



ulm university universität
uulm

Vorlesungsskript

Stochastik für Wirtschaftswissenschaftler

Dr. Katharina Best

Wintersemester 2010/2011

Inhaltsverzeichnis

1	Einleitung	5
1.1	Der Begriff <i>Stochastik</i>	5
1.2	Geschichtliche Entwicklung	5
1.3	Problemstellungen der Stochastik	6
2	Ergebnisse, Ereignisse und Wahrscheinlichkeit	7
2.1	Mengen	7
2.2	Mengensysteme	9
2.3	Zufallsereignisse	10
2.4	Wahrscheinlichkeiten	11
2.4.1	Axiome von Kolmogorov	11
2.4.2	Laplacescher Wahrscheinlichkeitsraum	12
2.4.3	Kombinatorische Wahrscheinlichkeiten	13
2.5	Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten	15
2.5.1	Unabhängige Ereignisse	15
2.5.2	Bedingte Wahrscheinlichkeit	16
3	Zufallsvariablen und Verteilungen	21
3.1	Grundbegriffe	21
3.2	Diskrete Zufallsvariablen	23
3.2.1	Charakterisierung diskreter Zufallsvariablen	24
3.2.2	Spezielle/Wichtige diskrete Verteilungen	27
3.2.3	Unabhängige diskrete Zufallsvariablen	33
3.3	Stetige Zufallsvariablen	36
3.3.1	Charakterisierung stetiger Zufallsvariablen	37
3.3.2	Spezielle/Wichtige stetige Verteilungen	38
3.4	Mehrdimensionale Verteilungen	40
3.4.1	Gemeinsamen und marginale Verteilungen	40
3.4.2	Erwartungswert und Varianz	43
3.4.3	Bedingte Verteilungen	46
3.4.4	Transformation von Dichten	48
4	Abschätzungen und Grenzwertsätze	53
4.1	Ungleichungen	53
4.2	Grenzwertsätze	55
4.2.1	Gesetze der großen Zahlen	58
4.2.2	Zentraler Grenzwertsatz	61
5	Grundkonzepte der Statistik	63
5.1	Stichproben	63
5.2	Beispiele für Stichprobenfunktionen	65
5.2.1	Stichprobenmittel	65
5.2.2	Stichprobenvarianz	66

Inhaltsverzeichnis

5.2.3	Empirische Verteilungsfunktion	67
6	Parameterschätzung	69
6.1	Parametrisierung	69
6.2	Punktschätzung	70
6.2.1	Eigenschaften von Schätzern	70
6.2.2	Methoden der Punktschätzung	72
6.2.3	<i>Einschub: Grundlagen der linearen Regression</i>	79
6.3	Prüfverteilungen	82
6.3.1	χ^2 -Verteilung	82
6.3.2	t -Verteilung	83
6.3.3	F -Verteilung	83
6.4	Intervallschätzung	85
6.4.1	Normalverteilungsannahme	86
6.4.2	Asymptotische Konfidenzintervalle	92
7	Statistische Tests	95
7.1	Prinzipien des Testens	95
7.2	Aufbau eines statistischen Tests	95
7.3	Parameter tests bei Normalverteilung	98
7.4	p-Wert versus Konfidenzniveau	103
7.5	Asymptotische Tests	104
7.5.1	χ^2 -Anpassungstest	104
7.5.2	χ^2 -Unabhängigkeitstest	105

1 Einleitung

»Im Grunde ist Wahrscheinlichkeitsrechnung nur normaler gesunder Menschenverstand, ausgedrückt durch Mathematik. Sie versetzt uns in die Lage, Dinge exakt abzuschätzen, die wir in einer Art Instinkt fühlen, aber nicht erklären können.«
Pierre-Simon Laplace

1.1 Der Begriff *Stochastik*

Die Wahrscheinlichkeitsrechnung ist eine Teildisziplin der Stochastik. Das Wort *Stochastik* kommt aus dem Griechischen *στωχαστική* – Die Kunst des Vermutens – abgeleitet von *στωχωξ* (Vermutung, Ahnung, Ziel). Der Begriff wurde von Jacob Bernoulli in seinem Buch *Ars conjectandi* 1713 geprägt, in dem das erste Gesetz der großen Zahlen bewiesen wurde.

Die Stochastik beschäftigt sich mit den Ausprägungen und quantitativen Merkmalen des Zufalls. Dabei ist noch zu klären, was Zufall ist. Gibt es Zufall überhaupt? Und was würde man als Zufall betrachten? Mit diesem Aspekt beschäftigt sich die Philosophie¹. Für die moderne Mathematik ist der Zufall eine Arbeitshypothese, die es ermöglicht, Vorgänge in der Natur, Technik oder Wirtschaft zu beschreiben und zu analysieren. Das Konzept des Zufalls eröffnet einen Weg, Vorgänge mit Wahrscheinlichkeiten zu belegen, diese zu berechnen, zu quantifizieren und zu vergleichen. Allerdings muss man an dieser Stelle erneut nachhaken: Was ist Wahrscheinlichkeit? Wie können wir Wahrscheinlichkeit interpretieren? Bei einem Spiel, für das wir die Regeln festlegen, ist die Antwort noch vergleichsweise einfach. *Zwei Ausgänge sind gleich wahrscheinlich, wenn keiner bevorzugt wird.* Allerdings muss man aus einer weitergehenden Quelle wissen, dass keiner bevorzugt ist. Bei vielen Fragestellungen steht der Begriff der Wahrscheinlichkeit für das Maß an Sicherheit resp. Unsicherheit². Diese, der Wahrscheinlichkeit innewohnende, Ambivalenz sollte man stets im Hinterkopf behalten.

1.2 Geschichtliche Entwicklung

Die Stochastik wurde zu Beginn über das Gebiet der Glücksspiele entwickelt. Die ersten Würfelspiele sind in Ägypten (ca. 3500 v. Chr.) nachgewiesen, und wurden in Griechenland und vor allem im römischen Reich fortgesetzt. Zugleich gab es im Handel in Bezug auf Versicherungen von Schiffstransporten das Bedürfnis nach Berechenbarkeit. Die älteste bekannte Form der Versicherungsverträge stammt aus Babylon (4-3 T. Jahre v. Chr.). Die ersten Sterbetafeln in der Lebensversicherung stammen von dem römischen Juristen Ulpian (220 v. Chr.), die erste datierte Police aus dem Jahr 1347 aus Genua.

¹siehe hierzu <http://plato.stanford.edu/entries/chance-randomness/>

²siehe hierzu <http://plato.stanford.edu/entries/probability-interpret/>
und <http://de.wikipedia.org/wiki/Wahrscheinlichkeit>

1 Einleitung



Abbildung 1.1: Von links nach rechts: Gerolamo Cardano, Jakob Bernoulli, Pierre-Simon Laplace und Andrej Nikolajewitsch Kolmogorov.

Der erste Wissenschaftler, der sich mit diesen Fragestellungen aus mathematischer Sicht beschäftigt hat, war Gerolamo Cardano. In seinem *Buch der Glücksspiele (Liber de Ludo Aleae)* beschreibt er 1524 zum ersten Mal für den Spieler vorteilhafte Ereignisse und deren Kombinationen beim Würfelspiel, verwendet Binomialkoeffizienten und auch als erster den Anteil

$$\frac{\text{Anzahl vorteilhafter Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

als Maß für die Wahrscheinlichkeit. Außerdem ist er auch ein Praktiker: Sein Universitätsgehalt bessert er mit dem durch sein Wissen beim Glücksspiel verdienten Geld auf. Etwa hundert Jahre später diskutieren Blaise Pascal und Pierre de Fermat die Wahrscheinlichkeit betreffende Probleme in ihrem Briefwechsel. Wieder etwa hundert Jahre später gelang Pierre-Simon Laplace 1812 in seinem Buch *Théorie Analytique des Probabilités* eine mathematische Behandlung der Wahrscheinlichkeit.

Allerdings erlaubten diese Betrachtungen noch keine Aussage bei Problemen, bei denen die Ereignisse gar nicht alle beschreibbar und damit auf eine Gesamtmenge aller gleichwahrscheinlicher Ereignisse reduzierbar waren. An dieser Stelle wurde von David Hilbert ein axiomatischer Ansatz, ähnlich wie er in anderen Gebieten der Mathematik bereits verwendet wurde, gefordert. Basierend auf der Maß- und Integrationstheorie von Émile Borel und Henri Léon Lebesgue, führte Andrei Nikolajewitsch Kolmogorow in seinem Werk *Grundbegriffe der Wahrscheinlichkeitsrechnung* 1933 Axiome der Wahrscheinlichkeitstheorie ein.

1.3 Problemstellungen der Stochastik

Zur Stochastik kann man grob die *Wahrscheinlichkeitstheorie*, die die Grundlagen lehrt, und die *Statistik*, die sich mit dem Umgang mit Daten befasst, zählen. Daraus lassen sich die typischen Fragestellungen wie

- (i) die Modellierung von Zufallsexperimenten, d. h. deren adäquate theoretische Beschreibung,
- (ii) die Bestimmung von Wahrscheinlichkeiten von Ereignissen,
- (iii) die Bestimmung von Mittelwerten, Varianzen und Verteilungsgesetzen von Zufallsvariablen,
- (iv) die Näherung mittels Grenzwertsätzen,
- (v) die Schätzung von Modellparametern,
- (vi) die Prüfung statistischer Hypothesen formulieren.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

In der Wahrscheinlichkeitsrechnung interessieren wir uns für Ausgänge von Zufallsexperimenten. Um sie beschreiben zu können, werden die verschiedenen Ausgänge als Mengen betrachtet und ihnen wird eine Zahl zugeordnet, die Wahrscheinlichkeit ihres Eintretens.

2.1 Mengen

Im folgenden definieren wir einige Grundbegriffe der Mengenlehre und betrachten zugehörige Mengenoperationen.

Definition 2.1 (Menge): Eine *Menge* ist eine Zusammenfassung verschiedener Objekte zu einem Ganzen. Die einzelnen Objekte werden *Elemente* genannt.

Mengen werden entweder durch eine Auflistung ihrer Elemente oder durch eine definierende Eigenschaft angegeben.

Beispiel 2.2: Sei A eine Menge der Zahlen auf einem Würfel. Wir können A als

$$A = \{1, 2, 3, 4, 5, 6\} \quad \text{oder} \quad A = \{n : n \text{ ist eine natürliche Zahl mit } 1 \leq n \leq 6\}$$

schreiben.

Notation: Die *leere* Menge wird mit \emptyset dargestellt.

Definition 2.3 (Grundbegriffe der Mengenlehre): (i) Die Eigenschaft x ist ein Element von A wird mit $x \in A$ notiert, das Gegenteil mit $x \notin A$.

(ii) Für $A \subset B$, d. h. A ist *Teilmenge* von B , gilt dass für alle $x \in A$ auch $x \in B$ gilt.

(iii) Die *Schnittmenge* $A \cap B$ ist die Menge aller Elemente, die sowohl in A als auch in B sind. Für alle $x \in A \cap B$ gilt $x \in A$ und $x \in B$.

(iv) Die *Vereinigungsmenge* $A \cup B$ ist die Menge aller Elemente, die in A oder in B sind. Für alle $x \in A \cup B$ gilt $x \in A$ oder $x \in B$.

(v) Die *Differenzmenge* $A \setminus B$ ist die Menge aller Elemente, die in A aber nicht in B sind. Für alle $x \in A \setminus B$ gilt $x \in A$ und $x \notin B$.

(vi) Die *Komplementärmenge* A^c von $A \subset \Omega$ wird bezüglich einer Grundmenge Ω definiert und ist die Menge aller Elemente, die in Ω sind, aber nicht in A .

