \sim \chi_m^2$, $Y \sim \chi_n^2$ unabhängig, dann gilt

$$\frac{\frac{1}{m}X}{\frac{1}{n}Y} \sim F_{m,n}.$$

Für $X \sim F_{m,n}$ gilt:

Momente

$$\begin{aligned} \mathbb{E}(X) &= \frac{n}{n-2} \text{ falls } n \geq 3, \\ \text{Var}(X) &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ falls } n \geq 5. \end{aligned}$$

Sonderfälle

$$m = 1: F_{1,n} \sim t_n^2$$

Quantile

Das γ -Quantil der $F_{m,n}$ -Verteilung bezeichnen wir $F_{m,n,\gamma}$,

$$\mathbb{P}(X \leq F_{m,n,\gamma}) = \gamma, \quad \gamma \in (0, 1).$$

Tabeliert sind die Quantile meistens für $\gamma > 0,9$. Die „unteren Quantile“ berechnen sich aus $F_{m,n,1-\gamma} = \frac{1}{F_{n,m,\gamma}}$.

$$m = 1 \quad F_{1,n,1-\alpha} = t_{n,1-\frac{\alpha}{2}}^2, \quad \lim_{n \rightarrow \infty} F_{1,n,1-\alpha} = u_{1-\alpha/2}^2$$

9.2 Konfidenzintervalle für Parameter der Normalverteilung

Wir nehmen hier an, dass gilt:

$$X_i \sim N(\mu, \sigma) \text{ für alle } i = 1, \dots, n.$$

Die interessierenden Schätzer für μ und σ^2 sind der Mittelwert und die empirische Varianz. Über deren Verteilung gibt der nächste Satz Auskunft.

Satz 9.2.1 (Student) Sind $X_1, \dots, X_n$ iid, $N(\mu, \sigma^2)$ verteilt, dann gilt für

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

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

$$(b) \quad (n-1) \frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$(c) \quad \bar{X}_n \text{ und } S_n^2 \text{ sind unabhängig.}$$

$$(d) \quad T_n = \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1}$$

Beweis [P], S. 21, Satz. □

9.2.1 Konfidenzintervall für μ mit bekannter Varianz

Aufgabenstellung Wir wissen, dass eine Zufallsvariable normalverteilt ist. Wir kennen die Varianz, nicht aber den Erwartungswert. Letzteren wollen wir anhand einer Stichprobe schätzen.

Als Schätzer für den Erwartungswert verwenden wir das Stichprobenmittel $\bar{X}_n$. Dessen Verteilung kennen wir aus Satz 9.2.1 (a). Somit ist die Standardisierung von $\bar{X}_n$

$$\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

standardnormalverteilt. Es gilt nun

$$\begin{aligned}
 \alpha &= \Phi(u_{1-\alpha/2}) - \Phi(-u_{1-\alpha/2}) \\
 &= \mathbb{P}\left(-u_{1-\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq u_{1-\alpha/2}\right) \\
 &= \mathbb{P}\left(\frac{-u_{1-\alpha/2}\sigma}{\sqrt{n}} - \bar{X}_n \leq -\mu \leq \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}} - \bar{X}_n\right) \\
 &= \mathbb{P}\left(\bar{X}_n - \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{u_{1-\alpha/2}\sigma}{\sqrt{n}}\right),
 \end{aligned}$$

d.h. unser Konfidenzintervall, das wir aus der Messreihe $x_1, \dots, x_n$ erhalten, ist

$$\left[\bar{x} - \frac{\sigma \cdot u_{1-\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{\sigma \cdot u_{1-\alpha/2}}{\sqrt{n}}\right]. \quad (9.1)$$

Die Bestimmung des Schätzintervalls bei bekannter Varianz verläuft also in vier Schritten:

- (a) Festlegung eines Konfidenzniveaus $1 - \alpha \in]0, 1[$.
- (b) Bestimmung des $(1-\alpha/2)$ Quantils der Standardnormalverteilung $u_{1-\alpha/2}$.
- (c) Berechnung des Stichprobenmittelwerts $\bar{x}$ und $\frac{\sigma \cdot u_{1-\alpha/2}}{\sqrt{n}}$.
- (d) Bestimmung des Konfidenzintervalls laut (9.1).

Bemerkung Das Intervall

$$\left[\bar{X} - \frac{\sigma \cdot u_{1-\alpha/2}}{\sqrt{n}}, \bar{X} + \frac{\sigma \cdot u_{1-\alpha/2}}{\sqrt{n}}\right]$$

enthält mit einer Wahrscheinlichkeit von $1 - \alpha$ den Erwartungswert. Man beachte, dass hier die Intervallgrenzen Zufallsvariablen sind (die Intervallgrenzen hängen von der Zufallsvariablen $\bar{X}_n$ ab). In (9.1) erhalten wir indes ein festes Intervall $[a, b]$ (der Mittelwert der Stichprobe wird eingesetzt) bei dem eine Wahrscheinlichkeitsaussage wie „das Intervall enthält mit einer Wahrscheinlichkeit von 95 % den gesuchten Erwartungswert“ keinen Sinn ergibt. Für ein festes Intervall $[a, b]$ gilt nämlich entweder $\mu \in [a, b]$ oder $\mu \notin [a, b]$. Man kann nur sagen, dass bei einem solchen Vorgehen in 95 % der Fälle der Erwartungswert in diesem Intervall liegt. In welchen Fällen das so ist, lässt sich aber nicht sagen!

Beispiel 9.2.2 Eine Maschine schneidet Stahlbleche automatisch auf die gewünschte Länge zu. Allerdings beträgt die Standardabweichung der Längen 2,2 cm. Es soll ein Schätzintervall für die erwartete Länge der Stahlbleche zu einem Konfidenzniveau von 95 % angegeben werden. Eine Stichprobe vom Umfang 30 ergab $\bar{x} = 80,50$ cm.

Lösung

- (a) Konfidenzniveau = 0,95, also $\alpha = 0,05$
- (b) Das 0,975-Quantil beträgt $u_{0,975} = 1,96$.
- (c) $\bar{x} = 80,50$ cm.
- (d) $\frac{2,2 \cdot 1,96}{\sqrt{30}} \approx 0,79$
- (e) Das Konfidenzintervall beträgt $[80,50-0,79; 80,50+0,79] = [79,71; 81,29]$.

□

Ergänzung: Quantile der Standardnormalverteilung

Wahrscheinlichkeit	Quantil
0,50 %	-2,58
1,00 %	-2,33
2,50 %	-1,96
5,00 %	-1,64
10,00 %	-1,28
90,00 %	1,28
95,00 %	1,64
97,50 %	1,96
99,00 %	2,33
99,50 %	2,58

Aufgabe 9.2.3 Bestimme für das vorhergehende Beispiel ein Schätzintervall zum Niveau 99 %.

Bestimmung des Stichprobenumfangs

Wir gehen der Frage nach, wie groß der Umfang n einer Stichprobe sein sollte damit das Intervall eine vorgegebene Länge L nicht überschreitet. Es wird eine möglichst kleine Länge L angestrebt.

Lemma 9.2.4 Wählt man n gemäß

$$n \geq \left(\frac{2\sigma u_{1-\alpha/2}}{L} \right)^2, \quad (9.2)$$

dann ist die Länge des Konfidenzintervall für den Erwartungswert von X kleiner als L .

Interpretation Je kleiner L gewählt wird, desto „genauer“ ist die Schätzung in dem Sinn, dass die Bandbreite in der der gesuchte Parameter liegt klein ist. Allerdings wächst der Stichprobenumfang n mit zunehmender Genauigkeit.

Begründung Das Intervall (9.1) hat die Länge

$$l = 2\sigma u_{1-\alpha/2} \frac{1}{\sqrt{n}}.$$

Löst man die Ungleichung

$$l = 2\sigma u_{1-\alpha/2} \frac{1}{\sqrt{n}} \leq L$$

nach n auf ergibt sich

$$\sqrt{n} \geq 2\sigma u_{1-\alpha/2} \frac{1}{L}.$$

Aufgabe 9.2.5 (Fortsetzung Beispiel 9.2.2) *Wieviele Beobachtungen wären bei den angegebenen Irrtumswahrscheinlichkeiten erforderlich, um ein Schätzintervall der angegebenen Länge zu erhalten?*

Irrtumswahrsch.	Intervalllänge	Anzahl Beobachtungen
5 %	1 cm	
5 %	0,5 cm	
1 %	1 cm	
1 %	0,5 cm	

9.2.2 Konfidenzintervall für μ mit unbekannter Varianz

Aufgabenstellung Wir wissen, dass eine Zufallsvariable normalverteilt ist. Wir kennen weder Varianz, noch Erwartungswert. Letzteren wollen wir anhand einer Stichprobe schätzen.

Wie im vorigen Abschnitt ist $\bar{X}_n$ der Schätzer für den Erwartungswert. Allerdings können wir $\bar{X}$ nicht mehr standardisieren, da die Varianz σ^2 unbekannt ist. Ein Ansatz ist es, anstelle der Standardabweichung σ den Schätzer

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

einzusetzen. Statt der Standardisierung von $\bar{X}_n$ erhalten wir

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

über dessen Verteilung der Satz von Student (Satz 9.2.1) Auskunft gibt:

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

Die zum vorigen Abschnitt analogen Überlegungen führen zum Konfidenzintervall für die Messreihe

$$\left[\bar{x} - \frac{s_n \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n}}, \bar{x} + \frac{s_n \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n}} \right]. \quad (9.3)$$

Die Bestimmung des Schätzintervalls für den Erwartungswert bei unbekannter Varianz erfolgt in vier Schritten:

- (a) Festlegung eines Konfidenzniveaus $1 - \alpha$.
- (b) Bestimmung des Quantils $1 - \frac{\alpha}{2}$ der t_{n-1} -Verteilung.
- (c) Berechnung des Stichprobenmittelwerts $\bar{x}$ und der empirischen Standardabweichung s_n .
- (d) Bestimmung des Konfidenzintervalls laut (9.3).

Bemerkung Das („zufällige“) Intervall

$$\left[\bar{X}_n - \frac{S_n \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{S_n \cdot t_{n-1, 1-\alpha/2}}{\sqrt{n}} \right]$$

umfasst mit einer Wahrscheinlichkeit von $1 - \alpha$ den gesuchten Erwartungswert. Das Intervall in (9.3) ist fest, die Aussage, dass es mit einer bestimmten Wahrscheinlichkeit den wahren Wert μ enthält ist nicht richtig. Siehe auch die Bemerkung auf S. 119.