(vii) Die *Potenzmenge* $\mathcal{P}(A)$ ist die Menge aller Teilmengen von A . Für alle $M \in \mathcal{P}(A)$ gilt $M \subset A$.

(viii) Die *Mächtigkeit* von A bezeichnet die Anzahl der Elemente von A . Notiert wird sie mit $|A| = \#\{x : x \in A\}$.

(ix) Die *Gleichheit* von zwei Mengen A und B ist gegeben, wenn $A \subset B$ und $B \subset A$ gilt.

(x) Zwei Mengen A und B sind *disjunkt*, wenn sie keine gemeinsamen Elemente haben, d. h. $A \cap B = \emptyset$.

(xi) Die Mengen A_1, A_2, \dots heißen *paarweise disjunkt*, falls für alle Paare i und j mit $i \neq j$ die Mengen A_i und A_j disjunkt sind.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

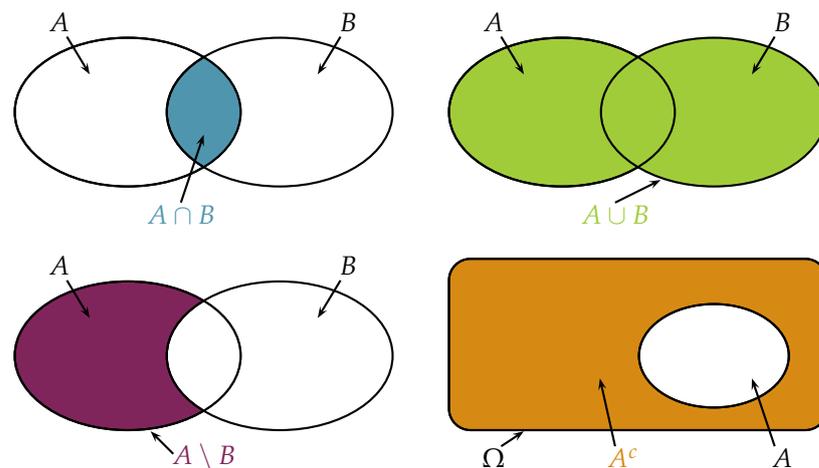


Abbildung 2.1: Venn-Diagramme zu einigen Mengenoperationen aus Definition 2.3 zur Verdeutlichung.

Zur Darstellung eignen sich häufig Venn-Diagramme, siehe auch die Abbildung 2.1 zu den Abschnitten (iii) – (vi).

Beispiel 2.4: (a) Sei $A = \{1, 3, 7, 8\}$ und $B = \{1, 2, 4, 7\}$. Dann gilt $7 \in B$, aber $3 \notin B$.

(b) Sei $\Omega = \{1, 3, 4\}$. Dann ist $\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{3\}, \{4\}, \{1, 3\}, \{1, 4\}, \{3, 4\}, \Omega\}$.

(c) Sei $\Omega = \{1, 2, 3, 4, 5\}$ und seien $A_1 = \{2, 4, 5\}$ und $A_2 = \{x \in \Omega : x \text{ ungerade}\}$. Dann gilt

$$A_1 \cup A_2 = \{1, 2, 3, 4, 5\} = \Omega \quad A_1 \cap A_2 = \{5\}$$

$$A_1 \setminus A_2 = \{2, 4\} \quad A_2 \setminus A_1 = \{1, 3\} \quad A_1^c = \{1, 3\}$$

Definition 2.5 (Symmetrische Differenz): Die *symmetrische Differenz* zweier Mengen A_1 und A_2 ist definiert als

$$A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$$

Um kompliziertere Mengen und Mengenbeziehungen beschreiben zu können, benötigen wir noch einige Rechenregeln.

Satz 2.6 (Rechenregeln für Mengen): Seien A, B und C Mengen.

(i) *Eindeutigkeitsgesetze*

$$A \cup \emptyset = A, \quad A \cap \emptyset = \emptyset \quad \text{und} \quad A \cup \Omega = \Omega, \quad A \cap \Omega = A$$

(ii) *Kommutativgesetze*

$$A \cap B = B \cap A \quad \text{und} \quad A \cup B = B \cup A$$

(iii) *Assoziativgesetze*

$$(A \cap B) \cap C = A \cap (B \cap C) \quad \text{und} \quad (A \cup B) \cup C = A \cup (B \cup C)$$

(iv) *Distributivgesetze*

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad \text{und} \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

(v) *De Morgansche Regeln*

$$(A \cap B)^c = A^c \cup B^c \quad \text{und} \quad (A \cup B)^c = A^c \cap B^c$$

(vi) Aus $A \subset B$ folgt für die Inklusion bei Negation $B^c \subset A^c$.

(vii) Die Differenzmenge $A \setminus B$ lässt sich als Schnitt darstellen durch $A \cap B^c$.

2.2 Mengensysteme

Für die Arbeit mit Mengen ist es sehr hilfreich, wenn die Familie von Mengen, die betrachtet wird, bezüglich der Operationen \cap, \cup und \setminus abgeschlossen ist.

Definition 2.7 (Algebra): Eine nichtleere Familie \mathcal{F} von Teilmengen von Ω heißt *Algebra*, wenn

(A1) aus $A \in \mathcal{F}$ folgt $A^c \in \mathcal{F}$,

(A2) aus $A_1, A_2 \in \mathcal{F}$ folgt $A_1 \cup A_2 \in \mathcal{F}$

gilt.

Beispiel 2.8: (a) Sei Ω gegeben. Dann ist $\mathcal{P}(\Omega)$ eine Algebra.

(b) Sei Ω gegeben. Dann ist $\mathcal{F} = \{\emptyset, \Omega\}$ eine Algebra.

(c) Sei $\Omega = \{1, 2, 3, 4, 5\}$ und $\mathcal{F} = \{\emptyset, \{1, 3\}, \{2, 4, 5\}, \Omega\}$. Dann ist \mathcal{F} eine Algebra.

(d) Sei $\Omega = \{2, 3, 4\}$ und $\mathcal{F} = \{\emptyset, \{2\}, \{3\}, \Omega\}$. Dann ist \mathcal{F} keine Algebra.

Lemma 2.9 (Eigenschaften einer Algebra): Sei \mathcal{F} eine Algebra und $A_1, A_2, \dots, A_n \in \mathcal{F}$. Dann gilt

(i) $\emptyset, \Omega \in \mathcal{F}$,

(ii) $A_1 \cap A_2 \in \mathcal{F}$,

(iii) $A_1 \setminus A_2 \in \mathcal{F}$,

(iv) $\bigcup_{i=1}^n A_i \in \mathcal{F}$ und $\bigcap_{i=1}^n A_i \in \mathcal{F}$.

BEWEIS (i) Da \mathcal{F} nicht leer ist, gibt es ein $A \subset \Omega$ mit $A \in \mathcal{F}$. Damit ist wegen (A1) $A^c \in \mathcal{F}$ und wegen (A2) $\Omega = A \cup A^c \in \mathcal{F}$. Wegen (A1) gilt auch $\emptyset = \Omega^c \in \mathcal{F}$.

(ii) $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{F}$.

(iii) $A_1 \setminus A_2 = A_1 \cap A_2^c = (A_1^c \cup A_2)^c \in \mathcal{F}$.

(iv) Beweis durch Induktion: Sei $n = 2$ und $A_1, A_2 \in \mathcal{F}$. Dann ist $A_1 \cup A_2 \in \mathcal{F}$ nach (A2).

Übergang $n \mapsto n + 1$: Seien $A_1, A_2, \dots, A_n, A_{n+1} \in \mathcal{F}$.

Dann ist nach Induktionsvoraussetzung $\bigcup_{i=1}^n A_i \in \mathcal{F}$ und nach (A2) gilt $\bigcup_{i=1}^{n+1} A_i = \left(\bigcup_{i=1}^n A_i \right) \cup A_{n+1} \in \mathcal{F}$.

Dass $\bigcap_{i=1}^n A_i \in \mathcal{F}$ gilt, wird analog mit Hilfe von (ii) gezeigt. □

Um Grenzwerte von Mengen bilden zu können, benötigen wir ein Mengensystem \mathcal{F} , das nicht nur bezüglich Vereinigung und Durchschnitt *endlich* vieler Mengen, sondern bezüglich Vereinigung und Durchschnitt von *abzählbar unendlich* vielen Mengen abgeschlossen ist. Dazu nimmt man die folgende Bedingung hinzu.

Definition 2.10 (σ -Algebra): Sei \mathcal{F} eine Algebra.

(A3) Aus $A_1, A_2, \dots \in \mathcal{F}$ folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Lemma 2.11: Sei \mathcal{F} eine σ -Algebra und $A_1, A_2, \dots \in \mathcal{F}$. Dann gilt $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

Beispiel 2.12: (a) Sei Ω gegeben. Dann ist $\mathcal{P}(\Omega)$ eine Algebra.

(b) Sei $\Omega = \mathbb{R}$ und \mathcal{F} die Klasse von Teilmengen aus Ω , die aus endlichen Vereinigungen von disjunkten Intervallen der Form $(-\infty, a]$, $(b, c]$ und (d, ∞) mit $a, b, c, d \in \mathbb{R}$ besteht. Dann ist \mathcal{F} eine Algebra. Aber $[b, c] = \bigcap_{n=1}^{\infty} \left(b - \frac{1}{n}, c\right] \notin \mathcal{F}$ und somit ist \mathcal{F} keine σ -Algebra.

2.3 Zufallsereignisse

Die Ereignisse, die bei einem Zufallsexperiment eintreten können, werden als Mengen modelliert. Sei E ein Grundraum.

Definition 2.13 (Ergebnisse): Die Menge $\Omega \subset E$ aller möglichen Ausgänge eines Zufallsexperiments nennen wir *Ergebnisraum*, *Grundraum* oder *Stichprobenraum*. Die Elemente $\omega \in \Omega$ heißen *Ergebnisse* oder *Elementarereignisse*.

Beispiel 2.14: (a) *Bernoulli-Experiment*: Einmaliger Münzwurf. $\Omega = \{\text{Kopf, Zahl}\}$ oder $E = \mathbb{N}$ und $\Omega = \{0, 1\}$.

(b) Einmaliger Würfelwurf. $E = \mathbb{N}$ und $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(c) n -maliger Würfelwurf. $E = \mathbb{N}^n$ und

$\Omega = \{1, 2, 3, 4, 5, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, 2, 3, 4, 5, 6\} \text{ für } 1 \leq i \leq n\}$. Zu beachten ist, dass hier nicht jeweils ein ω_i das Ergebnis ist, sondern $\omega = (\omega_1, \dots, \omega_n)$.

(d) Unendlich oft ausgeführter Münzwurf. $\Omega = \{0, 1\}^{\mathbb{N}} = \{(\omega_i)_{i \in \mathbb{N}} : \omega_i \in \{0, 1\} \text{ für } i \in \mathbb{N}\}$.

(e) Kurs einer Aktie innerhalb eines Jahres. $\Omega = \{f(t) : 0 \leq t \leq 365, f \in C(\mathbb{R})\}$, wenn $t = 0$ genau dem Beginn des Jahres, 1. Januar 2010 um 0 Uhr entspricht, und eine Zeiteinheit einem Tag.

Die letzten zwei Beispiele zeigen, dass als Ergebnis eines Zufallsexperiments auch Folgen und Funktionen auftreten können und Ω deswegen nicht nur endlich sein kann, sondern auch unendlich abzählbar oder sogar überabzählbar.

Oft interessiert man sich nicht für einzelne Ergebnisse, sondern für Mengen von Ergebnissen mit einer bestimmten Eigenschaft. Diese Mengen nennen wir *Ereignisse*.