Beispiel 9.2.6 *Wir setzen das vorherige Beispiel fort: Eine Maschine schneidet Stahlbleche automatisch auf die gewünschte Länge zu. Allerdings ist die Standardabweichung der Länge unbekannt. Es soll ein Schätzintervall für die erwartete Länge der Stahlbleche mit einem Konfidenzniveau von 95 % angegeben werden. Eine Stichprobe vom Umfang 30 ergab $\bar{x} = 80,50$ cm und $s = 2,30$ cm. Man bestimme ein Schätzintervall zum Konfidenzniveau 95 %.*

Lösung

- (a) Konfidenzniveau 0,95, $\alpha = 0,05$
- (b) Das 0,975-Quantil der t_{29} -Verteilung beträgt $t_{29,0,975} = 2,045$
- (c) $\bar{x} = 80,50\text{cm}$ und $s = 2,30\text{cm}$.
- (d) $\frac{2,3 \cdot 2,045}{30} \approx 0,85$
- (e) Das Konfidenzintervall beträgt $[80,50 - 0,85; 80,50 + 0,85] = [79,65; 81,35]$

Das Konfidenzintervall ist etwas größer als im Fall, dass die Varianz bekannt ist, es liegt ja auch weniger Information vor. $\square$

Aufgabe 9.2.7 Bestimmen Sie mit den Angaben des vorigen Beispiels ein Schätzintervall zum Niveau 99%.

9.2.3 Konfidenzintervall für σ^2

Aufgabenstellung Wir wissen, dass eine Zufallsvariable normalverteilt ist. Wir kennen weder Varianz, noch Erwartungswert. Die Varianz wollen wir anhand einer Stichprobe schätzen.

Idee Wir verwenden den erwartungstreuen Schätzer

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

für die Varianz. Laut Satz von Student (Satz 9.2.1, S. 118) gilt

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

Wir gehen nicht weiter auf die Herleitung der Vorgehensweise in diesem Fall ein, sie ist ähnlich zu der in den vorherigen Abschnitten vorgestellten. Die Bestimmung eines Schätzintervalls für die Varianz bei unbekanntem Erwartungswert μ :

- (a) Festlegung eines Konfidenzniveaus $1 - \alpha$
- (b) Bestimmung der Quantile $\chi_{n-1;\alpha/2}^2$ und $\chi_{n-1;1-\alpha/2}^2$.
- (c) Berechnung der Größe $(n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ aus der Stichprobe

(d) Berechnung der Werte $\frac{(n-1)s^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1;\alpha/2}^2}$.

(e) Gesuchtes Schätzintervall ist

$$\left[\frac{(n-1)s^2}{\chi_{n-1;1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1;\alpha/2}^2} \right] \quad (9.4)$$

Beispiel 9.2.8 *Vorheriges Beispiel: Eine Maschine schneidet Stahlbleche automatisch auf die gewünschte Länge zu. Allerdings sind die Standardabweichung σ und der Erwartungswert der Länge μ unbekannt. Es soll ein Schätzintervall für die Varianz der Länge der Stahlbleche mit einem Konfidenzniveau von 95 % angegeben werden. Eine Stichprobe vom Umfang 30 ergab $\bar{x} = 80,50\text{cm}$ und $s = 2,30\text{cm}$.*

Lösung

(a) Konfidenzniveau = 0,95,

(b) Das 0,025-Quantil der χ_{29}^2 -Verteilung beträgt 16,05 und das 0,975-Quantil 45,72

(c) $29 \cdot 2,3^2 = 153,41$

(d) Untere Grenze: $\frac{153}{45,72} \approx 3,36$

Obere Grenze: $\frac{153}{16,05} \approx 9,56$.

(e) Gesuchtes Schätzintervall für die Varianz ist $[3,36; 9,56]$

□

Aufgabe 9.2.9 (a) Man bestimme ein Schätzintervall im obigen Beispiel zum Niveau 99%.

(b) Angenommen die Ergebnisse wurden bei einer Stichprobe vom Umfang $n = 20$ erzielt. Wird das Schätzintervall zum Niveau von 99% länger als in (a) ausfallen? (ohne Rechnung, mit Begründung).

(c) Überprüfen Sie Ihre Vermutung durch nachrechnen.

9.3 Konfidenzintervalle für eine Wahrscheinlichkeit

Liegt keine Normalverteilung vor, kann bei großen Stichproben ein approximatives Konfidenzintervall angegeben werden.

Für die Wahrscheinlichkeit $p = \mathbb{P}(A)$ eines zufälligen Ereignisses A ist ein Konfidenzintervall zu konstruieren. Letztlich geht es wie in den vorangegangenen Abschnitten um die Konstruktion eines Konfidenzintervalls für einen Parameter der Wahrscheinlichkeitsverteilung der Zufallsvariable 1_A .

Die Wahrscheinlichkeitsfunktion von 1_A genügt einer $B(1, p)$ -Verteilung mit

$$\begin{aligned}\mathbb{P}(1_A = 1) &= \mathbb{P}(A) = p, & \mathbb{P}(1_A = 0) &= P(A^c) = 1 - p \\ \mathbb{E}(1_A) &= p, & \text{Var}(1_A) &= p(1 - p)\end{aligned}$$

Sei nun $x_1, \dots, x_n$ eine Realisierung von $X = (X_1, \dots, X_n)$ wobei die X_i iid sind, mit $X_i \sim B(1, p)$. Als Schätzer wird

$$\hat{p} := \frac{1}{n} \sum_{i=1}^n X_i$$

gewählt, mit

$$\mathbb{E}(\hat{p}) = p, \quad \text{Var}(\hat{p}) = \frac{p(1-p)}{n}. \quad (9.5)$$

$\hat{p}$ ist die relative Häufigkeit des Eintretens des Ereignisses A bei n unabhängigen Versuchen. Gemäß dem zentralen Grenzwertsatz ist $\hat{p}$ für hinreichend große n näherungsweise normalverteilt den Parametern aus (9.5). Somit gilt

$$\forall p \in (0, 1): \quad \mathbb{P} \left(-u_{1-\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\alpha/2} \right) \approx 1 - \alpha$$

Nach Umformungen erkennt man, dass die Ungleichung

$$-c \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq c$$

erfüllt ist, für $p \in I(X) = [U(X), O(X)]$ mit

$$\begin{aligned}U(X) &= \frac{n}{n+c^2} \left(\hat{p} + \frac{c^2}{n} - \frac{c}{\sqrt{n}} \sqrt{\hat{p}(1-\hat{p}) + \frac{c^2}{4n}} \right) \\ O(X) &= \frac{n}{n+c^2} \left(\hat{p} + \frac{c^2}{n} + \frac{c}{\sqrt{n}} \sqrt{\hat{p}(1-\hat{p}) + \frac{c^2}{4n}} \right).\end{aligned}$$

Daraus ergibt sich mit $c = u_{1-\alpha/2}$ das gesuchte Konfidenzintervall. Für große n erhält man die folgende Approximation:

$$\tilde{I}(X) = \left[\hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]. \quad (9.6)$$

Beispiel 9.3.1 In den Jahren 1950 bis 1970 wurden in der Schweiz 1.944.700 Kinder geboren, darunter 997.600 Knaben. Für die Wahrscheinlichkeit p , dass bei der Geburt eines Kindes in der Schweiz männliches Geschlecht festgestellt wird, bestimme man nach dem durch (9.6) gegebenen Konfidenzschätzverfahren ein konkretes Schätzintervall

(a) zum Konfidenzniveau 0,99,

(b) zum Konfidenzniveau 0,999.

Lösung (a) Wegen $u_{0.995} = 2.58$ erhalten wir mit $n = 1.944.700$ und $k = 997600$ für den unteren Endpunkt

$$\tilde{U} = \frac{1}{1.944.700} \left(997.600 - 2.58 \sqrt{\frac{997.600 \cdot 947.100}{1.944.700}} \right) = 0.5121$$

und für den oberen

$$\tilde{O} = \frac{1}{1.944.700} \left(997.600 + 2.58 \sqrt{\frac{997.600 \cdot 947.100}{1.944.700}} \right) = 0.5139$$

Wir behaupten daher, dass der wahre Parameter 0 im konkreten Schätzintervall $[0.5121, 0.5139]$ liegt. Wir wissen zwar nicht, ob diese Behauptung richtig ist, stellen sie aber doch mit einer gewissen Überzeugung auf, da wir aufgrund eines Verfahrens zu dieser Behauptung gelangt sind, das mit mindestens 99%-tiger Wahrscheinlichkeit richtige Behauptungen liefert.

(b) Wegen $u_{0.9995} = 3.29$ erhalten wir $\tilde{U} = 0.5118$ und $\tilde{O} = 0.5142$. Wir behaupten daher, dass der wahre Parameter 0 im konkreten Schätzintervall $[0.5118, 0.5142]$ liegt. Wir stellen diese Behauptung mit noch größerer Überzeugung auf als die Behauptung in (a), erkennen aber, dass diese Behauptung weniger präzise ist, da das angegebene konkrete Schätzintervall länger ist.

10 Hypothesentests

Ein **Test** ist ein Verfahren zur Überprüfung von Vermutungen über Verteilungen bzw. deren Parameter, die das Zustandekommen von Beobachtungsdaten beschreiben. Solche Vermutungen werden **Hypothesen** genannt. Bei diesen Tests werden Regeln zur Verwerfung von Hypothesen formuliert. Die Wahrscheinlichkeit einer fälschlichen Ablehnung soll eine vorgegebene Schranke $\alpha \in (0, 1)$ nicht überschreiten.

10.1 Konzepte

In diesem Abschnitt nehmen wir an, dass die Beobachtungswerte $x_1, \dots, x_n$ als Realisierungen von iid Zufallsvariablen $X_1, \dots, X_n$ angesehen werden können. Ihre Verteilungsfunktion F_ϑ hänge von $\vartheta \in \Theta \subset \mathbb{R}^d$ ab, und wir setzen wie bisher $X = (X_1, \dots, X_n)$. Liegt eine Vermutung über den Wert eines Parameters ϑ vor, formuliert man diese als **Nullhypothese**

$$H_0 : \vartheta \in \Theta_0, \emptyset \neq \Theta_0 \subset \Theta. \quad [\text{Nullhypothese}]$$

Wir verwenden folgende Begriffe:

Nullhypothese $H_0 : \vartheta \in \Theta_0, \emptyset \neq \Theta_0 \subset \Theta$

Alternative Hypothese $H_1 : \vartheta \in \Theta \setminus \Theta_0$.

Teststatistik : $T \circ X : \Omega \longrightarrow \mathbb{R}$ wobei $T : \mathbb{R}^n \longrightarrow \mathbb{R}$ messbar ist.

Signifikanzniveau : $\alpha \in (0, 1)$

Verwerfungsbereich : $B \subset \mathbb{R}$ so, dass gilt:

$$\forall \vartheta \in \Theta_0 : \mathbb{P}_\vartheta(T(X) \in B) \leq \alpha \quad (10.1)$$

B hängt also von Θ_0 und α ab.

Ein Signifikanztest verläuft in folgenden Schritten:

1. Formulierung einer Nullhypothese H_0 gemäß einer Vermutung (Behauptung) über den Wert von ϑ und einer Alternative H_1 .
2. Auswahl eines Signifikanzniveaus α ($\alpha = 0.10, \alpha = 0.05, \alpha = 0.01, \alpha = 0.001$ sind übliche Werte).
3. Auswahl einer Teststatistik T und eines Verwerfungsbereich B laut (10.1).

4. Mit einer Realisation (Stichprobe) $x_1, \dots, x_n$ Bestimmung von $T(x)$ (dies ist also eine Realisation der Zufallsvariablen $T \circ X$).
5. H_0 wird zugunsten von H_1 , verworfen, falls $T(x) \in B$, sonst nicht.

10.1.1 Fehlerwahrscheinlichkeiten

Als einen **Fehler 1. Art** bezeichnet man die fälschliche Verwerfung der Hypothese H_0 . Die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, also

$$\mathbb{P}_\vartheta(T(X) \in B), \quad \vartheta \in \Theta_0$$

ist wegen (10.1) höchstens α . Einen **Fehler 2. Art** begeht man, wenn man H_0 nicht verwirft, obwohl H_0 falsch ist (H_1 also richtig ist). Die Wahrscheinlichkeit für einen Fehler 2. Art bezeichnet man mit

$$\beta(\vartheta) := \mathbb{P}_\vartheta(T(X) \notin B).$$

In der Regel wird $\beta(\vartheta)$ umso kleiner, je weiter der „wahre“ Parameterwert ϑ von Θ_0 entfernt liegt.

10.1.2 Gütefunktion

Die **Gütefunktion** $G : \Theta \rightarrow \mathbb{R}$

$$G(\vartheta) := \mathbb{P}_\vartheta(T(X) \in B)$$

gibt für jeden Wert ϑ die Wahrscheinlichkeit an, dass H_0 abgelehnt wird. Laut (10.1) gilt

$$\begin{aligned} G(\vartheta) &\leq \alpha, & \text{für } \vartheta \in \Theta_0 \\ G(\vartheta) &= 1 - \beta(\vartheta), & \text{für } \vartheta \in \Theta_1. \end{aligned}$$

10.1.3 p -Werte

Angenommen $T(x) \notin B$, d.h. H_0 wird zum Niveau α nicht verworfen. Dann wird H_0 auch nicht verworfen zu einem Niveau $\tilde{\alpha} > \alpha$. Für ein $\tilde{\alpha} < \alpha$ kann jedoch beides eintreten, Verwerfung oder nicht. Der **p -Wert** ist das kleinste Niveau so, dass H_0 angenommen wird.

Der Verwerfungsbereich B hängt wie aus (10.1) ersichtlich von α ab, deshalb schreiben wir auch B_α . Nimmt man $B_{\alpha_1} \subset B_{\alpha_2}$ für $\alpha_1 < \alpha_2$ an (das ist bei den Tests, die später vorgestellt werden der Fall), dann erfüllt der p -Wert

$$p\text{-Wert} = \inf\{\tilde{\alpha} \in (0, 1) | T(x) \notin B_{\tilde{\alpha}}\} = \sup\{\tilde{\alpha} \in (0, 1) | T(x) \in B_{\tilde{\alpha}}\}.$$

Gibt man sich das Niveau α vor und gilt $\alpha < p$ -Wert, dann wird H_0 verworfen, denn für $\tilde{\alpha} \in (\alpha, p\text{-Wert})$ gilt $T(x) \in B_{\tilde{\alpha}}$.

10.1.4 Zusammenhang mit Konfidenzintervallen

Aus jedem Konfidenzintervall zum Niveau α , kann man einen Signifikanztest zum Niveau α gewinnen:

- Sei $I(x)$ ein Konfidenzintervall zum Niveau α .
- Verwerfe $H_0: \vartheta = \vartheta_0$ genau dann, wenn $\vartheta_0 \notin I(x)$.

Umgekehrt kann man aus einem Signifikanztest einen Konfidenzbereich konstruieren.

Definition 10.1.1 Ein **Konfidenzbereich** für $\vartheta \in \Theta$ zum Niveau $\alpha \in (0, 1)$ ist eine Abbildung

$$K : \mathbb{R}^n \longrightarrow \mathcal{P}(\Theta).$$

(d.h. zu jeder Stichprobe x wird eine Teilmenge $K(x) \subset \Theta$ konstruiert) so, dass

$$\begin{aligned} \forall \vartheta \in \Theta : \quad \{x \in \mathbb{R}^n | \vartheta \in K(x)\} &\in \mathcal{B}^n \\ \mathbb{P}_{\vartheta}(\vartheta \in K(X)) &\geq 1 - \alpha \end{aligned} \quad (10.2)$$

gilt. (Die erste Bedingung stellt sicher, dass (10.2) definiert ist).

Konfidenzintervalle sind also Konfidenzbereiche, bei denen $K(x)$ jeweils ein Intervall ist.

Ist $T(\vartheta_0, x)$ die Teststatistik zur Stichprobe x , B_{ϑ_0} der Verwerfungsbereich für $H_0: \vartheta = \vartheta_0$ zum Niveau α , dann ist

$$K(x) := \{\vartheta_0 \in \Theta | \vartheta = \vartheta_0 \text{ wird nicht abgelehnt}\}$$

ein Konfidenzbereich zum Niveau α .

Konkret würde man folgendermaßen vorgehen:

- Berechne $T(\vartheta_0, x)$ für alle $\vartheta_0 \in \Theta$
- Bestimme B_{ϑ_0} für alle ϑ_0
- Untersuche, ob $T(\vartheta_0, x) \in B_{\vartheta_0}$ (Ablehnung) oder $T(\vartheta_0, x) \notin B_{\vartheta_0}$ (Annahme).

10.2 Test bei Normalverteilungsannahme

Generell seien alle Zufallsvariablen in diesem Abschnitt normal verteilt.

Die nun folgenden Tests gibt es in für zwei Fälle:

Eine Stichprobe $x = (x_1, \dots, x_n)$ wird als Realisierungen von unabhängigen, identisch Zufallsvariablen $X_1, \dots, X_n$ angesehen. Ihre Verteilungsfunktion F_ϑ hänge von $\vartheta \in \Theta \subset \mathbb{R}^d$ ab, und wir setzen wie bisher $X = (X_1, \dots, X_n)$.

Zwei Stichproben Es liegen zwei Stichproben $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_m)$ vor, die Realisierungen von $X = (X_1, \dots, X_n)$ und $Y = (Y_1, \dots, Y_m)$ sind. Die Zufallsvariablen $X_1, \dots, X_n, Y_1, \dots, Y_m$ seien unabhängig, $X_i \sim F_{\vartheta_1}$, $i = 1, \dots, n$ und $Y_j \sim F_{\vartheta_2}$, $j = 1, \dots, m$. Insbesondere sind die X_i und Y_j also jeweils iid.

10.2.1 Gauß-Tests

Gauß Test gehen von normalverteilten Zufallsvariablen aus, bei **bekannter Varianz**. Die Hypothesen betreffen die Erwartungswerte.

(a) Eine Stichprobe, Zweiseitiger Test

$X_i \sim N(\mu, \sigma_0^2)$, die Standardabweichung $\sigma_0 > 0$ sei bekannt.

1. $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$
2. $T(x) := \frac{\bar{x} - \mu_0}{\sigma_0} \sqrt{n}, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$

Laut Satz 9.2.1 gilt, wenn H_0 zutrifft: $T(X) \sim N(0, 1)$

3. $B := (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty)$, d.h. H_0 wird genau dann abgelehnt, wenn $|T(x)| > u_{1-\alpha/2}$.

Unter H_0 gilt dann

$$\mathbb{P}_{\mu_0}(|T(X)| > u_{1-\alpha/2}) = \mathbb{P}_{\mu_0}(T(X) < -u_{1-\alpha/2}) + 1 - \mathbb{P}_{\mu_0}(T(X) < u_{1-\alpha/2}) = \alpha$$

also erfüllt B die gewünschte Ungleichung (10.1).

(b) Eine Stichprobe, Einseitiger Test Wie (a) mit folgenden Unterschieden:

1. $H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$
2. $T(x) := \frac{\bar{x} - \mu_0}{\sigma_0} \sqrt{n}, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$

$$3. B = (u_{1-\alpha}, \infty)$$

Wie in (a) gilt $T(X) \sim N\left(\frac{\mu - \mu_0}{\sigma_0} \sqrt{n}, 1\right)$ unter H_0 , wenn $X_i \sim N(\mu, \sigma_0^2)$. (10.1) folgt aus

$$\mathbb{P}_\mu(T(X) > u_{1-\alpha}) \leq \mathbb{P}_{\mu_0}(T(X) > u_{1-\alpha}) = \alpha.$$

(c) **Zwei Stichproben** $X_1, \dots, X_n \sim N(\mu_1, \sigma_1^2)$, $Y_1, \dots, Y_m \sim N(\mu_2, \sigma_2^2)$, σ_1^2, σ_2^2 bekannt.

$$1. H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$$

$$2. T(x) := \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sqrt{n}, \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i. \text{ Unter } H_0 \text{ gilt } T(X, Y) \sim N(0, 1)$$

$$3. B := (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty).$$

Beispiel 10.2.1 Eine Maschine schneidet Stahlbleche automatisch auf die gewünschte Länge zu. Die Standardabweichung der Längen beträgt 2,2 cm. Der Erwartungswert der Länge ist allerdings unbekannt. Eine Stichprobe ergab $\bar{x} = 80,50$ cm. Der Hersteller der Maschine behauptet, der Erwartungswert würde 80 cm betragen.