Beispiel 2.15: (a) Zum Beispiel 2.14 (b) ist *Es wird eine ungerade Zahl gewürfelt* ein Ereignis,

$$A_1 = \{1, 3, 5\}.$$

(b) Zum Beispiel 2.14 (d) ist *Die ersten 10 Würfe ergeben Zahl* ein Ereignis,

$$A_2 = \{(\omega_i)_{i \in \mathbb{N}} : \omega_i = 1 \text{ für } 1 \leq i \leq 10 \text{ und } \omega_i \in \{0, 1\} \text{ für } i \geq 11\}.$$

(c) Zum Beispiel 2.14 (e) ist *Der Kurs bleibt unter dem Wert 400* ein Ereignis,

$$A_3 = \left\{ f(t) : 0 \leq t \leq 365, f \in C(\mathbb{R}) \text{ und } \sup_t f(t) < 400 \right\}.$$

Mit dieser Modellierung ist es klar, wieso wir vorher die Mengensysteme definiert haben. Bevor wir uns der Wahrscheinlichkeit zuwenden, schauen wir uns die Formulierung der Ereignisse als Mengen noch etwas genauer an.

Anmerkung: (a) Wird bei einem Zufallsexperiment das Ergebnis ω erzielt, so sagen wir: Das Ereignis A tritt ein, falls $\omega \in A$.

(b) Interessiert man sich für das Ereignis A^c , so sagt man, dass A nicht eintritt.

(c) Ist $A = \emptyset$, so wird A das *unmögliche Ereignis* genannt.

(d) Ist $A = \Omega$, so wird A das *sichere* oder *wahre Ereignis* genannt.

(e) Gilt $A \subset B$, so folgt aus dem Eintreten von A auch, dass B eintritt.

2.4 Wahrscheinlichkeiten

Den auf diese Weise definierten Ereignissen wollen wir jetzt eine Kennzahl zuordnen, die wir Wahrscheinlichkeit nennen. Eine solche Zahl sollte nicht-negativ sein. Sie sollte normiert sein, damit wir die Ereignisse bewerten können. Außerdem wollen wir Ereignisse sinnvoll verbinden können.

2.4.1 Axiome von Kolmogorov

Die oben genannten Forderungen wurden von Kolmogorov in drei Axiomen festgehalten, die für den allgemeinen Fall wie folgt aussehen:

Definition 2.16 (Kolmogorov-Axiome): Ein *Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, \mathbb{P})$ ist ein Tripel bestehend aus einer Menge Ω , einer σ -Algebra \mathcal{F} und einer Funktion $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$, *Wahrscheinlichkeitsmaß* genannt, mit den Eigenschaften

- (i) *Nichtnegativität* $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{F}$.
- (ii) *Normiertheit* $\mathbb{P}(\Omega) = 1$.
- (iii) *σ -Additivität* Sind die Mengen $A_1, A_2, \dots \in \mathcal{F}$ paarweise disjunkt, so gilt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Anmerkung: Im abzählbar unendlichen Fall kann man die geforderte σ -Algebra einfach durch die Potenzmenge $\mathcal{P}(\Omega)$ ersetzen. Im endlichen Fall kann man das Axiom der σ -Additivität vereinfachend noch durch die Additivität ersetzen, da sich jede unendliche Vereinigung in eine endliche Vereinigung umschreiben lässt.

Aus diesen Axiomen lassen sich nun Rechenregeln ableiten.

Satz 2.17 (Rechenregeln für Wahrscheinlichkeiten): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, A_1, A_2, \dots \in \mathcal{F}$. Dann gilt

- (i) $\mathbb{P}(\emptyset) = 0$,
- (ii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- (iii) Aus $A_1 \subset A_2$ folgt $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$,
- (iv) $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$,
- (v) $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.

BEWEIS (i) Wir wissen, dass $\Omega = \Omega \cup \emptyset$ und $\Omega \cap \emptyset = \emptyset$. Dann ist

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset),$$

und damit $\mathbb{P}(\emptyset) = 0$.

(ii) Auch hier wissen wir, dass $A \cap A^c = \emptyset$ und $A \cup A^c = \Omega$ und erhalten

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$$

(iii) Da $A_1 \subset A_2$ ist, lässt sich A_2 schreiben als $A_2 = (A_2 \setminus A_1) \cup A_1$, wobei $A_2 \setminus A_1$ und A_1 disjunkt sind und damit folgt

$$\mathbb{P}(A_2) = \mathbb{P}((A_2 \setminus A_1) \cup A_1) = \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_1) \geq \mathbb{P}(A_1)$$

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

(iv) Wegen $A_1 \cup A_2 = (A_1 \setminus (A_1 \cap A_2)) \cup (A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2))$, wobei die drei Mengen disjunkt sind, gilt

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1 \setminus (A_1 \cap A_2)) + \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_2 \setminus (A_1 \cap A_2)) \\ &= (\mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2)) + \mathbb{P}(A_1 \cap A_2) + (\mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \end{aligned}$$

mit den Erkenntnissen aus (iii).

(v) Dies folgt durch Induktion aus (iii). \square

Eine Verallgemeinerung des Satzes 2.17 (iv) ist der *Allgemeine Additionssatz*, auch *Siebformel von Poincaré-Sylvester* genannt, bei dem für n Mengen alle k -weisen Schnittmengen (Paare, Tripel, usw.) betrachtet und entsprechend addiert oder subtrahiert werden.

Satz 2.18 (Siebformel von Poincaré-Sylvester): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots, A_n \in \mathcal{F}$. Dann gilt

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} \mathbb{P}(A_{k_1} \cap \dots \cap A_{k_i}) \quad (2.1)$$

BEWEIS durch Induktion mit Hilfe des Satzes 2.17 (iv). \square

2.4.2 Laplacescher Wahrscheinlichkeitsraum

Sei Ω endlich, also $|\Omega| < \infty$. Dann gilt natürlich auch für alle $A \subset \Omega$, dass $|A| < \infty$. Dann heißt ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ ein *endlicher Wahrscheinlichkeitsraum*.

Definition 2.19 (Laplacescher Wahrscheinlichkeitsraum): Sei Ω endlich und $\mathcal{F} = \mathcal{P}(\Omega)$. Ein endlicher Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, bei dem alle Elementarereignisse gleichwahrscheinlich sind, d. h. $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$, heißt *Laplacescher Wahrscheinlichkeitsraum*.

Anmerkung: Sei $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein Laplacescher Wahrscheinlichkeitsraum. Für alle $A \subset \Omega$ gilt dann $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ wegen der σ -Additivität von Wahrscheinlichkeitsmaßen. Die so definierte Wahrscheinlichkeit heißt *Laplace-Wahrscheinlichkeit*.

Beispiel 2.20 (Zweimaliges Würfeln): Als erstes muss man sich über Ω klar werden:

$$\Omega = \{\omega = (i, j) : 1 \leq i, j \leq 6, i = \text{Augenzahl beim 1. Wurf}, j = \text{Augenzahl beim 2. Wurf}\}$$

Dann ist $|\Omega| = 36$ und wir nehmen $\mathcal{F} = \mathcal{P}(\Omega)$. Sei $\mathbb{P} : \Omega \mapsto [0, 1]$ ein Laplace-Wahrscheinlichkeitsmaß, d. h. für alle $\omega_1, \omega_2 \in \Omega$ ist $\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\})$ und $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$, somit $\mathbb{P}(\{\omega\}) = \frac{1}{36}$ für alle $\omega \in \Omega$. Das Tripel $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ bildet einen Laplaceschen Wahrscheinlichkeitsraum.

Sei nun $A = \{\text{Gesamtaugenzahl} \geq 10\} = \{(4, 6), (5, 6), (6, 6), (5, 5), (6, 5), (6, 4)\}$. Dann ist $|A| = 6$ und $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{1}{6}$.

Wir wollen noch ein Beispiel betrachten, welches vielleicht etwas unintuitiv ist.

Beispiel 2.21 (Geburtstagsproblem): Auf einer Party sind 30 Personen. Jemand wettet, dass dabei zwei Personen mit dem Geburtstag am gleichen Tag sind. Wetten Sie dagegen? (Nachdem Sie sich durch einen raschen Blick in die Runde versichert haben, dass keine offensichtlichen Zwillinge mitfeiern.)

Mathematisch formuliert ist $N = 365$ die Anzahl der möglichen Geburtstage und $n = 30$ die Anzahl der Leute. Dann ist $\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\}\}$ und $|\Omega| = N^n$. Das Ereignis

$$A_n = \{\text{mindestens 2 Personen haben am gleichen Tag Geburtstag}\}$$

lässt sich schreiben als

$$= \left\{ \omega = (\omega_1, \dots, \omega_n) \in \Omega : \text{es gibt } 1 \leq i, j \leq n \text{ mit } i \neq j : \omega_i = \omega_j \right\}.$$

Zur Berechnung der Wahrscheinlichkeit wählen wir den Ansatz über das Gegenereignis $\mathbb{P}(A_n) = 1 - \mathbb{P}(A_n^c)$ und

$$A_n^c = \left\{ \omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_i \neq \omega_j \text{ für alle } 1 \leq i, j \leq n \text{ mit } i \neq j \right\}$$

Die Mächtigkeit des Ereignisses ist dann

$$|A_n^c| = \underbrace{N}_{\text{frei Wahl}} \cdot \underbrace{(N-1)}_{\text{nicht am gewählten Tag}} \cdot \underbrace{(N-2)}_{\text{nicht an den 2 gewählten Tagen}} \cdot \dots \cdot \underbrace{(N-(n-1))}_{\text{nicht an den } n-1 \text{ gewählten Tagen}}$$

und somit

$$\mathbb{P}(A_n^c) = \frac{N \cdot (N-1) \cdot \dots \cdot (N-n+1)}{N^n}$$

Für die Party von oben folgt $\mathbb{P}(A_{30}) = 1 - \mathbb{P}(A_{30}^c) \approx 0,7063162$, die Wette wird also in etwa 71% der Fälle verloren.

2.4.3 Kombinatorische Wahrscheinlichkeiten

Fragestellungen wie im vorangehenden Beispiel 2.21 kann man aus der Sicht von *Urnenmodellen* betrachten. Zur Berechnung der Wahrscheinlichkeit behilft man sich hier der *Kombinatorik*. Bevor wir damit anfangen, benötigen wir erst einmal einige Begriffe aus der Kombinatorik.

Definition 2.22 (Fakultät): Für eine natürliche Zahl n definieren wir n Fakultät als

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1,$$

weiterhin ist $0! = 1$ (als leeres Produkt).

Definition 2.23 (Binomialkoeffizient): Für zwei natürliche Zahlen n_1 und n_2 mit $n_1 > n_2$ definieren wir den *Binomialkoeffizienten* als

$$\binom{n_1}{n_2} := \frac{n_1!}{n_2! \cdot (n_1 - n_2)!}.$$

Damit kommen wir jetzt zu den Urnenmodellen.

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

Definition und Lemma 2.24 (Urnenmodell): In einer Urne liegen N durchnummerierte Kugeln. Es werden n Kugeln zufällig gezogen. Die Ziehung kann mit oder ohne Zurücklegen und mit oder ohne Beachtung der Reihenfolge geschehen.

(i) *Ziehen mit Reihenfolge und mit Zurücklegen:* Damit ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\}\}$$

mit der Mächtigkeit $|\Omega| = N^n$.

(ii) *Ziehen mit Reihenfolge und ohne Zurücklegen:* Hier ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ mit } \omega_i \neq \omega_j \text{ für } i \neq j\}$$

mit der Mächtigkeit $|\Omega| = \frac{N!}{(N-n)!}$.

(iii) *Ziehen ohne Reihenfolge und ohne Zurücklegen:* Der Ergebnisraum ist

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ und } \omega_1 < \omega_2 < \dots < \omega_n\}$$

mit der Mächtigkeit $|\Omega| = \frac{N!}{n!(N-n)!} = \binom{N}{n}$.

(iv) *Ziehen ohne Reihenfolge und mit Zurücklegen:* In diesem Fall ist der Ergebnisraum

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, N\} \text{ und } \omega_1 \leq \omega_2 \leq \dots \leq \omega_n\}$$

mit der Mächtigkeit $|\Omega| = \frac{(N+n-1)!}{n!(N-1)!} = \binom{N+n-1}{n}$.

Beispiel 2.25: (a) Ein Verein hat 22 Mitglieder und plant seine jährliche Weihnachtsfeier. Das Planungsgremium besteht aus 5 Mitgliedern. Der Verein könnte $\binom{22}{5} = 26334$ Jahre mit jeweils anders besetzten Gremien die Feier veranstalten.¹ Wir haben hier den Fall ohne Reihenfolge und ohne Zurücklegen.

(b) Bei einer Tombola machen 87 Personen mit und es gibt 10 Preise zu gewinnen. Für den ersten Preis wird der Name des Gewinners aus einer Urne gezogen, dann wird aus der gleichen Urne der Name des Gewinners des zweiten Preises gezogen usw. bis alle 10 Preise vergeben sind. Wir haben $\frac{87!}{77!}$ mögliche Gewinner-10-tupel. Unter den Teilnehmern ist das Ehepaar Müller. Wie groß ist die Wahrscheinlichkeit, dass beide Müllers unter den Gewinnern sind? Wir haben $|\Omega| = \frac{87!}{77!}$ und betrachten wieder das Gegenereignis. Seien

$$\begin{aligned} A &= \{\text{kein Herr Müller}\} & \text{mit} & \quad |A| = \frac{86!}{76!} \\ B &= \{\text{keine Frau Müller}\} & \text{mit} & \quad |A| = \frac{86!}{76!} \\ C &= \{\text{gar kein Müller}\} & \text{mit} & \quad |A| = \frac{85!}{75!} \end{aligned}$$

Dann ist

$$\begin{aligned} \mathbb{P}(\text{beide Müller}) &= 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(C)) \\ &= 1 - \frac{\frac{86!}{76!} + \frac{86!}{76!} - \frac{85!}{75!}}{\frac{87!}{77!}} \approx 0,012 = 1,2\% \end{aligned}$$

¹Wir müssen hier von sehr langlebigen Mitgliedern ausgehen, außerdem gibt es keine Ein- und Austritte.

(c) Bei dem Beispiel 2.21 wollen wir erst einmal jeder Person einen Geburtstag zuordnen. Wir ziehen also mit Reihenfolge und mit Zurücklegen, da jede Person eindeutig einen Geburtstag bekommt und diese sich auch wiederholen können. Damit haben wir unser Ω . Dann interessieren wir uns für das Ereignis mit Reihenfolge, da wieder jede einzelne Person betrachtet wird, und ohne Zurücklegen, weil wir im mehrfache Geburtstage nicht zulassen (Gegenereignis).

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

Wir wollen uns nicht nur mit den Ereignissen selbst beschäftigen, sondern auch ihre Zusammenhänge beschreiben. Als erstes schauen wir uns den Fall an, dass es keine Verbindung zwischen ihnen gibt.

Wir setzen den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ voraus.

2.5.1 Unabhängige Ereignisse

Wir nennen zwei Ereignisse A und B unabhängig voneinander, wenn das Eintreten von A die Wahrscheinlichkeit des Eintretens von B nicht verändert.

Definition 2.26 (Unabhängigkeit zweier Ereignisse): Zwei Ereignisse $A, B \in \mathcal{F}$ heißen *unabhängig*, wenn gilt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B). \quad (2.2)$$

Wir sehen, dass diese Eigenschaft symmetrisch ist.

Anmerkung: Beim Laplaceschen Wahrscheinlichkeitsraum sehen wir zur Interpretation, dass sich die relative Häufigkeit nicht ändert, wenn wir die Bezugsmenge von Ω zu A ändern

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

das Ereignis B in der Menge A anteilig also genauso oft auftritt, wie in der Grundgesamtheit Ω .

Lemma 2.27: Das sichere und das unmögliche Ereignis sind von allen anderen unabhängig, insbesondere sind sie von sich selbst unabhängig. Umgekehrt gilt für ein Ereignis, welches von sich selbst unabhängig ist, dass seine Wahrscheinlichkeit entweder 0 oder 1 ist.

BEWEIS Sei $A \in \mathcal{F}$. Dann gilt

$$\begin{aligned} \mathbb{P}(A \cap \Omega) &= \mathbb{P}(A) = \mathbb{P}(A) \cdot 1 = \mathbb{P}(A) \cdot \mathbb{P}(\Omega), \\ \mathbb{P}(A \cap \emptyset) &= \mathbb{P}(\emptyset) = 0 = 0 \cdot \mathbb{P}(A) = \mathbb{P}(\emptyset) \cdot \mathbb{P}(A), \end{aligned}$$

wobei A keinerlei Einschränkung unterworfen war, insbesondere gilt die Argumentation für $A = \Omega$ und $A = \emptyset$. Sei nun $B \in \mathcal{F}$ von sich selbst unabhängig, also

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

falls $\mathbb{P}(B) \neq 0$, können wir beide Seiten durch $\mathbb{P}(B)$ teilen

$$1 = \mathbb{P}(B),$$

andernfalls ist gerade $\mathbb{P}(B) = 0$. □

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

Wir können die Definition auch auf Familien von Ereignissen ausweiten.

Definition 2.28: Seien $A_1, \dots, A_n \in \mathcal{F}$. Die Ereignisse A_1, \dots, A_n heißen unabhängig, wenn für jedes $1 \leq k \leq n$ und jede Auswahl $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ gilt

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_k}). \quad (2.3)$$

Anmerkung: (a) Wir müssen die Eigenschaft der Unabhängigkeit für jede Auswahl der Mengen der Familie fordern, denn sonst wäre jede Familie unabhängig, wenn ein $A_i = \emptyset$ für $1 \leq i \leq n$ wäre.

(b) Betrachten wir das Beispiel des 2-maligen Wurfs mit einer fairen Münze, also

$$\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$$

und dabei die Ereignisse

$$A = \{\text{beim ersten Wurf Kopf}\} = \{(0,0), (0,1)\}$$

$$B = \{\text{beim zweiten Wurf Kopf}\} = \{(0,0), (1,0)\}$$

$$C = \{\text{Anzahl der Würfe mit Kopf ist gerade}\} = \{(0,0), (1,1)\}$$

Dann gilt

$$\mathbb{P}(A) = \frac{1}{2} \qquad \mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

$$\mathbb{P}(B) = \frac{1}{2} \qquad \mathbb{P}(A \cap C) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(C)$$

$$\mathbb{P}(C) = \frac{1}{2} \qquad \mathbb{P}(B \cap C) = \frac{1}{4} = \mathbb{P}(B) \cdot \mathbb{P}(C)$$

aber

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

Deswegen ist paarweise Unabhängigkeit nicht ausreichend für die Unabhängigkeit einer Familie von Mengen.

2.5.2 Bedingte Wahrscheinlichkeit

Definition 2.29 (Bedingte Wahrscheinlichkeit): Seien A und B zwei Ereignisse mit $\mathbb{P}(A) > 0$. Dann ist

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad (2.4)$$

die *bedingte Wahrscheinlichkeit* von B gegeben A .

Anmerkung (Rechenregeln für die bedingte Wahrscheinlichkeit): Sei $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$. Dann gilt, dass die Abbildung

$$\begin{aligned} \mathbb{P}(\cdot|A) : \mathcal{F} &\rightarrow [0,1] \\ B &\mapsto \mathbb{P}(B|A) \end{aligned}$$

ein Wahrscheinlichkeitsmaß ist. (Die Axiome von Kolmogorov können entsprechend nachgeprüft werden.)

Lemma 2.30: Sind zwei Ereignisse A und B unabhängig, so ist $\mathbb{P}(B|A) = \mathbb{P}(B)$

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

BEWEIS Da $\mathbb{P}(A) > 0$ ist, gilt

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(B). \quad \square$$

Beispiel 2.31: Wir betrachten zweimaliges Ziehen ohne Zurücklegen aus einer Urne mit N Kugeln, R roten und $N - R$ blauen, und definieren die Ereignisse

$$\begin{aligned} A &= \{\omega = (\omega_1, \omega_2) : \omega_1 = \text{rot}\}, \\ B &= \{\omega = (\omega_1, \omega_2) : \omega_2 = \text{rot}\} \end{aligned}$$

Wir kennen $\mathbb{P}(A) = \frac{R}{N}$ und $\mathbb{P}(A \cap B) = \frac{R(R-1)}{N(N-1)}$ und somit $\mathbb{P}(B|A) = \frac{R-1}{N-1}$.

Es folgen drei wichtige Aussagen über bedingte Wahrscheinlichkeiten.

Satz 2.32 (Multiplikationsregel): Seien A_1, \dots, A_n Ereignisse mit $\mathbb{P}(A_1 \cap \dots \cap A_n) \neq 0$. Dann gilt

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \quad (2.5)$$

BEWEIS Auf jeden der Faktoren auf der rechten Seite wenden wir die Definition der bedingten Wahrscheinlichkeit an und erhalten

$$\begin{aligned} &\mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \cdot \dots \cdot \frac{\mathbb{P}(A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})} \\ &= \mathbb{P}(A_1 \cap \dots \cap A_n), \end{aligned}$$

die rechte Seite. □

Definition 2.33 (Partition): Die Mengenfamilie $B_1, \dots, B_n \subset \Omega$ ist eine Partition von Ω , falls $B_1 \cup \dots \cup B_n = \Omega$ und $B_i \cap B_j = \emptyset$ für alle $1 \leq i, j \leq n$ mit $i \neq j$ gilt.

Satz 2.34 (Satz von der totalen Wahrscheinlichkeit): Es seien $B_1, \dots, B_n \in \mathcal{F}$ eine Partition von Ω und gelte $\mathbb{P}(B_i) > 0$ für alle $0 \leq i \leq n$. Dann gilt

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i) \quad (2.6)$$

für alle $A \in \mathcal{F}$.

BEWEIS Definiere $C_i = A \cap B_i$ für $1 \leq i \leq n$. Dann sind die Mengen C_1, \dots, C_n disjunkt und $A = C_1 \cup \dots \cup C_n$ und es gilt

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(C_i) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i). \quad \square$$

Der Satz von der totalen Wahrscheinlichkeit wird vor allem zu Modellierung mehrstufiger Experimente verwendet.

Beispiel 2.35 (Zweistufiges Experiment): Betrachte 10 Urnen mit jeweils 9 Kugeln. Die erste Urne enthält 9 rote Kugeln, die zweite 8 rote und eine blaue, usw., die zehnte Urne enthält 9 blaue Kugeln. Es wird zufällig eine Urne ausgewählt. Anschließend wird zweimalig aus dieser Urne mit Zurücklegen gezogen. Dann ist

$$\Omega = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 \in \{\text{Urne } 1, \dots, \text{Urne } 10\}, \omega_2, \omega_3 \in \{\text{rot, blau}\}\}$$

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

die betrachtete Grundgesamtheit. Wir definieren die Ereignisse

$$\begin{aligned} B_i &= \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_1 = \text{Urne } i\}, \\ A_1 &= \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_2 = \text{blau}\}, \\ A_2 &= \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_3 = \text{blau}\} \end{aligned}$$

und interessieren uns für $\mathbb{P}(A_1)$ und $\mathbb{P}(A_2|A_1)$. Wir kennen $\mathbb{P}(B_i) = \frac{1}{10}$ und $\mathbb{P}(A_1|B_i) = \frac{i-1}{9}$. Damit berechnen wir

$$\mathbb{P}(A_1) = \sum_{i=1}^{10} \mathbb{P}(A_1|B_i)\mathbb{P}(B_i) = \frac{1}{10} \sum_{i=1}^{10} \frac{i-1}{9} = \frac{1}{90} \sum_{i=0}^9 i = \frac{1}{2}.$$

Weiter gilt

$$\begin{aligned} \mathbb{P}(A_2|A_1) &= \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = 2 \cdot \mathbb{P}(A_1 \cap A_2) = 2 \cdot \sum_{i=1}^{10} \mathbb{P}(A_1 \cap A_2|B_i)\mathbb{P}(B_i) \\ &= \frac{2}{10} \sum_{i=1}^{10} \left(\frac{i-1}{9}\right)^2 = \frac{2}{10 \cdot 9 \cdot 9} \sum_{i=0}^9 i^2 = \frac{2}{10 \cdot 9 \cdot 9} \cdot \frac{9 \cdot 10 \cdot 19}{6} = \frac{19}{27}. \end{aligned}$$

Satz 2.36 (Bayes-Formel): Es seien $B_1, \dots, B_n \in \mathcal{F}$ eine Partition von Ω und es gelte $\mathbb{P}(B_i) > 0$ für alle $1 \leq i \leq n$. Dann gilt

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)} \quad (2.7)$$

für alle $A \in \mathcal{F}$ mit $\mathbb{P}(A) > 0$ und alle $1 \leq k \leq n$.