(a) Der Umfang der Stichprobe sei $n = 30$. Geben Sie H_0, H_1 an. Wird H_0 bei einem Niveau von 5% abgelehnt?

(b) Wird H_0 abgelehnt, wenn der Umfang 100 betragen hätte?

(c) Wird H_0 abgelehnt, wenn der Umfang 100 und das Niveau 1 % betragen hätte?

(d) Nun behauptet der Hersteller, dass der Erwartungswert höchstens 80 cm beträgt. Wie fällt der Test aus, bei einem Niveau von 5 % und $n = 30$?

Lösung (a) Hypothese $H_0 : \mathbb{E}(X) = 80$, Alternative Hypothese $H_1 : \mathbb{E}(X) \neq 80$. Das Niveau für einen Test der Hypothese H_0 wird auf 5 % festgelegt und damit beträgt die Irrtumswahrscheinlichkeit $\alpha = 5 \%$, d.h. die Wahrscheinlichkeit H_0 zu unrecht abzulehnen:

$$\frac{\bar{x} - \mu_0}{\sigma_0} \sqrt{n} = \frac{80,5 - 80}{2,2} \sqrt{30} \approx 1,24 \in [-1,96; 1,96]$$

d.h. H_0 wird nicht abgelehnt (muss aber nicht richtig sein).

(b)

$$\frac{\bar{x} - \mu_0}{\sigma_0} \sqrt{n} = \frac{80,5 - 80}{2,2} \sqrt{100} \approx 2,27 \notin [-1,96; 1,96]$$

d.h. H_0 wird abgelehnt (muss aber nicht falsch sein).

(c)

$$\frac{\bar{x} - \mu_0}{\sigma_0} \sqrt{n} = \frac{80,5 - 80}{2,2} \sqrt{100} \approx 2,27 \in [-2,58; 2,58]$$

d.h. H_0 wird nicht abgelehnt (muss aber nicht richtig sein).(d) $T(x) = 1,24 < 1,64$, H_0 wird nicht abgelehnt. □

10.2.2 t-Tests

t -Tests sind Tests für Hypothesen über den Erwartungswert, für den Fall dass die Standardabweichung unbekannt ist (im vorigen Abschnitt hatten wir sie als bekannt vorausgesetzt). Die Teststatistiken sind ähnlich jenen des Gauß-Tests, die dort bekannte Varianz wird hier durch einen Schätzer dafür ersetzt.

(a) Eine Stichprobe, Zweiseitiger Test

$X_i \sim N(\mu, \sigma^2)$, μ, σ seien unbekannt.

$$1. H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$$

$$2. T(x) := \frac{\bar{x} - \mu_0}{s} \sqrt{n}, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Laut Satz 9.2.1 gilt, wenn H_0 zutrifft: $T(X) \sim t_{n-1}$

$$3. B := (-\infty, -t_{n-1, 1-\alpha/2}) \cup (t_{n-1, 1-\alpha/2}, \infty), \text{ d.h. } H_0 \text{ wird genau dann abgelehnt, wenn } |T(x)| > t_{n-1, 1-\alpha/2}.$$

Wie beim Gauß-Test folgt unter H_0 gilt dann

$$\mathbb{P}_{\mu_0}(|T(X)| > t_{n-1, 1-\alpha/2}) = \alpha.$$

(b) Eine Stichprobe, Einseitiger Test Wie (a) mit folgenden Unterschieden:

$$1. H_0 : \mu \leq \mu_0, H_1 : \mu > \mu_0$$

$$2. T(x) := \frac{\bar{x} - \mu_0}{s} \sqrt{n}, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

$$3. B = (t_{n-1, 1-\alpha}, \infty)$$

(c) **Zwei Stichproben** $X_1, \dots, X_n \sim N(\mu_1, \sigma_1^2)$, $Y_1, \dots, Y_m \sim N(\mu_2, \sigma_2^2)$, σ_1^2, σ_2^2 bekannt.

$$1. H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$$

$$2. T(x) := \frac{\bar{x} - \bar{y}}{s} \sqrt{\frac{mn}{m+n}}, \text{ mit}$$

$$s^2 := \frac{(n-1)s_x^2 + (m-1)s_y^2}{m+n-2}$$

wobei s_x^2 und s_y^2 die empirischen Varianzen der Stichproben x und y sind. Falls H_0 zutrifft, gilt $T(X, Y) \sim t_{n+m-2}$.

$$3. B := (-\infty, -t_{m+n-2, 1-\alpha/2}) \cup (t_{m+n-2, 1-\alpha/2}, \infty).$$

Beispiel 10.2.2 Eine Maschine schneidet Stahlbleche automatisch auf die gewünschte Länge zu. Der Erwartungswert und die Varianz der Länge sind unbekannt. Der Hersteller der Maschine behauptet, der Erwartungswert würde 80 cm betragen. Eine Stichprobe vom Umfang $n = 30$ ergab $s = 2,3$ cm und $\bar{x} = 80,5$ cm. Nehmen Sie die Hypothese bei einem Niveau von 5 % an?

Lösung

$$\frac{80,5 - 80}{2,3} \sqrt{30} = 1,19, \quad t_{29;0,975} = 2,045 \implies H_0 \text{ wird nicht abgelehnt}$$

□

Beispiel 10.2.3 In einer Versuchsanstalt erhielten 9 von 22 Masttieren (Gruppe I) Grünfutterzumischung, während die übrigen 13 Tiere (Gruppe II) ausschließlich mit proteinhaltigen Mastfutter gefüttert wurden. Nach einer gewissen Zeit wurde die Gewichtszunahme in kg bei den Tieren festgestellt:

<i>I</i>	7.0	11.8	10.1	8.5	10.7	13.2	9.4	7.9	11.1				
<i>II</i>	13.4	14.6	10.4	11.9	12.7	16.1	10.7	8.3	13.2	10.3	11.3	12.9	9.7

Führen beide Fütterungen zu gleichen Mastergebnissen?

Lösung Mittelwerte: $\bar{x} = 9.97$, $\bar{y} = 11.96$.

Empirische Varianzen: $s_x^2 = 3.90$ und $s_y^2 = 4.57$ sowie $T(x, y) = 2.21$

Wird die Nullhypothese gleicher Erwartungswerte auf dem Niveau $\alpha = 0.05$ geprüft, so führt dieser Wert zu einer Ablehnung, da er das Quantil $t_{20,0.975} = 2.09$ übersteigt. □

10.2.3 χ^2 -Streuungstest

Dies ist ein Test für die Varianz bei einer Stichprobe, wobei Erwartungswert und Varianz unbekannt sind.

$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ iid, μ und σ unbekannt.

1. $H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 \neq \sigma_0^2$
2. $T(x) = \frac{(n-1)s^2}{\sigma_0^2}, s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$

Nach dem Satz von Student gilt $T(X) \sim \chi_{n-1}^2$

3. $B = [0, \chi_{n-1, \alpha/2}^2) \cup (\chi_{n-1, 1-\alpha/2}^2, \infty)$ ist, d.h.

$$H_0 \text{ wird abgelehnt} \iff \frac{(n-1)s^2}{\sigma_0^2} \notin [\chi_{n-1, \alpha/2}^2; \chi_{n-1, 1-\alpha/2}^2] \quad (10.3)$$

Beispiel 10.2.4 Wiegt man einen Probekörper mit einer Feinwaage mehrere Male, so stellen sich unterschiedliche Ergebnisse ein. Wir nehmen an, dass die Wägungen durch unabhängige $N(\mu, \sigma^2)$ -verteilte Zufallsvariablen beschrieben werden können, wobei μ dem unbekannten Gewicht des Probekörpers entspricht und $\sigma > 0$ eine vom Hersteller der Waage angegebene positive Zahl ist.

- (a) Für μ soll mit dem arithmetischen Mittel der Messergebnisse ein Schätzwert angegeben werden. Der Hersteller gibt $\sigma^2 = 5 \cdot 10^{-5}$ als Wert an. Der Schätzwert für μ soll mit einer Wahrscheinlichkeit > 0.95 vom wahren Wert um nicht mehr als $5 \cdot 10^{-3}$ abweichen. Wie viele Messergebnisse werden benötigt?
- (b) Die Qualitätsangabe des Herstellers soll auf dem Niveau $\alpha = 0.05$ überprüft werden. Bei 16 Wägungen ergibt sich für die Teststatistik der Wert 18.3. Wie beurteilen Sie die Aussage des Herstellers?

Lösung

(a) Mindestens 8 Messergebnisse.

(b) Das ist ein Wert, der zwischen $\chi_{15, 0.025}^2 = 6.26$ und $\chi_{15, 0.975}^2 = 27.49$. Auf dem 5%-Niveau ergibt sich also kein Widerspruch zur Qualitätsangabe des Herstellers. $\square$

10.2.4 F-Test

Dies ist ein Test für die Varianz bei zwei Stichproben.

$X_1, \dots, X_n \sim N(\mu_1, \sigma_1^2)$ iid, $Y_1, \dots, Y_m \sim N(\mu_2, \sigma_2^2)$ iid $\mu_1, \mu_2, \sigma_1, \sigma_2$ unbekannt.

$$1. H_0 : \sigma_1^2 = \sigma_2^2, H_1 : \sigma_1^2 \neq \sigma_2^2$$

$$2. T(x, y) = \frac{s_x^2}{s_y^2}, s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, s_y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$$

Nach dem Satz von Student gilt $T(X, Y) \sim F_{n-1, m-1}$

$$3. B = [0, F_{n-1, m-1, \alpha/2}) \cup (F_{n-1, m-1, 1-\alpha/2}, \infty).$$

Beispiel 10.2.5 Wir greifen Beispiel 10.2.3 auf. Bei der Überprüfung der dort betrachteten Nullhypothese $H_0 : \mu_1 = \mu_2$ waren wir von der Annahme gleicher Varianzen der zugrunde liegenden Zufallsvariablen ausgegangen, um die Anwendung des Zweistichproben t -Tests rechtfertigen zu können. Wir betrachten also jetzt (mit den Bezeichnungen des F -Tests) die Nullhypothese $H_0 : \sigma_1^2 = \sigma_2^2$. Was ergibt der Signifikanztest bei einem Niveau von 0.1?

Lösung $s_x^2 = 3.90$ und $s_y^2 = 4.57$. Daraus ergibt sich als Wert der Teststatistik des F -Tests $s_x^2/s_y^2 = 0.85 \in (F_{8,12;0.05}, F_{8,12;0.95}) = (0.3, 2.85)$. Der F Test führt auf dem Niveau 0.10 nicht zu einer Ablehnung der Nullhypothese.

Hätte der F -Test zu einer Ablehnung der Nullhypothese geführt, wäre die Prüfung auf Gleichheit der Erwartungswerte mit dem t -Test nicht sinnvoll gewesen. $\square$

10.3 Asymptotischer Binomialtest

Wir gehen von Zufallsvariablen $X_1, \dots, X_n \sim B(1, p)$ iid. aus und einer Realisierung $x = (x_1, \dots, x_n)$.

$$1. H_0: p = p_0, H_1: p \neq p_0$$

$$2. T(x) := \sum_{i=1}^n x_i$$

$$3. T(X) \sim B(n, p) \text{ wegen Satz 5.5.16 (c). Ist } n \text{ groß, dann gilt } B(n, p) \approx N(np, np(1-p))$$

$$4. B = (-\infty, np_0 - u_{1-\alpha/2} \sqrt{np_0(1-p_0)}) \cup (np_0 + u_{1-\alpha/2} \sqrt{np_0(1-p_0)}, \infty).$$

Mit der Normalverteilungsapproximation sieht man wie in den vorherigen Fällen die Gültigkeit von (10.1).

10.4 Anpassungstests

In den vorhergehenden Tests wurde der Verteilungstyp als bekannt vorausgesetzt. Analytische Methoden zur Nachprüfung von Hypothesen über einen Verteilungstyp werden Anpassungstest genannt. Es wird geprüft, wie gut eine theoretische Verteilung zu der Häufigkeitsverteilung einer gegebenen Stichprobe passt.

10.4.1 Kolmogorov-Smirnov

Dieser Test beruht auf dem Satz von Kolmogorov (Satz 7.3.3, S. 100) und kann für stetige Verteilungsfunktionen angewendet werden.

Sei $x_1, \dots, x_n$ eine Realisierung von Zufallsvariablen $X_1, \dots, X_n$ mit $X_i \sim F$ iid, wobei F eine stetige Verteilungsfunktion ist.

1. $H_0 : F = F_0, H_1 : F \neq F_0$
2. $T(x) = \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$, wobei F_n die empirische Verteilungsfunktion von $x_1, \dots, x_n$ ist. Wegen Satz 7.3.3 ist für große n

$$\mathbb{P}(\sqrt{n}T(X) \leq t) \approx K(t)$$

unter der Nullhypothese. K ist hierbei die Kolmogorovsche Verteilungsfunktion.

3. H_0 wird zum Niveau α verworfen, wenn

$$T(x) \geq \frac{k_{1-\alpha}}{\sqrt{n}}, \quad k_{1-\alpha} := K^{-1}(1 - \alpha).$$

Unter H_0 gilt dann:

$$\mathbb{P}(\sqrt{n}T(X) > k_{1-\alpha}) = 1 - \mathbb{P}(\sqrt{n}T(X) \leq k_{1-\alpha}) \approx 1 - K(k_{1-\alpha}) = \alpha.$$

Die Bestimmung von $T(x)$ ist durchführbar. Da F stetig ist und F_n eine Treppenfunktion ist, wird $\sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$ an einer der Sprungstellen von F_n angenommen, also an einer der Stellen $x_i, i = 1, \dots, n$:

$$T(x) = \max_i \{|F(x_i) - F_n(x_i)|, |F(x_i) - F_n(x_i - 0)| \mid i = 1, \dots, n\}.$$

Beispiel 10.4.1 (Test auf Gleichverteilung) Gegeben seien folgende Zufallszahlen (die Excel liefert):

0,953 0,055 0,534 0,540 0,916 0,100 0,447 0,158 0,176 0,716

Was ergibt der Test auf $U(0, 1)$ -Verteilung zum Niveau 0,1?

Lösung Die Nullhypothese lautet

$H_0 : F = F_0$ mit $F(x) = x$ für $x \in [0, 1]$. Mit $d_+ := |F(x_i) - F_{10}(x_i)|$, $d_- := |F(x_i) - F_{10}(x_i - 0)|$ erhalten wir

x_i	0,055	0,100	0,158	0,176	0,447	0,534	0,540	0,716	0,916	0,953
$F_{10}(x_i)$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$F_{10}(x_i -)$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$F(x_i)$	0,055	0,100	0,158	0,176	0,447	0,534	0,540	0,716	0,916	0,953
d_+	0,045	0,100	0,142	0,224	0,053	0,066	0,160	0,084	0,016	0,047
d_-	0,055	0,000	0,042	0,124	0,047	0,034	0,060	0,016	0,116	0,053

und somit $d_{10} = 0,224$. Wegen $0,224 \leq \frac{1,22}{\sqrt{10}} \approx 0,38$ wird die Nullhypothese nicht abgelehnt. $\square$

10.4.2 χ^2 -Anpassungstest

Multinomialverteilung

Sei $n \in \mathbb{N}$, $p = (p_1, \dots, p_r) \in (0, \infty)^r$ und $\sum_{i=1}^r p_i = 1$. Dann heißt die Zufallsvariable $Y = (Y_1, \dots, Y_r) \in \mathbb{N}_0^r$ **multinomial verteilt** $M(n, p)$ falls gilt

$$\mathbb{P}(Y_1 = y_1, \dots, Y_r = y_r) = \begin{cases} \frac{n!}{y_1! \dots y_r!} \prod_{i=1}^r p_i^{y_i} & \sum_{i=1}^r y_i = n \\ 0 & \text{sonst.} \end{cases}$$

Die Multinomial Verteilung modelliert ein Urnenmodell:

- Urne enthält Kugeln mit r verschiedenen Farben.
- p_i ist der Anteil der Kugeln der i -ten Farbe, $i = 1, \dots, r$.
- Es wird n -mal unabhängig, mit Zurücklegen gezogen.
- Y_i ist die Anzahl der gezogenen Kugeln der Farbe i , $i = 1, \dots, r$.

Falls $r = 2$ liegt die Binomialverteilung mit Parametern n, p_1 vor. Ferner gilt $Y_i \sim B(n, p_i)$ und somit

$$\mathbb{E}(Y_i) = np_i, \quad \text{Var}(Y_i) = np_i(1 - p_i).$$

Anpassungstest für $M(n, p)$

Wir betrachten ein Zufallsexperiment, das eine von r Alternativen $A_1, \dots, A_r$ auswählt, und zwar mit den Wahrscheinlichkeiten

$$p_1 = \mathbb{P}(A_1), \dots, p_r = \mathbb{P}(A_r), \quad \sum_{i=1}^r p_i = 1.$$

und das n mal unabhängig wiederholt wird. Es ergeben sich $y_1, \dots, y_r$ Häufigkeiten, mit denen die Alternativen $A_1, \dots, A_r$ vorkommen, $\sum_{i=1}^r y_i = n$.

1. $H_0 : p_1 = p_1^{(0)}, \dots, p_r = p_r^{(0)}, H_1 : p_i \neq p_i^{(0)}$ für ein $i = 1, \dots, r$.
2. $T(y) = \sum_{i=1}^r \frac{(y_i - np_i^{(0)})^2}{np_i^{(0)}}$. Man kann zeigen ([WN], S. 88), dass für alle $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} \mathbb{P}_{p^{(0)}}(T(Y) \leq x) = \chi_{r-1}^2(x)$$

gilt, also für große n $T(Y) \sim \chi_{r-1}^2$. Als **Faustregel** soll $np_i^{(0)} \geq 5$ für alle i gelten.

3. H_0 wird zum Niveau α verworfen, wenn gilt:

$$T(y_1, \dots, y_r) > \chi_{r-1, 1-\alpha}^2$$

Bemerkung 10.4.2 Bei einer Summe von r Quadraten, würde man für $T(Y)$ eine χ_r^2 -Verteilung erwarten. Wegen der Nebenbedingung $\sum_{i=1}^r p_i = 1$ erhält man aber einen Freiheitsgrad weniger.

Beispiel 10.4.3 (Idealer oder gezinkter Würfel) Das einmalige Werfen eines vorgelegten Würfels wird durch eine $M(1, p_1, \dots, p_6)$ Verteilung modelliert, wobei $p_i > 0$ die Eintrittswahrscheinlichkeit für die Augenzahl i darstellt, $i = 1, \dots, 6$. Bei n -facher Wiederholung besitzt dann das Tupel $Y = (Y_1, \dots, Y_6)$ der Eintrittshäufigkeiten Y_i für die Augenzahlen i $M(n, p_1, \dots, p_6)$ -Verteilung.

Bei $n = 30$ ergab sich das Ergebnis $y = (5, 7, 2, 5, 8, 3)$. Wird die Hypothese, dass es sich um einen fairen Würfel bei einem Niveau von $\alpha = 0.05$ bzw. $\alpha = 0.1$ verworfen?

Lösung $H_0 : (p_1, \dots, p_6) = (1/6, \dots, 1/6), H_1 : (p_1, \dots, p_6) \neq (1/6, \dots, 1/6)$. Die Faustregel $np_j^{(0)} = 30/6 \geq 5$ ist erfüllt.

$$T(y) = \sum_{i=1}^6 \frac{(y_i - 5)^2}{5} = \frac{1}{5} (0 + 4 + 9 + 0 + 9 + 4) = 5,2.$$

Die Nullhypothese wird also wegen $\chi_{5,0,95}^2 = 11,07 > 5,2 = T(y)$ nicht verworfen. (Hieraus können wir jedoch nicht schließen, dass der benutzte Würfel ein fair ist.) Auch zum Niveau 0,1 wird H_0 nicht verworfen, denn $\chi_{5,0,9}^2 = 9,24$. $\square$

χ^2 -Anpassungstest für eine Verteilungsfunktion F_0

Gegeben sei $x = (x_1, \dots, x_n)$ eine Realisierung von $X = (X_1, \dots, X_n)$, X_i iid mit (unbekannter) Verteilungsfunktion F . Der Wertebereich von X_1 wird zerlegt in r disjunkte Intervalle $I_1, \dots, I_r$ und

$$\mathbb{P}_F(X_1 \in I_i) = p_i, \quad i = 1, \dots, r. \quad (10.4)$$

Das Problem der Überprüfung ob $F = F_0$ für eine bestimmte Verteilungsfunktion gilt, wird auf die Fragestellung

$$(p_1, \dots, p_r) = (p_1^{(0)}, \dots, p_r^{(0)}), \text{ mit } p_i^{(0)} := \mathbb{P}_{F_0}(X_i \in I_i)$$

reduziert.