BEWEIS Mit zweimaliger Anwendung der Definition der bedingten Wahrscheinlichkeit wissen wir, dass

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^n \mathbb{P}(A|B_i)\mathbb{P}(B_i)},$$

wobei die letzte Gleichheit aus dem Satz 2.34 folgt. \square

Beispiel 2.37: Wir betrachten das Modell aus dem Beispiel 2.35 und interessieren uns für $\mathbb{P}(B_k|A_1)$:

$$\mathbb{P}(B_k|A_1) = \frac{\mathbb{P}(A_1|B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{10} \mathbb{P}(A_1|B_i)\mathbb{P}(B_i)} = \frac{\frac{k-1}{9} \frac{1}{10}}{\sum_{i=1}^{10} \frac{i-1}{9} \frac{1}{10}} = \frac{k-1}{\sum_{i=0}^9 i} = \frac{k-1}{45}$$

Beispiel 2.38: Wir betrachten eine Krankheit, unter der 0,1% der Menschen leidet. Es wird für sie ein medizinischer Test entwickelt. Der Test ist für 99% der Erkrankten positiv und auch für 1% der Gesunden. Wir definieren die Ereignisse

$$\begin{aligned} A &= \{\text{Person erkrankt}\} \\ B &= \{\text{Person hat einen positiv ausgefallenen Test}\} \end{aligned}$$

Wir interessieren uns für $\mathbb{P}(A|B)$:

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)} \\ &= \frac{0,99 \cdot 0,001}{0,99 \cdot 0,001 + 0,01 \cdot 0,999} = 0,09016393 \approx 9,0\%, \end{aligned}$$

2.5 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

d. h. von allen Personen, bei denen der Test positiv ausfällt, sind nur 9,0% erkrankt.

Für einen medizinischen Test mit $\mathbb{P}(B|A) = 0,98$ und $\mathbb{P}(B|A^c) = 0,03$ erhalten wir $\mathbb{P}(A|B) = 0,032$, also nur 3,2%.

Beispiel 2.39: Wir betrachten die Zinsentwicklung einer Anlageform innerhalb eines vorgegebenen Zeitraumes und definieren die Ereignisse

$$\begin{aligned}A_1 &= \{\text{Der Zins fällt um } 0,5\%\} \\A_2 &= \{\text{Der Zins bleibt unverändert}\} \\A_3 &= \{\text{Der Zins steigt um } 0,5\%\}\end{aligned}$$

der Einfachheit halber sei $\Omega = A_1 \cup A_2 \cup A_3$ und

$$B = \{\text{Anlageberater sagt, der Zins steige}\}$$

Wir kennen die folgenden Einschätzungen

$$\mathbb{P}(A_1) = 0,1 \qquad \mathbb{P}(A_2) = 0,6 \qquad \mathbb{P}(A_3) = 0,3$$

und die Erfahrungswerte

$$\mathbb{P}(B|A_1) = 0,15 \qquad \mathbb{P}(B|A_2) = 0,3 \qquad \mathbb{P}(B|A_3) = 0,75$$

Wir interessieren uns für $\mathbb{P}(A_1|B)$, $\mathbb{P}(A_2|B)$ und $\mathbb{P}(A_3|B)$.

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) + \mathbb{P}(B|A_3)\mathbb{P}(A_3) \\ &= 0,15 \cdot 0,1 + 0,3 \cdot 0,6 + 0,75 \cdot 0,3 = 0,42\end{aligned}$$

und damit

$$\mathbb{P}(A_1|B) = \frac{0,15 \cdot 0,1}{0,42} \approx 0,036 \qquad \mathbb{P}(A_2|B) \approx 0,429 \qquad \mathbb{P}(A_3|B) \approx 0,536$$

Unsere Unsicherheit bezüglich des Eintretens des uns interessierenden Ereignisses A_3 ist also mit Hilfe der Information B gestiegen. (Wir haben noch keine Aussage, was passiert, wenn der Anlageberater etwas anderes vermutet.)

2 Ergebnisse, Ereignisse und Wahrscheinlichkeit

3 Zufallsvariablen und Verteilungen

Zufallsvariablen sind ein weiteres Instrument bei der Beschäftigung mit Zufallsexperimenten. Sie kommen immer dann zum Einsatz, wenn wir uns nicht für die Einzelheiten der Ergebnisse und Ereignisse interessieren, sondern nur eine Zusammenfassung brauchen.

3.1 Grundbegriffe

Eine Zufallsvariable ordnet jedem Elementarergebnis eine Zahl zu.

Anmerkung: Zur Definition der Zufallsvariable benötigen wir eine Regularitätsbedingung, die *Messbarkeit*. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \rightarrow \mathbb{R}$ heißt bezüglich der σ -Algebra \mathcal{F} *messbar*, wenn

$$\{\omega : X(\omega) \leq y\} \in \mathcal{F} \tag{3.1}$$

für alle $y \in \mathbb{R}$ gilt.

Definition 3.1 (Zufallsvariable): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine *Zufallsvariable* ist eine messbare Funktion $X : \Omega \rightarrow \mathbb{R}$.

Anmerkung: Eine Zufallsvariable heißt zwar Variable, ist aber eine Funktion.

Anmerkung: Oft interessiert man sich nicht für einen speziellen Wert, den X annimmt, sondern ob sie in einen Bereich fällt, also für $X \in B$, wobei $B \subset \mathbb{R}$ ist, z. B. $B = (-\infty, c]$ oder $B = [c_1, d_1] \cup [c_2, d_2]$. Deshalb betrachten wir auch im Bildraum \mathbb{R} von X eine σ -Algebra. Dazu nehmen wir die sogenannte *Borel- σ -Algebra*, die als die kleinste σ -Algebra von Teilmengen von \mathbb{R} , die alle offenen Intervalle von \mathbb{R} enthält, definiert ist, d. h.

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{(a, b) : -\infty < a < b < \infty\}\right)$$

Durch Bildung von Schnitten über unendlich viele Mengen enthält $\mathcal{B}(\mathbb{R})$ auch alle halboffenen und abgeschlossenen Intervalle, denn

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + 1/n) \in \mathcal{B}(\mathbb{R})$$

$$[a, b) = \bigcap_{n=1}^{\infty} (a - 1/n, b) \in \mathcal{B}(\mathbb{R})$$

$$[a, b] = \bigcap_{n=1}^{\infty} (a - 1/n, b + 1/n) \in \mathcal{B}(\mathbb{R})$$

und damit sind auch einelementige Mengen

$$\{c\} = \bigcap_{n=1}^{\infty} (c - 1/n, c + 1/n) \in \mathcal{B}(\mathbb{R})$$

3 Zufallsvariablen und Verteilungen

enthalten. Damit können wir Gleichung (3.1) schreiben als

$$\{\omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{für alle } B \in \mathcal{B}(\mathbb{R}), \quad (3.2)$$

was äquivalent ist.

Notation: Bei der Definition von $\mathcal{B}(\mathbb{R})$ haben wir eine Notation verwendet, die der Erläuterung bedarf. Zu einer Grundmenge Ω und einer Menge von Mengen \mathcal{G} aus Ω , also für alle $G \in \mathcal{G}$ gilt $G \subseteq \Omega$ bezeichnet $\sigma(\mathcal{G})$ eine σ -Algebra, die von \mathcal{G} erzeugt wird. Sie enthält alle Mengen aus \mathcal{G} , d. h. für alle $G \in \mathcal{G}$ gilt $G \in \sigma(\mathcal{G})$ und darüber hinaus enthält sie alle Mengen, die sich durch Anwendung der Komplementbildung und (unendlicher) Vereinigung aus diesen zusammensetzen lassen. Damit erfüllt $\sigma(\mathcal{G})$ die Eigenschaften (A1), (A2) und (A3) aus den Definitionen 2.7 und 2.10 und ist eine σ -Algebra, und zwar, da bei dieser Konstruktion zusätzlich nichts „überflüssiges“ hinzukam, die kleinste σ -Algebra, welche \mathcal{G} enthält.

Eine alternative Interpretation ist, dass $\sigma(\mathcal{G})$ der Schnitt aller σ -Algebren ist, die \mathcal{G} enthalten. Damit fallen auch alle nicht „notwendigen“ Mengen heraus und sie ist die kleinste σ -Algebra, welche \mathcal{G} enthält.

Ein Spezialfall ist die von einer Menge erzeugte σ -Algebra. Für eine Menge $A \subseteq \Omega$ ist $\sigma(A) := \sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}$ und damit die kleinste σ -Algebra, die A enthält.

Notation: Es ist üblich eine Zufallsvariable mit einem großen lateinischen Buchstaben zu bezeichnen. Der Wert $X(\omega)$ einer Zufallsvariablen X wird mit dem entsprechenden Kleinbuchstaben, hier x bezeichnet, und *Realisierung von X* genannt.

Beispiel 3.2 (Indikatorfunktion): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ gegeben und $A \in \mathcal{F}$ ein Ereignis. Die *Indikatorfunktion* ist definiert als

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \notin A. \end{cases} \quad (3.3)$$

und wir betrachten die Zufallsvariable $X(\omega) = \mathbb{1}_A(\omega)$ und haben

$$\{\omega : X(\omega) \in B\} = \begin{cases} A, & \text{falls } 1 \in B \text{ und } 0 \notin B, \\ A^c, & \text{falls } 0 \in B \text{ und } 1 \notin B, \\ \Omega, & \text{falls } 0 \in B \text{ und } 1 \in B, \\ \emptyset, & \text{falls } 0 \notin B \text{ und } 1 \notin B \end{cases}$$

für alle $B \in \mathcal{B}(\mathbb{R})$.

Definition 3.3 (Verteilung): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable.

(i) Dann heißt die Funktion

$$F_X(y) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq y\})$$

für alle $y \in \mathbb{R}$ *Verteilungsfunktion* von X . Es gilt $F_X : \mathbb{R} \rightarrow [0, 1]$.

(ii) Die Funktion

$$\mathbb{P}_X(B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

für alle Mengen B der σ -Algebra $\mathcal{B}(\mathbb{R})$ heißt *Verteilung* von X und es gilt $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$.

Notation: Die folgenden Schreibweisen sind üblich:

$$\begin{aligned} F_X(y) &= \mathbb{P}(X \leq y) \\ \mathbb{P}_X(B) &= \mathbb{P}(X \in B) \end{aligned}$$

Anmerkung: Die Funktion \mathbb{P}_X ist ein Wahrscheinlichkeitsmaß, auch *von X induziertes Wahrscheinlichkeitsmaß* genannt, und wir erhalten den Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.