Aus $x_1, \dots, x_n$ werden die Häufigkeiten

$$y_i = |\{x_k | x_k \in I_i\}|, \quad i = 1, \dots, r$$

bestimmt. Nun wird der Test des vorherigen Abschnitts auf $y_1, \dots, y_r$ angewendet, wobei $(y_1, \dots, y_r)$ eine Realisierung einer $M(n, p_1, \dots, p_r)$ Zufallsvariable ist.

1. Hypothesen:

$$H_0 : (p_1, \dots, p_r) = (p_1^{(0)}, \dots, p_r^{(0)}) \quad (10.5)$$

$$H_1 : (p_1, \dots, p_r) \neq (p_1^{(0)}, \dots, p_r^{(0)}) \quad (10.6)$$

$$2. \quad T(y_1, \dots, y_r) = \sum_{i=1}^r \frac{(y_i - np_i^{(0)})^2}{np_i^{(0)}}.$$

3. $B = (\chi_{r-1, 1-\alpha}^2, \infty)$, also wird H_0 abgelehnt, wenn $T(y_1, \dots, y_r) > \chi_{r-1, 1-\alpha}^2$ gilt.

Man beachte, dass bei Anwendung des χ^2 -Anpassungstests als Test für das Vorliegen einer Verteilungsfunktion F_0 nicht die Nullhypothese $F = F_0$ überprüft wird, sondern nur, ob die Verteilung $\mathbb{P}_F$ mit der vermuteten Verteilung $\mathbb{P}_{F_0}$ auf $I_1, \dots, I_r$ übereinstimmt.

Anders als der Kolmogorov-Smirnov-Test, kommt dieser Test ohne die Voraussetzung der Stetigkeit aus.

Anpassungstest für einen Verteilungstyp

Der χ^2 -Anpassungstest lässt sich auch zum Prüfen eines bestimmten Verteilungstyps wie z.B. $N(\mu, \sigma^2)$, $\text{Poi}(\lambda)$ etc. verwenden. Angenommen der Verteilungstyp wird mittels Parametern $\vartheta \in \Theta \subset \mathbb{R}^d$ und Verteilungsfunktionen F_ϑ beschrieben. Die Nullhypothese wird sich dann nicht auf eine spezielle Auswahl von F_{ϑ_0} beziehen, sondern eher $F = F_\vartheta$ für geeignetes ϑ lauten. In (10.5) und (10.6) stehen dann anstelle der Zahlenwerte $p^{(0)}$ vom Parameter ϑ abhängende Größen $p_i(\vartheta)$. Um dennoch mit Zahlen arbeiten zu können setzt man dort $p_i(\hat{\vartheta})$ ein, wobei $\hat{\vartheta}$ aus den Daten $y_1, \dots, y_r$ mittels Maximum Likelihood geschätzt wird, also durch maximieren der Likelihood

$$L(\vartheta) = \prod_{i=1}^r p_i(\vartheta)^{y_i}. \quad (10.7)$$

Dann wird der χ^2 Anpassungstest durchgeführt.

Dieser „vorgeschaltete“ Schätzvorgang bleibt nicht ohne Einfluss auf die Verteilung der Testgröße. Man kann unter gewissen, sehr allgemeinen Voraussetzungen jedoch zeigen, dass sich wiederum näherungsweise eine χ^2_{r-1-d} -Verteilung ergibt. Für die Entscheidungsregel wird dann das Quantil $\chi^2_{r-d-1, 1-\alpha}$ herangezogen (siehe [WN], S. 89).

1. Hypothesen:

$$H_0 : (p_1, \dots, p_r) = (p_1(\hat{\vartheta}), \dots, p_r(\hat{\vartheta}))$$

$$H_1 : (p_1, \dots, p_r) \neq (p_1(\hat{\vartheta}), \dots, p_r(\hat{\vartheta}))$$

wobei $\hat{\vartheta}$ der ML-Schätzer gemäß (10.7) ist.

$$2. T(y_1, \dots, y_r) = \sum_{i=1}^r \frac{(y_i - np_i(\hat{\vartheta}))^2}{np_i(\hat{\vartheta})}.$$

3. $B = (\chi^2_{r-d-1, 1-\alpha}, \infty)$, also wird H_0 abgelehnt, wenn $T(y_1, \dots, y_r) > \chi^2_{r-d-1, 1-\alpha}$ gilt.

Bemerkung 10.4.4 (Forts. Bemerkung 10.4.2) *Dadurch, dass man d Parameter aus den Daten schätzt, reduziert sich die Anzahl der Freiheitsgrade von $r - 1$ um d , es verbleiben also $r - d - 1$ Freiheitsgrade.*

Beispiel 10.4.5 *Die Anzahl der durch Hufschlag verursachten Todesfälle pro Jahr in preußischen Kavallerieregimentern ist in der folgenden Tabelle gegeben.:*

Anzahl der Todesfälle	0	1	2	3	4
beobachtete Häufigkeit	109	65	22	3	1

Man untersuche, ob für dieses Phänomen eine der Poisson-Verteilungen mit Parameter $\lambda > 0$ angemessen ist, mit Hilfe des χ^2 -Anpassungstests.

Lösung Sei X die Anzahl der toten pro Regiment. Wählt man als Unterteilung des Wertebereichs $\mathbb{N}_0$ der Poisson-verteilten Zufallsvariablen X die Menge

$$I_1 = \{0\}, \quad I_2 = \{1\}, \quad I_3 = \{3, 4, \dots\}$$

so ist $r = 4$,

$$\begin{aligned} p_1(\lambda) &= e^{-\lambda} \\ p_2(\lambda) &= e^{-\lambda} \lambda \\ p_3(\lambda) &= e^{-\lambda} \frac{\lambda^2}{2} \\ p_4(\lambda) &= 1 - e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2} \right) \end{aligned}$$

Die zu den Anzahlen $y_1 = 109$, $y_2 = 65$, $y_3 = 22$ und $y_4 = 4$ gehörende Likelihood (siehe (10.7)) ist

$$\begin{aligned} L(\lambda) &= e^{-109\lambda} e^{-65\lambda} \lambda^{65} e^{-22\lambda} \frac{\lambda^{44}}{2^{22}} \left(1 - e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2} \right) \right)^4 \\ &= e^{-196\lambda} \frac{\lambda^{109}}{2^{22}} \left(1 - e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2} \right) \right)^4. \end{aligned}$$

Eine numerische Rechnung ergibt, dass das Maximum dieser Funktion bei $\hat{\lambda} = 0.61$ liegt. Daraus ergeben sich die geschätzten Wahrscheinlichkeiten

$$p_1(\hat{\lambda}) = 0,543, \quad p_2(\hat{\lambda}) = 0,331, \quad p_3(\hat{\lambda}) = 0,101 \quad p_4(\hat{\lambda}) = 0,024.$$

Da $d = 1$ ist, ist die χ^2_{4-1-1} -Verteilung zugrunde zu legen. Bei $\alpha = 0,1$ ergibt sich $\chi^2_{2,0.9} = 4,60$. Der Wert der Testgröße ist

$$\begin{aligned} T(y_1, \dots, y_4) &= \frac{(109 - 108,6)^2}{108,6} + \frac{(65 - 66,2)^2}{66,2} + \frac{(22 - 20,2)^2}{20,2} + \frac{(4 - 4,8)^2}{4,8} \\ &= 0,0015 + 0,0218 + 0,1604 + 0,1333 = 0,3170. \end{aligned}$$

Wir erhalten einen Wert, der unter dem Quantil $\chi^2_{2,0.9} = 4,60$ liegt, H_0 wird nicht verworfen.

10.5 χ^2 -Unabhängigkeitstest

Für eine zweidimensionale Zufallsvariable (X, Y) soll überprüft werden, ob ihre Komponenten X und Y unabhängig sind. Entsprechend dem Vorgehen beim χ^2 -Anpassungstest zerlegen wir den Wertebereich von X bzw. Y in disjunkte Mengen $I_1, \dots, I_k$ bzw. $J_1, \dots, J_l$. Es seien für $i \in \{1, \dots, k\}$ und $j \in \{1, \dots, l\}$

$$\begin{aligned} p_{ij} &:= \mathbb{P}(X \in I_i, Y \in J_j) \\ p_{i\bullet} &:= \mathbb{P}(X \in I_i) = \sum_{j=1}^l p_{ij} \\ p_{\bullet j} &:= \mathbb{P}(Y \in J_j) = \sum_{i=1}^k p_{ij} \end{aligned}$$

Sind X und Y unabhängig, dann gilt für alle $i \in \{1, \dots, k\}$ und $j \in \{1, \dots, l\}$

$$\mathbb{P}(X \in I_i, Y \in J_j) = \mathbb{P}(X \in I_i) \cdot \mathbb{P}(Y \in J_j)$$

also

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j}.$$

Die Eigenschaft der Unabhängigkeit von X und Y wird auf die Gültigkeit dieser Gleichungen reduziert.

Seien $(x_1, y_1), \dots, (x_n, y_n)$ unabhängige Realisierungen von (X, Y) und

$$n_{ij} := |\{(x_p, y_p) \in I_i \times J_j\}|, \quad \text{für } i \in \{1, \dots, k\}, j \in \{1, \dots, l\}$$

die Anzahl der Realisierungen in $I_i \times J_j$. Diese Werte n_{ij} werden in eine **Kontingenztafel** eingetragen und die Spalten- und Zeilensummen

$$n_{\bullet j} = \sum_{i=1}^k n_{ij} \quad \text{bzw.} \quad n_{i\bullet} = \sum_{j=1}^l n_{ij}$$

berechnet.

	J_1	J_2	$\dots$	J_l	
I_1	n_{11}	n_{12}	$\dots$	n_{1l}	$n_{1\bullet}$
I_2	n_{21}	n_{22}	$\dots$	n_{2l}	$n_{2\bullet}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_k	n_{k1}	n_{k2}	$\dots$	n_{kl}	$n_{k\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$\dots$	$n_{\bullet l}$	n

Entsprechend definiert man von $(X_1, Y_1), \dots, (X_n, Y_n)$ iid wie (X, Y) verteilt ausgehend, die Zufallsvariablen N_{ij} , $N_{i\bullet}$ und $N_{\bullet j}$.

1. $H_0 : p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$ für alle Paare (i, j) bei der Gegenhypothese
 $H_1 : p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j}$ für mindestens ein Paar (i, j) .
2. Die Testgröße ist

$$T((x_1, y_1), \dots, (x_n, y_n)) := \sum_{i=1}^k \sum_{j=1}^l \frac{(n \cdot n_{ij} - n_{i\bullet} \cdot n_{\bullet j})^2}{n \cdot n_{i\bullet} \cdot n_{\bullet j}}.$$

Man erhält sie aus der Testgröße für den χ^2 -Anpassungstest (siehe S. 139, 2.): Schätzer für $p_{i\bullet}$, $p_{\bullet j}$ sind $\hat{p}_{i\bullet} = \frac{n_{i\bullet}}{n}$, $\hat{p}_{\bullet j} = \frac{n_{\bullet j}}{n}$

$$\begin{aligned} \sum_{i,j} \frac{\left(n_{ij} - n \frac{n_{i\bullet}}{n} \frac{n_{\bullet j}}{n}\right)^2}{n \frac{n_{i\bullet}}{n} \frac{n_{\bullet j}}{n}} &= \\ &= \sum_{i,j} \frac{\left(n_{ij} - n \frac{n_{i\bullet}}{n} \frac{n_{\bullet j}}{n}\right)^2}{n \frac{n_{i\bullet}}{n} \frac{n_{\bullet j}}{n}} = \sum_{i,j} \frac{\left(\frac{1}{n} (n \cdot n_{ij} - n_{i\bullet} \cdot n_{\bullet j})\right)^2}{\frac{1}{n} n_{i\bullet} n_{\bullet j}} \\ &= \sum_{i,j} \frac{(n \cdot n_{ij} - n_{i\bullet} \cdot n_{\bullet j})^2}{n \cdot n_{i\bullet} n_{\bullet j}} \end{aligned}$$

$T((X_1, Y_1), \dots, (X_n, Y_n))$ ist unter H_0 für großes n , wie man zeigen kann, näherungsweise $\chi^2_{(k-1)(l-1)}$ -verteilt (siehe [WN] S. 91).

3. H_0 wird abgelehnt, wenn $T((x_1, y_1), \dots, (x_n, y_n)) > \chi^2_{(k-1)(l-1), 1-\alpha}$ gilt.

Bemerkung 10.5.1 Da die Testgröße eine Summe mit $k \cdot l$ Summanden ist, scheint zunächst die χ^2_{kl-1} -Verteilung als Grenzverteilung angemessen. Da jedoch $p_{1\bullet}, \dots, p_{k-1\bullet}$ sowie $p_{\bullet 1}, \dots, p_{\bullet l-1}$ aufgrund von Maximum Likelihood-Schätzungen geschätzt werden ($p_{\bullet l}, p_{k\bullet}$ ergeben sich aus $p_{\bullet l} = 1 - \sum_{j=1}^{l-1} p_{\bullet j}$, $p_{k\bullet} =$

$1 - \sum_{i=1}^{k-1} p_{i\bullet}$) ist es wegen $kl - (k-1) - (l-1) - 1 = (k-1)(l-1)$ plausibel, dass man auf eine $\chi^2_{(k-1)(l-1)}$ -Verteilung geführt wird.

Beispiel 10.5.2 Besteht ein Zusammenhang zwischen dem Alter eines Autofahrers und der Anzahl der Unfälle, in die er im Laufe eines Jahres verwickelt wird? Die folgende Kontingenztafel sei das Ergebnis einer Untersuchung zu dieser Fragestellung:

<i>Alter</i>	<i>Anzahl der Unfälle pro Jahr</i>				
	0	1	2	≥ 3	
18 – 29	748	74	31	9	862
30 – 39	821	60	25	10	916
40 – 49	786	51	22	6	865
50 – 59	720	66	16	5	807
≥ 60	672	50	15	7	744
	3747	301	109	37	4194

Lösung Hier ist $k = 5$ und $l = 4$. Wählen wir $\alpha = 0,1$ so finden wir $\chi^2_{12,0.9} = 18,55$. Als Wert der Testgröße ergibt sich 14,40. Die Annahme der Unabhängigkeit von Alter und Unfallneigung wird auf dem Niveau nicht verworfen. $\square$

11 Lineare Regression

Mit der linearen Regression versucht man lineare Zusammenhänge zwischen Merkmalen zu modellieren, z.B. die Abhängigkeit des Blutdrucks vom Alter:

$$\text{Blutdruck} = \alpha + \beta \cdot \text{Alter} + \text{zufällige Abweichungen}$$

11.1 Modell und Parameterschätzung

Wir gehen von n Messungen der Y -Variablen an den Stellen $x_1, \dots, x_n$ aus und erhalten so eine bivariate Stichprobe $(y_1, x_1), \dots, (y_n, x_n)$.

Wir treffen die Modellannahme

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (11.1)$$

wobei die Fehlervariablen ε_i unabhängig sind, mit $\mathbb{E}(\varepsilon_i) = 0$ und $\text{Var}(\varepsilon_i) = \sigma^2$, für alle i . Somit gilt $\mathbb{E}(Y_i) = \alpha + \beta x_i$ und $\text{Var}(Y_i) = \sigma^2$.

Schätzer für die unbekannten Parameter a und b werden mit der Methode der **kleinsten Quadrate** bestimmt: Sei $Q : \mathbb{R}^2 \rightarrow [0, \infty)$

$$Q(a, b) := \sum_{i=1}^n (y_i - (a + bx_i))^2. \quad (11.2)$$

$Q(a, b)$ ist die Summe der quadratischen Abstände zwischen den Beobachtungen (x_i, y_i) und der Gerade $y = a + bx$. Gesucht werden a, b , die Q minimieren, also die Ausgleichsgerade der Punkte (x_i, y_i) . Die so gewonnenen Schätzer heißen **kleinste Quadrate Schätzer**.

Satz 11.1.1 *Die kleinsten Quadrate Schätzer für α und β für (11.1) sind gegeben durch*

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\alpha} &= \bar{Y} - \hat{\beta} \bar{x} \end{aligned}$$

wobei $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ die Mittelwerte von $x_1, \dots, x_n$ und $Y_1, \dots, Y_n$

sind. Sie sind erwartungstreu und ihre Varianzen sind gegeben durch

$$\text{Var}(\hat{\alpha}) = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \quad (11.3)$$

$$\text{Var}(\hat{\beta}) = \sigma^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (11.4)$$

Ein erwartungstreuer Schätzer für die unbekannte Varianz σ^2 ist

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - (\hat{\alpha} + \hat{\beta}x_i))^2. \quad (11.5)$$

Beweis [LW], S. 193, Satz 3.63. Beweisidee für die erste Aussage ist Differenzieren von Q in (11.2) nach a, b und explizite Bestimmung des Minimums.

□

Sei $\hat{y}_i := \hat{\alpha} + \hat{\beta}x_i$. Setzt man $\hat{\alpha}$ und $\hat{\beta}$ in (11.2) ein, erhält man für die Stichprobe das Minimum für Q :

$$Q(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

11.2 Tests und Konfidenzintervalle

Wird zusätzlich eine Normalverteilung für die Fehlervariablen ε_i in (11.1) angenommen, also

$$\varepsilon_i \sim N(0, \sigma^2), \quad (11.6)$$

so können wir die folgenden Signifikanztests zum Niveau α (hier nicht zu verwechseln mit dem Ordinatenabschnitt α) und Konfidenzintervalle zum Niveau $= 1 - \alpha$ aufstellen. Zunächst eine Verteilungsaussage für die später verwendeten Testgrößen:

Satz 11.2.1 *Mit den Bezeichnungen aus Satz 11.1.1 und der Verteilungsannahme (11.6) gilt*

$$\frac{\hat{\alpha} - \alpha}{\text{se}(\hat{\alpha})} \sim t_{n-2}, \quad \frac{\hat{\beta} - \beta}{\text{se}(\hat{\beta})} \sim t_{n-2}$$

mit

$$\begin{aligned} \text{se}(\hat{\alpha}) &:= \hat{\sigma} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \\ \text{se}(\hat{\beta}) &:= \hat{\sigma} \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \end{aligned}$$

Beweis [LW], S. 193, Satz 3.63. $\square$

$\text{se}(\hat{\alpha})$ bzw. $\text{se}(\hat{\beta})$ erhält man aus (11.3) bzw. (11.4), in dem man dort σ^2 durch den Schätzer $\hat{\sigma}^2$ aus (11.5) für die Varianz ersetzt und heißt Standardfehler von $\hat{\alpha}$ bzw. $\hat{\beta}$.

11.2.1 Testen der Regressionsparameter

1. $H_0 : \beta = \beta_0, H_1 : \beta \neq \beta_0$.

2. Die Teststatistik

$$t = \frac{\hat{\beta} - \beta_0}{\text{se}(\hat{\beta})}, \quad \text{mit } \text{se}(\hat{\beta}) = \hat{\sigma} \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (11.7)$$

ist unter H_0 t_{n-2} verteilt. (s. Satz 11.2.1)

3. H_0 wird abgelehnt, wenn $|t| > t_{n-2, 1-\alpha/2}$.

Der Test für α ist analog bis auf Änderung der Testgröße wie in Satz 11.2.1.