Anmerkung: Wir haben zuerst eine neue Grundmenge genommen, uns dann eine σ -Algebra definiert und im Anschluss ein Wahrscheinlichkeitsmaß auf demselben mit Hilfe von X gebastelt. Die Abbildung $\mathbb{P} \rightarrow \mathbb{P}_X$ heißt auch *Maßtransport*.

Wir wollen zuerst diskrete Zufallsvariablen untersuchen.

3.2 Diskrete Zufallsvariablen

Dazu benötigen wir die folgende

Definition 3.4 (Diskrete Zufallsvariable): Eine Zufallsvariable heißt *diskret*, wenn es eine endliche oder abzählbar unendliche Teilmenge $C \subset \mathbb{R}$ gibt mit $\mathbb{P}(X \in C) = 1$.

Im Abschnitt 2.4 haben wir diskrete Wahrscheinlichkeitsverteilungen definiert. Eine Zufallsvariable ist genau dann diskret, wenn ihre Verteilung diskret ist. Analog gilt für diskrete Zufallsvariablen, dass sie vollständig durch ihre Wahrscheinlichkeitsfunktion beschrieben werden.

Definition 3.5 (Wahrscheinlichkeitsfunktion): Sei X eine diskrete Zufallsvariable mit Wertebereich $X(\Omega) = \{x_1, x_2, \dots\}$. Dann heißt die Funktion $p : X(\Omega) \rightarrow [0, 1]$ definiert durch

$$p(x_i) := \mathbb{P}(\{\omega : X(\omega) = x_i\}) = \mathbb{P}(X = x_i), \quad (3.4)$$

die *Wahrscheinlichkeitsfunktion* von X , auch *Zähldichte* genannt.

Anmerkung: Gelegentlich wird p zu einer Funktion auf ganz \mathbb{R} ausgedehnt, indem $p(x) = 0$ für $x \in \mathbb{R} \setminus X(\Omega)$ gesetzt wird.

Beispiel 3.6: Wir betrachten wieder die Indikatorfunktion aus dem Beispiel 3.2. Sie ist eine diskrete Zufallsvariable mit dem Wertebereich $\{0, 1\}$. Die zugehörige Wahrscheinlichkeitsfunktion ist gegeben durch $p(1) = \mathbb{P}(A)$ und $p(0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Was uns noch fehlt, ist die tatsächliche Überprüfung der Kolmogorovschen Axiome.

Satz 3.7: Sei X eine diskrete Zufallsvariable mit Wertebereich $X(\Omega) = \{x_1, x_2, \dots\}$. Dann erfüllt \mathbb{P}_X die Axiome von Kolmogorov, i. e.

$$\mathbb{P}_X(\{x_i\}) \geq 0 \quad \text{für alle } x_i \in X(\Omega) \quad (3.5)$$

$$\mathbb{P}_X(B) = \sum_{x_i \in B} \mathbb{P}_X(\{x_i\}) \quad (3.6)$$

$$\mathbb{P}_X(X(\Omega)) = 1 \quad (3.7)$$

BEWEIS Wir wissen, dass

$$\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i) = p(x_i) \geq 0$$

da \mathbb{P} ein Wahrscheinlichkeitsmaß ist. Weiter ist

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(\bigcup_{x_i \in B} \{X = x_i\}\right) = \sum_{x_i \in B} \mathbb{P}(X = x_i) = \sum_{x_i \in B} p(x_i) = \sum_{x_i \in B} \mathbb{P}_X(\{x_i\}),$$

3 Zufallsvariablen und Verteilungen

was die σ -Additivität zeigt. Außerdem gilt

$$1 = \mathbb{P}(X \in X(\Omega)) = \mathbb{P}\left(\bigcup_i \{X = x_i\}\right) = \sum_i \mathbb{P}(X = x_i) = \sum_i p(x_i) = \mathbb{P}_X(X(\Omega)),$$

womit auch die Normiertheit gezeigt wäre. \square

Beispiel 3.8 (Münzwurf): Wir betrachten 4-maliges Werfen einer fairen Münze. Sei X definiert durch $X(\omega) = \omega_1 + \omega_2 + \omega_3 + \omega_4$ die interessierende Zufallsvariable. Sie ist diskret mit dem Wertebereich $\{0, 1, 2, 3, 4\}$. Wir erhalten dann die Wahrscheinlichkeitsfunktion

k	0	1	2	3	4
$p(k)$	1/16	4/16	6/16	4/16	1/16

Damit lässt sich ganz leicht die Wahrscheinlichkeit aller Ereignisse der Form $\{X \in A\}$ berechnen, z. B. ist $\mathbb{P}(X \leq 2) = p(0) + p(1) + p(2) = 11/16$.

3.2.1 Charakterisierung diskreter Zufallsvariablen

Es gibt eine Vielzahl praktisch auftretender Verteilungen, deren „Form“ sich stark unterscheidet. Wir wollen uns im Folgenden Möglichkeiten ansehen, diese „Form“ zu charakterisieren. Dazu führen wir die Begriffe des Erwartungswertes und der Varianz ein und beschreiben deren Eigenschaften.

Der Erwartungswert einer Zufallsvariablen ist der Wert, der sich gemittelt über alle Realisierungen, gewichtet mit der Wahrscheinlichkeit ihres Auftretens, ergibt. Er sollte nicht mit dem wahrscheinlichsten Einzelwert verwechselt werden. Das Konzept stammt aus der Untersuchung von Glücksspielen, wo der Erwartungswert den durchschnittlichen Gewinn oder Verlust bei praktisch unendlicher Wiederholung des Spiels beschreibt.

Definition 3.9 (Erwartungswert): Sei X eine diskrete Zufallsvariable und p ihre Wahrscheinlichkeitsfunktion. Der Erwartungswert von X existiert, falls $\sum_x |x| \cdot p(x) < \infty$. Der *Erwartungswert* ist dann definiert als

$$\mathbb{E}(X) := \sum_{x \in X(\Omega)} x \cdot p(x), \quad (3.8)$$

als Symbol wählt man manchmal μ oder μ_X .

Beispiel 3.10: Betrachten wir die Zufallsvariable aus Beispiel 3.8. Dann ist

$$\mathbb{E}(X) = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = 2$$

der Erwartungswert in dem Spiel, d. h. von vier Würfeln mit einer fairen Münze werde zweimal *Kopf* erwartet. Das entspricht auch dem common sense.

Satz 3.11 (Transformationformel für den Erwartungswert): Sei X eine diskrete Zufallsvariable, p ihre Wahrscheinlichkeitsfunktion, sei weiter $u : X(\Omega) \rightarrow \mathbb{R}$ eine Abbildung, für die $\sum_{x \in X(\Omega)} |u(x)|p(x) < \infty$ gilt. Dann ist

$$\mathbb{E}(u(X)) = \sum_{x \in X(\Omega)} u(x) \cdot p(x) \quad (3.9)$$

der Erwartungswert der transformierten Zufallsvariablen $u(X)$.

BEWEIS Wir definieren $Y := u(X)$ und untersuchen die zugehörige Wahrscheinlichkeitsfunktion

$$p_Y(y) = \mathbb{P}(u(X) = y) = \sum_{x:u(x)=y} p(x),$$

mit der wir den Erwartungswert

$$\mathbb{E}(Y) = \sum_y y \cdot p_Y(y) = \sum_y y \sum_{x:u(x)=y} p(x) = \sum_y \sum_{x:u(x)=y} u(x)p(x) = \sum_x u(x)p(x)$$

berechnen. □

Satz 3.12 (Dreiecksungleichung für den Erwartungswert): Sei X eine diskrete Zufallsvariable, deren Erwartungswert existiert, d. h. $\sum_{x \in X(\Omega)} |x|p(x) < \infty$. Dann gilt

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|). \quad (3.10)$$

BEWEIS Die Transformationsformel aus Satz 3.11 mit $u(x) = |x|$ liefert

$$\mathbb{E}(|X|) = \sum_{x \in X(\Omega)} |x|p(x) = \sum_{x \in X(\Omega)} |xp(x)| \leq \left| \sum_{x \in X(\Omega)} xp(x) \right| = |\mathbb{E}(X)|$$

wobei bei der Ungleichheit die Dreiecksungleichung für Summen reeller Zahlen verwendet wurde. □

Satz 3.13 (Linearität des Erwartungswertes): Seien X und Y zwei Zufallsvariablen, deren Erwartungswerte existieren. Dann gilt für $a, b \in \mathbb{R}$

- (i) $\mathbb{E}(aX) = a\mathbb{E}(X)$,
- (ii) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
- (iii) $\mathbb{E}(b) = b$.

BEWEIS (i) Aus der Transformationsformel haben wir mit der Funktion $u(x) = a \cdot x$

$$\mathbb{E}(aX) = \sum_x (ax)p(x) = a \sum_x xp(x) = a\mathbb{E}(X)$$

(ii) Eine hinreichende Bedingung für die Gleichheit von Summen ist die Gleichheit und das Auftreten aller Summanden.

$$\mathbb{E}(X + Y) = \sum_z z \mathbb{P}(X + Y = z),$$

aus der Summe wir mit dem Satz von der totalen Wahrscheinlichkeit (Satz 2.34) eine Doppelsumme bilden

$$= \sum_z \sum_x z \mathbb{P}(X + Y = z, X = x),$$

die wir umstellen können

$$= \sum_x \sum_z ((z - x) + x) \mathbb{P}(Y = z - x, X = x)$$

und jetzt $y := z - x$ definieren und wegen $x, z \in \mathbb{R}$ gilt auch $y \in \mathbb{R}$, weswegen wir nun nur die Reihenfolge der Summanden ändern

$$\begin{aligned} &= \sum_x \sum_y (y + x) \mathbb{P}(Y = y, X = x) \\ &= \sum_x \sum_y x \mathbb{P}(Y = y, X = x) + \sum_x \sum_y y \mathbb{P}(Y = y, X = x), \\ &= \sum_x x \sum_y \mathbb{P}(Y = y, X = x) + \sum_y y \sum_x \mathbb{P}(Y = y, X = x), \end{aligned}$$

3 Zufallsvariablen und Verteilungen

der Satz von der totalen Wahrscheinlichkeit liefert

$$\begin{aligned} &= \sum_X x \mathbb{P}(X = x) + \sum_y y \mathbb{P}(Y = y, X = x), \\ &= \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

(iii) Die konstante Zufallsvariable $X \equiv b$ hat genau eine Realisierung, und zwar mit Wahrscheinlichkeit 1. Die Wahrscheinlichkeitsfunktion ist dann $p(b) = 1$ und $p(x) = 0$ falls $x \neq b$. \square

Anmerkung: Wir haben in diesem Beweis verwendet, dass $X + Y$ eine Zufallsvariable ist. Allgemeiner formuliert sagen wir, dass $u(X, Y)$ eine Zufallsvariable ist, wenn u messbar ist. Das ist eine Generalisierung des Transformationssatzes für mehrdimensionale Zufallsvariablen, die wir im Abschnitt 3.4 kennenlernen werden. Die Aussage selbst werden hier hin und wieder vorwegnehmen.

Neben der Erwartung benötigen wir noch die Varianz einer Zufallsvariablen.

Definition 3.14 (Varianz): Sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann ist die *Varianz* von X gegeben durch

$$\text{Var}X := \mathbb{E} (X - \mathbb{E}(X))^2. \quad (3.11)$$

Als Symbol wird oft σ^2 oder σ_X^2 verwendet. Die positive Wurzel aus der Varianz nennt man die *Standardabweichung* von X .

Anmerkung: Aus der Bedingung $\mathbb{E}(X^2) < \infty$ folgt, dass $\mathbb{E}(|X|) < \infty$, also die Existenz des Erwartungswertes.