11.2.2 Konfidenzintervalle

Mit dem Standardfehler $\text{se}(\hat{\beta})$ aus (11.7) lautet ein Konfidenzintervall für den Parameter β

$$\left[\hat{\beta} - t_{n-2, 1-\alpha/2} \cdot \text{se}(\hat{\beta}), \hat{\beta} + t_{n-2, 1-\alpha/2} \cdot \text{se}(\hat{\beta}) \right].$$

Ein individuelles Konfidenzintervall für den Wert $\alpha + \beta x = \mathbb{E}(Y)$ der (wahren) Regressionsgeraden an der Stelle x erhalten wir

$$\left[\hat{\alpha} + \hat{\beta}x - t_{n-2, 1-\alpha/2} \cdot \text{se}(\hat{\alpha} + \hat{\beta}x), \hat{\alpha} + \hat{\beta}x + t_{n-2, 1-\alpha/2} \cdot \text{se}(\hat{\alpha} + \hat{\beta}x) \right]$$

mit

$$\text{se}(\hat{\alpha} + \hat{\beta}x) = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (11.8)$$

11.2.3 Prognoseintervalle

Ist man an einer Prognose zu einem neuen Wert der Zielvariablen Y_{n+1} an der Stelle x_{n+1} interessiert, dann geht man wie folgt vor: Y_{n+1} ist unabhängig von $Y_1, \dots, Y_n$ und wird mit dem gleichen Regressionsmodell wie in (11.1) modelliert, also

$$Y_{n+1} = \alpha + \beta x_{n+1} + \varepsilon_{n+1}$$

mit $\varepsilon_{n+1} \sim N(0, \sigma^2)$. Dann ist

$$\mathbb{P} \left(|Y_{n+1} - (\hat{\alpha} + \hat{\beta}x_{n+1})| \leq t_{n-2, 1-\alpha/2} \cdot \sqrt{\text{se}(\hat{\alpha} + \hat{\beta}x_{n+1})^2 + \hat{\sigma}^2} \right) \geq 1 - \alpha.$$

Daraus erhält man das gewünschte Prognoseintervall zum Niveau α (nicht nur für den Schätzer $\hat{\alpha} + \hat{\beta}x_{n+1}$ von $\mathbb{E}(X_{n+1})$). Wegen

$$\text{se}(\hat{\alpha} + \hat{\beta}x_{n+1})^2 + \sigma^2 = \hat{\sigma}^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^n (x_k - \bar{x})^2} \right).$$

kommt im Vergleich zu (11.8) eine 1 unter der Wurzel dazu, das Prognoseintervall ist also breiter als das Konfidenzintervall.

Beispiel 11.2.2 Eine Messreihe bei der Alter x und Blutdruck y von 30 Patienten gemessen wurde, ergab:

$$\begin{aligned} \bar{x} &= 44,9, \bar{y} = 143,2, \sum_{i=1}^n (x_i - \bar{x})^2 = 6783,4, \\ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= 6584,3, \sum_{i=1}^n (y_i - \bar{y})^2 = 14788,4, \end{aligned}$$

- (a) Bestimmen Sie Schätzer $\hat{\alpha}, \hat{\beta}$ für die Regressionsgerade und $\hat{\sigma}^2$ für die Varianz.
- (b) Es wird vermutet, dass $\alpha = 50$ bzw. $\beta = 0$ gilt. Testen Sie die Hypothesen zum Niveau 0,05.

- (c) Bestimmen Sie ein Konfidenzintervall zum Niveau 0,95 für β und α .
- (d) Man bestimme ein Prognoseintervall zum Niveau 0,95 für den prognostizierten Blutdruck einer Person mit Alter $x = 50$.

Lösung (a) $\hat{\beta} = 0.97$, $\hat{\alpha} = 99.6$.

$$\begin{aligned} \sum_{i=1}^n (y_i - (\hat{\alpha} + \hat{\beta}x_i))^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= 14788,0 + \frac{6584,3^2}{6783,4} = 8397,0 \\ \hat{\sigma}^2 &= \frac{8397,0}{28} = 299,9 \end{aligned}$$

(b) Für die Hypothese $\beta = 0$ ergibt sich:

$$\begin{aligned} \text{se}(\hat{\beta}) &= 0,21 \\ t &= 4,61. \end{aligned}$$

Also wird $\beta = 0$ verworfen, denn $t_{28,0.975} = 2,05$. Im Fall $\alpha = 50$ ergibt sich die Testgröße 9,95, auch diese Hypothese wird abgelehnt.

(c) Konfidenzintervall für β : $[0,54;1,4]$

Konfidenzintervall für α : $[79,2;120]$

(d) Konfidenzintervall für die Prognose: $[111,9; 184,3]$

□

Literatur

- [B-MT] BAUER, HEINZ: *Maß- und Integrationstheorie*, 2. Auflage, Walter de Gruyter, 1992.
- [B-WT] BAUER, HEINZ: *Wahrscheinlichkeitstheorie*, 4. Auflage, Walter de Gruyter, 1990.
- [BM] BEICHELT, MONTGOMERY (Hrsg.): *Teubner-Taschenbuch der Stochastik*, Teubner, 2003.
- [El] ELSTRODT, JÜRGEN: *Maß- und Integrationstheorie*, 4. Auflage, Springer, 2004.
- [GS] GÄNSSLER, P., STUTE, W.: *Wahrscheinlichkeitstheorie*, Springer, 1977.
- [Kr] KRENGEL, ULRICH: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 6. Auflage, Vieweg 2002.
- [LW] LEHN, JÜRGEN UND WEGMANN, HELMUT: *Einführung in die Statistik*, 3. Auflage, Teubner 2000
- [P] PRUSCHA, HELMUT: *Vorlesungen über mathematische Statistik*, Teubner, 2000
- [WN] WITTING, NÖLLE: *Angewandte Mathematische Statistik*, Teubner, 1970

Index

- 1_A , 6
- $L^1(\Omega)$, 21
- $L^2(\Omega)$, 21
- $\mathbb{E}^{\mathcal{C}}(X)$, 79
- $\mathcal{B}^n$, 8
- $\mathcal{Q}^n$, 10
- $\sigma(X)$, 68
- $\sigma(\mathcal{E})$, 7

- asymtотisch erwartungstreu, 104
- asymtотisch normalverteilt, 105

- Bayessche Formel, 66
- bedingte
 - Dichte, 86
 - Erwartung, 79
 - Kovarianz, 81
 - Varianz, 81
 - Wahrscheinlichkeit, 62, 90
- Bias, 104
- Bildma, 26
- Borelmengen, 8

- χ^2 -Streuungstest, 134

- dichotom, 102
- Dichte, 24
 - bedingte, 86
- Dirac-Ma, 17
- diskrete ZV, 37
- dominierten Konvergenz, 21
- $\cap$ -stabil, 10

- Ereignis, 29
- Erwartung
 - bedingte, 79
 - Versionen der bedingten-, 79
- erwartungstreu, 104
- Erwartungswert, 38

- Erzeuger, 7

- F -Test, 135
- Faktorisierungslemma, 82
- Faltung, 72
- fast berall, 19
- fast sicher, 19
- Fehler
 - 1. Art, 128
 - 2. Art, 128
- Feller Bedingung, 98
- fr fast alle, 19
- Funktion
 - Beta-, 61
 - erzeugende
 - charakteristische, 54
 - momenten-, 52
 - wahrscheinlichkeits-, 51
 - Gamma-, 61
 - konvexe, 60

- Gtefunktion, 128
- Gau-Test, 130
- gemeinsame
 - Verteilung, 56
 - Verteilungsfunktion, 56
- Gesetz der groen Zahlen
 - schwaches, 96
- Grundgesamtheit, 101

- Hypothese
 - Alternative-, 127
 - Null-, 127

- iid, 95
- Indikatorfunktion, 6
- integrierbar, 16

- kleinsten Quadrate, 145

- Konfidenzbereich, 129
- Konfidenzintervall, 114
- konsistent, 105
- Konvergenz
 - fast sicher, 96
 - stochastisch, 96
- Korrelationskoeffizient, 58
- korreliert
 - entgegengesetzt, 58
 - vollständig, 58
- Kovarianz, 58
 - bedingte, 81
- Kovarianzmatrix, 59
- L-B Maß, 11
- Laplacesche Wahrscheinlichkeit, 30
- Lebesgue-Borel Maß, 11
- Likelihoodfunktion, 109
- Lindeberg-Bedingung, 97
- Lyapunov Bedingung, 98
- Maßraum, 4
- Maß, 3
- Median, 42
- Merkmal, 101
- messbare Funktion, 11
- Messraum, 3
- mittlere quadratische Fehler, 104
- Moment, 42
- monotone Konvergenz, 21
- multinomial, 137
- Normalverteilung
 - mehrdimensionale, 60
- Nullhypothese, 127
- Nullmenge, 18
- p -Wert, 128
- Punktschätzer, 103
- Quader, 10
- Quantil, 42
- Randverteilung, 57
- Realisation, 34
- Satz von
 - Bayes, 66
 - der dominierten Konvergenz, 21
 - der monotonen Konvergenz, 21
 - der totalen Wahrscheinlichkeit, 66
 - Fubini, 23
 - Glivenko-Cantelli, 100
 - Kolmogorov, 100
 - Radon-Nykodym, 25
 - Student, 118
- Schätzer, 103
 - kleinste Quadrate-, 145
 - Maximum Likelihood-, 110
 - Momenten-, 107
- Schiefe, 42
- σ -Algebra, 3
 - erzeugte, 7, 68
- σ -additiv, 4
- Signifikanzniveau, 127
- Standardabweichung, 41
- Standardisierung, 41
- stetige ZV, 37
- Stetigkeit
 - von oben, 9
 - von unten, 9
- Stichprobe, 101, 102
- Test
 - asymptotischer Binomial-, 135
 - χ^2 -Streuungs-, 134
 - F -, 135
 - Gauß-, 130
 - Kolmogorov-Smirnov-, 136
 - t -, 132
- Teststatistik, 127
- Treppenfunktion, 14
- Unabhängigkeit von
 - Mengensystemen, 64

- Ereignissen, 64
- σ -Algebren, 68
- Zufallsvariablen, 69
- Ungleichung von
 - Jensen, 44
 - Markov, 43
 - Tschebyschev, 43
- unkorreliert, 58
- Variable, 101
- Varianz, 41
 - bedingte, 81
- Versionen der bedingten Erwartung, 79
- Verteilung
 - Bernoulli-, 45
 - Binomial-, 45
 - χ^2 , 115
 - Exponential-, 49
 - $F_{m,n}$, 117
 - Gamma-, 50
 - Gauß-, 49
 - geometrische, 46
 - Gleich-, 48
 - hypergeometrische-, 47
 - negative Binomial-, 46
 - Normal
 - mehrdimensional, 60
 - Normal-, 49
 - Poisson-, 46
 - Rechteck-, 48
 - Standardnormal-, 49
 - t_n , 116
 - uniforme-, 48
- Verteilungsfunktion, 35
 - empirische, 99
 - Kolmogorovsche, 100
- Verwerfungsbereich, 127
- Wahrscheinlichkeitsfunktion, 37, 56
- Wahrscheinlichkeitsmaß, 30
- Wahrscheinlichkeitsraum, 30
- Zentraler Grenzwertsatz, 98
- Zufallsvariable, 33
 - n -dimensionale, 56
 - diskrete, 37
 - stetig, 57
 - stetig (verteilte), 37