Mit dieser Definition als mittlere quadratische Abweichung ist die Varianz ein Maß für die Streuung.

Satz 3.15: Sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann gilt

- (i) $\text{Var}(aX + b) = a^2 \text{Var}(X)$ für $a, b \in \mathbb{R}$,
- (ii) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

BEWEIS (i) Wegen der Linearität des Erwartungswertes gilt $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ und es folgt

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

(ii) Ebenfalls aus der Linearität des Erwartungswertes folgt

$$\text{Var}(X) = \mathbb{E} (X - \mathbb{E}(X))^2 = \mathbb{E} (X^2 - 2(\mathbb{E}(X))X + (\mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \quad \square$$

Die Varianz und den Erwartungswert in Beziehung setzend sehen wir, dass der Erwartungswert eine Minimumeigenschaft besitzt.

Satz 3.16: Sei X eine Zufallsvariable, für die $\mathbb{E}(X^2) < \infty$ gilt, und sei $a \in \mathbb{R}$. Dann gilt

$$\mathbb{E}(X - a)^2 = \text{Var}(X) + (\mathbb{E}(X) - a)^2 \quad (3.12)$$

und damit

$$\mathbb{E}(X - a)^2 \geq \text{Var}(X). \quad (3.13)$$

BEWEIS Wegen der Linearität des Erwartungswertes ist $\mathbb{E}(X - \mathbb{E}X) = 0$ und damit folgt

$$\begin{aligned} \mathbb{E}(X - a)^2 &= \mathbb{E}(X - \mathbb{E}X + \mathbb{E}X - a)^2 \\ &= \mathbb{E}\left((X - \mathbb{E}X)^2 + 2\mathbb{E}(X - \mathbb{E}X)(\mathbb{E}X - a) + (\mathbb{E}X - a)^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}X)^2\right) + 2\mathbb{E}\left((X - \mathbb{E}X)(\mathbb{E}X - a)\right) + \mathbb{E}\left((\mathbb{E}X - a)^2\right) \\ &= \mathbb{E}\left((X - \mathbb{E}X)^2\right) + 2(\mathbb{E}X - a) \cdot \mathbb{E}(X - \mathbb{E}X) + \mathbb{E}\left((\mathbb{E}X - a)^2\right) \\ &= \mathbb{E}(X - \mathbb{E}X)^2 + (\mathbb{E}X - a)^2. \end{aligned} \quad \square$$

3.2.2 Spezielle/Wichtige diskrete Verteilungen

Im Folgenden wollen wir uns einige in der Praxis häufig vorkommende Verteilungen ansehen.

Notation: Ist eine Zufallsvariable X einer Verteilung $F(x) := \mathbb{P}(X \leq x)$ entsprechend verteilt, so sagt man $X \sim F$.

Bernoulli-Verteilung

Sei $A \in \Omega$ mit $\mathbb{P}(A) = \eta$ mit $\eta \in (0, 1)$ gegeben. Sei X eine Zufallsvariable, definiert durch

$$X(\omega) = \mathbb{1}_A(\omega)$$

wobei wir das Eintreten von A mit *Erfolg*, dasjenige von A^c mit *Misserfolg* identifizieren. Der Wertebereich $X(\Omega) = \{0, 1\}$ ¹ und X ist diskret mit der Wahrscheinlichkeitsfunktion

$$p(k) = \begin{cases} \eta & \text{für } k = 1, \\ 1 - \eta & \text{für } k = 0. \end{cases}$$

Die Verteilung heißt dann *Bernoulli-Verteilung* mit Parameter η .

Als Erwartungswert erhalten wir

$$\mathbb{E}X = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = \eta$$

und als Varianz

$$\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = (1^2 \cdot \mathbb{P}(X = 1) + 0^2 \cdot \mathbb{P}(X = 0)) - \eta^2 = \eta(1 - \eta).$$

Binomial-Verteilung

Seien X_i Zufallsvariablen für $i = 1, \dots, n$ Bernoulli-verteilt mit Parameter η , die nicht zusammenhängen.² Dann definiert $S_n := X_1 + \dots + X_n$ eine Zufallsvariable, die die Anzahl der Erfolge in n Experimenten

¹Streng genommen haben wir, dass $\mathbb{P}(X \in \{0, 1\}) = 1$ gilt. Man sagt dann, dass $X(\Omega) = \{0, 1\}$ *fast sicher* gilt. Damit tragen wir der Tatsache Rechnung, dass wir über Ereignisse, deren Wahrscheinlichkeit 0 ist, keine Aussage treffen können.

²Wir werden noch sehen, dass man die Eigenschaft der Unabhängigkeit, die wir von Ereignissen kennen, auch auf Zufallsvariablen übertragen kann.

3 Zufallsvariablen und Verteilungen

angibt. Die Wahrscheinlichkeitsfunktion ist

$$p(k) = \mathbb{P}(S_n = k) = \binom{n}{k} \eta^k (1 - \eta)^{n-k} \quad \text{für } k = 0, \dots, n,$$

was aus der Überlegung folgt, dass wir, um genau k Erfolge zu erhalten, k Erfolge und $n - k$ Misserfolge benötigen, wobei deren Reihenfolge unwichtig ist.

Notation: Die Binomialverteilung mit den Parametern n und η wird mit $\text{Bin}(n, \eta)$ abgekürzt.

Bevor wir den Erwartungswert ausrechnen, müssen wir uns noch über folgende Identität Gedanken machen.

Lemma 3.17: Für $k, n \in \mathbb{N}$ mit $n \geq k$ gilt

$$k \binom{n}{k} = n \binom{n-1}{k-1}. \quad (3.14)$$

Der Erwartungswert ist dann

$$\mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i = \sum_{i=1}^n \eta = n\eta$$

und weiter

$$\begin{aligned} \mathbb{E}(S_n(S_n - 1)) &= \sum_{k=0}^n k(k-1) \binom{n}{k} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n k(k-1) \binom{n}{k} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n n(k-1) \binom{n-1}{k-1} \eta^k (1-\eta)^{n-k} \\ &= \sum_{k=2}^n n(n-1) \binom{n-2}{k-2} \eta^k (1-\eta)^{n-k} \\ &= n(n-1) \cdot \eta^2 \sum_{k=2}^n \binom{n-2}{k-2} \eta^{k-2} (1-\eta)^{(n-2)-(k-2)} \\ &= n(n-1) \cdot \eta^2 \underbrace{\sum_{j=0}^{n-2} \binom{n-2}{j} \eta^j (1-\eta)^{(n-2)-j}}_{=1, \text{ da } \text{Bin}(n-2, \eta)} \\ &= n(n-1) \cdot \eta^2 \end{aligned}$$

und damit die Varianz

$$\begin{aligned} \text{Var}(S_n) &= \mathbb{E}(S_n^2) - (\mathbb{E}S_n)^2 \\ &= \mathbb{E}(S_n(S_n - 1)) + \mathbb{E}S_n - (\mathbb{E}S_n)^2 \\ &= n(n-1)\eta^2 + n\eta - (n\eta)^2 = n\eta - n\eta^2 = n\eta(1-\eta) \end{aligned}$$

Beispiel 3.18: Im Beispiel 3.8 hatten wir es mit einer Binomialverteilung mit $n = 4$ und $\eta = 0,5$ zu tun. Damit ist $\mathbb{E}S_4 = 4 \cdot 0,5 = 2$ (was wir schon hatten) und $\text{Var}S_4 = 4 \cdot 0,5 \cdot 0,5 = 1$.

Satz 3.19 (Symmetrieeigenschaft): Sei $X \sim \text{Bin}(n, \eta)$ und sei $Y := n - X$. Dann gilt

$$Y \sim \text{Bin}(n, 1 - \eta). \quad (3.15)$$

BEWEIS Seien p_X und p_Y die Wahrscheinlichkeitsfunktion von X resp. Y . Dann ist

$$\begin{aligned} p_Y(k) &= \mathbb{P}(Y = k) = \mathbb{P}(n - X = k) = \mathbb{P}(X = n - k) = p_X(n - k) \\ &= \binom{n}{n - k} \eta^{n - k} (1 - \eta)^{n - (n - k)} \\ &= \binom{n}{k} (1 - \eta)^k (1 - (1 - \eta))^{n - k} \end{aligned}$$

was die Wahrscheinlichkeitsfunktion von $\text{Bin}(n, 1 - \eta)$ ist. \square

Die Parameter der Binomialverteilung lassen sich aus dem Erwartungswert und der Varianz bestimmen. Definiere $\mu := \mathbb{E}S_n$ und $\sigma^2 := \text{Var}S_n$. Dann gilt

$$\begin{aligned} n &= \frac{\mu^2}{\mu - \sigma^2}, \\ \eta &= 1 - \frac{\sigma^2}{\mu}. \end{aligned}$$

Hypergeometrische Verteilung

Wir betrachten eine Urne mit N Kugeln, von denen R rot und $N - R$ weiß sind. Daraus ziehen wir ohne Zurücklegen und ohne Beachtung der Reihenfolge n Kugeln. Wir bezeichnen mit X die Anzahl der gezogenen roten Kugeln. Dann ist $X(\Omega) = \{\max\{n - (N - R), 0\}, \dots, \min\{n, R\}\}$ und die Wahrscheinlichkeitsfunktion von X ist

$$p(r) = \mathbb{P}(X = r) = \frac{\binom{R}{r} \binom{N - R}{n - r}}{\binom{N}{n}} \quad \text{für } r = 0, \dots, n,$$

wenn wir beachten, dass für $l < 0$ und $l > k$ der Binomialkoeffizient definitionsgemäss $\binom{k}{l} = 0$ ist. Der Nenner ist klar, da er sich aus dem Urnenmodell ohne Zurücklegen und ohne Reihenfolge ergibt. Für den Zähler teilen wir die Urne im Geiste in die roten und in die weißen Kugeln. Aus der rotkugeligen Urne mit R Kugeln ziehen wir r Kugeln, aus der weißkugeligen Urne mit $N - R$ Kugeln ziehen wir $n - r$ Kugeln. Damit stellen wir sicher, dass genau r der n gezogenen Kugeln rot sind.

Diese Verteilung heißt hypergeometrische Verteilung mit den Parametern N, R und n und wir sagen $X \sim \text{Hyper}(N, R, n)$.

Für den Erwartungswert und die Varianz benötigen wir noch die folgende Identität:

Lemma 3.20: Für die natürlichen Zahlen k, l, m gilt

$$\binom{l + m}{k} = \sum_{i=0}^k \binom{l}{i} \binom{m}{k - i}.$$

3 Zufallsvariablen und Verteilungen

Der Erwartungswert ist

$$\begin{aligned}\mathbb{E}X &= \sum_{r=0}^n r \frac{\binom{R}{r} \binom{N-R}{n-r}}{\binom{N}{n}} \\ &= \frac{1}{\binom{N}{n}} \sum_{r=0}^n R \binom{R-1}{r-1} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \sum_{r=0}^n \binom{R-1}{r-1} \binom{N-R}{(n-1)-(r-1)} \\ &= \frac{R}{\binom{N}{n}} \binom{N-1}{n-1} = \frac{R}{N} n.\end{aligned}$$

Dabei entspricht $\frac{R}{N}$ dem Anteil der roten Kugeln in der Urne vor dem Ziehen, die Wahrscheinlichkeit, als erste Kugel eine rote Kugel zu ziehen ist $\frac{R}{N}$ und damit ist der Erwartungswert analog dem einer Binomial-verteilten Zufallsvariable $Y \sim \text{Bin}(n, \frac{R}{N})$.

Zur Berechnung der Varianz betrachten wir

$$\begin{aligned}\mathbb{E}X^2 &= \frac{1}{\binom{N}{n}} \sum_{r=0}^n r^2 \binom{R}{r} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \sum_{r=0}^n r \binom{R-1}{r-1} \binom{N-R}{n-r} \\ &= \frac{R}{\binom{N}{n}} \left(\sum_{r=0}^n (r-1) \binom{R-1}{r-1} \binom{N-R}{n-r} + \sum_{r=0}^n \binom{R-1}{r-1} \binom{N-R}{n-r} \right) \\ &= \frac{R(R-1)}{\binom{N}{n}} \sum_{k=0}^n \binom{R-2}{r-2} \binom{N-R}{(n-2)-(r-2)} + \frac{R}{N} n \\ &= \frac{R(R-1)}{\binom{N}{n}} \binom{N-2}{n-2} + \frac{R}{N} n = \frac{R(R-1)}{N(N-1)} n(n-1) + \frac{R}{N} n\end{aligned}$$

und somit ist

$$\text{Var}X = \frac{R}{N} n \left(\frac{R-1}{N-1} (n-1) + 1 - \frac{R}{N} n \right) = \frac{R}{N} n \left(1 - \frac{R}{N} \right) \frac{N-n}{N-1}.$$

Auch hier sehen wir eine Ähnlichkeit zur Binomialverteilung. Wir haben $\text{Var}X = \text{Var}Y \cdot \frac{N-n}{N-1}$, d. h. für $n \geq 2$ wird die Varianz kleiner. Da wir ohne Zurücklegen ziehen, wissen wir schon etwas mehr über den neuen Urneninhalt und „nutzen diese Information“.

Insbesondere folgt daraus, dass die Verteilung einer Zufallsvariablen $Z \sim \text{Hyper}(N, [N\eta], n)$ sich für $n \rightarrow \infty$ einer $\text{Bin}(n, \eta)$ -verteilten Zufallsvariablen annähert. (Dabei ist $[x]$ die Gauss-Klammer und bezeichnet den ganzzahligen Anteil von x .)

Geometrische Verteilung

Seien X_i für $i = 1, 2, \dots$ eine unendliche Folge Bernoulli-verteilter Zufallsvariablen mit Parameter η und definieren

$$T := \min\{j : X_j = 1\} - 1$$

die Anzahl der Misserfolge, die vor dem ersten Erfolg auftreten. Dann ist $T(\Omega) = \{0, 1, 2, \dots\}$ und T ist eine diskrete Zufallsvariable. Die Wahrscheinlichkeitsfunktion von T ist

$$p(k) = \mathbb{P}(T = k) = (1 - \eta)^k \eta \quad \text{für } k = 0, 1, 2, \dots$$

Die Verteilung heißt geometrische Verteilung mit Parameter η .

Bevor wir den Erwartungswert und die Varianz ausrechnen, benötigen wir die geometrische Reihe.

Lemma 3.21 (Geometrische Reihe): Für $0 \leq x < 1$ gilt

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad (3.16)$$

die Ableitung ist

$$\frac{d}{dx} \sum_{k=0}^{\infty} x^k = \sum_{k=0}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}, \quad (3.17)$$

und die zweite Ableitung ist

$$\frac{d^2}{dx^2} \sum_{k=0}^{\infty} x^k = \sum_{k=0}^{\infty} k(k-1) x^{k-2} = \frac{2}{(1-x)^3}. \quad (3.18)$$

Außerdem definieren wir uns eine Hilfszufallsvariable

$$Y := \min\{j : X_j = 1\} = T + 1,$$

über die wir den Erwartungswert und die Varianz berechnen. Die Wahrscheinlichkeitsfunktion von Y ist $p_Y(k) = (1 - \eta)^{k-1} \eta$. Man kann Y interpretieren als die Anzahl der Versuche bis zum ersten Erfolg. Es gilt

$$\begin{aligned} \mathbb{E}Y &= \sum_{k=1}^{\infty} k \eta (1 - \eta)^{k-1} \\ &= \eta \sum_{k=0}^{\infty} k (1 - \eta)^{k-1} \\ &= \eta \cdot \frac{1}{(1 - (1 - \eta))^2} = \frac{1}{\eta} \end{aligned}$$

und

$$\begin{aligned} \mathbb{E}Y^2 &= \sum_{k=1}^{\infty} k^2 \eta (1 - \eta)^{k-1} \\ &= \eta \sum_{k=0}^{\infty} k(k+1) (1 - \eta)^{k-1} - \eta \sum_{k=0}^{\infty} k (1 - \eta)^{k-1} \\ &= \eta \cdot \frac{2}{(1 - (1 - \eta))^3} - \frac{1}{\eta} = \frac{2}{\eta^2} - \frac{1}{\eta} \end{aligned}$$

3 Zufallsvariablen und Verteilungen

und damit

$$\begin{aligned}\text{Var}Y &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \\ &= \frac{2 - \eta - 1}{\eta^2} = \frac{1}{\eta} \left(\frac{1}{\eta} - 1 \right).\end{aligned}$$

Da $T = Y - 1$ ist, gilt

$$\mathbb{E}T = \mathbb{E}(Y - 1) \stackrel{\text{Satz 3.13}}{=} \mathbb{E}Y - 1 = \frac{1 - \eta}{\eta} \quad (3.19)$$

$$\text{Var}T = \text{Var}(Y - 1) \stackrel{\text{Satz 3.15}}{=} \text{Var}Y = \frac{1}{\eta} \left(\frac{1}{\eta} - 1 \right) \quad (3.20)$$

Außerdem lässt sich die Wahrscheinlichkeit, mindestens k Wiederholungen warten zu müssen, einfach angeben

$$\mathbb{P}(T \geq k) = \sum_{j=k}^{\infty} (1 - \eta)^j \eta \stackrel{i:=j-k}{=} (1 - \eta)^k \eta \sum_{i=0}^{\infty} (1 - \eta)^i = (1 - \eta)^k \eta \frac{1}{1 - (1 - \eta)} = (1 - \eta)^k \quad (3.21)$$

Anmerkung: Die geometrische Verteilung wird angewendet, wenn es um Lebensdauer geht. Dabei wird eine Grundmenge an baugleichen Produkten oder eine Population modelliert, die mit der Wahrscheinlichkeit η ausfallen resp. sterben.

Anmerkung: Je nach Literatur und Anwendung wird die Verteilung von T oder auch manchmal von Y geometrische Verteilung genannt.

Negative Binomial-Verteilung

Siehe Übungsaufgabe 1 auf Blatt 4.

Poisson-Verteilung

Wir betrachten eine Zufallsvariable X , die die Anzahl der Erfolge angibt, die innerhalb einer Zeiteinheit eintreten. Die Wahrscheinlichkeitsfunktion ist

$$p(k) = \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{für } k = 0, 1, 2, \dots$$

und sagen, dass X Poisson-verteilt ist mit Parameter λ bezeichnen die Verteilung mit $\text{Poisson}(\lambda)$.

Auch hier benötigen wir eine Reihe, bevor wir den Erwartungswert ausrechnen können.

Lemma 3.22: Für ein $x \in \mathbb{R}$ ist

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

die *Exponentialreihe* mit dem Exponenten x .

Dann ist der Erwartungswert

$$\mathbb{E}X = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

und

$$\mathbb{E}X^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 + \lambda$$

und damit die Varianz

$$\mathbb{V}\text{ar}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda.$$

3.2.3 Unabhängige diskrete Zufallsvariablen

Wir wollen das Konzept der Stochastischen Unabhängigkeit aus Abschnitt 2.5 auf Zufallsvariablen ausweiten.

Definition 3.23 (Unabhängige Zufallsvariablen): Seien X_1, \dots, X_n diskrete Zufallsvariablen. Dann heißen X_1, \dots, X_n (stochastisch) unabhängig, falls

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i) \quad (3.22)$$

für alle $B_1, \dots, B_n \subset \mathbb{R}$ gilt.

Lemma 3.24: Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn die Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$ unabhängig sind für alle $B_1, \dots, B_n \subset \mathbb{R}$.

BEWEIS Die Aussage der Rückrichtung folgt direkt aus der Unabhängigkeit der Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$.

Umgekehrt müssen wir zeigen, dass für alle Indexkombinationen $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ die Ereignisse $\{X_{i_1} \in B_{i_1}\}, \dots, \{X_{i_k} \in B_{i_k}\}$ unabhängig sind.

Setze dafür $B_j = \mathbb{R}$ für alle $j \in \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Dann gilt

$$\mathbb{P}(X_{i_1} \in B_{i_1}, \dots, X_{i_k} \in B_{i_k}) = \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i)$$

und damit sind die Ereignisse $\{X_{i_1} \in B_{i_1}\}, \dots, \{X_{i_k} \in B_{i_k}\}$ zur Indexmenge $\{i_1, \dots, i_k\}$ unabhängig. Da dies für jede Wahl der Indexmenge gilt, gilt es auch für den vollständigen Satz der Ereignisse $\{X_1 \in B_1\}, \dots, \{X_n \in B_n\}$. \square

Satz 3.25: Für zwei unabhängige Zufallsvariablen X und Y gilt

$$\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y, \quad (3.23)$$

falls $\mathbb{E}X$ und $\mathbb{E}Y$ existieren.

3 Zufallsvariablen und Verteilungen

BEWEIS Diese Aussage zeigen wir durch Nachrechnen:

$$\begin{aligned}
 \mathbb{E}(XY) &= \sum_z z \mathbb{P}(XY = z) = \sum_z \sum_x z \mathbb{P}(XY = z, X = x) \\
 &\stackrel{x \neq 0}{=} \sum_x \sum_y \frac{z}{x} x \mathbb{P}\left(Y = \frac{z}{x}, X = x\right) \\
 &\stackrel{y = \frac{z}{x}}{=} \sum_x \sum_y y x \mathbb{P}(Y = y, X = x) \\
 &= \sum_x \sum_y x y \mathbb{P}(X = x) \mathbb{P}(Y = y) \\
 &= \sum_x \left(x \mathbb{P}(X = x) \sum_y y \mathbb{P}(Y = y) \right) \\
 &= \left(\sum_x x \mathbb{P}(X = x) \right) \cdot \left(\sum_y y \mathbb{P}(Y = y) \right) \\
 &= \mathbb{E}X \cdot \mathbb{E}Y
 \end{aligned}$$

□

Satz 3.26: Für unabhängige Zufallsvariablen X_1, \dots, X_n gilt

$$\mathbb{V}\text{ar}(X_1 + \dots + X_n) = \mathbb{V}\text{ar}X_1 + \dots + \mathbb{V}\text{ar}X_n, \quad (3.24)$$

falls die Varianzen existieren.

BEWEIS Zuerst sehen wir uns die gemischten Terme an. Für $1 \leq i < j \leq n$ gilt

$$\begin{aligned}
 \mathbb{E}\left((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right) &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i \cdot \mathbb{E}X_j) - \mathbb{E}(\mathbb{E}X_i \cdot X_j) + \mathbb{E}(\mathbb{E}X_i \mathbb{E}X_j) \\
 &= \mathbb{E}(X_i X_j) - \mathbb{E}X_i \mathbb{E}X_j - \mathbb{E}X_i \mathbb{E}X_j + \mathbb{E}X_i \mathbb{E}X_j \stackrel{\text{Unabhängigkeit}}{=} 0
 \end{aligned} \quad (3.25)$$

und damit folgt

$$\begin{aligned}
 \mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i - \mathbb{E}\left(\sum_{i=1}^n X_i\right)\right)^2\right) \\
 &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}X_i\right)^2\right) \\
 &= \mathbb{E}\left(\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right)^2\right) \\
 &= \mathbb{E}\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i)^2 + 2 \sum_{1 \leq i < j \leq n} (X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right) \\
 &= \sum_{i=1}^n \mathbb{E}(X_i - \mathbb{E}X_i)^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}\left((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right) \\
 &= \sum_{i=1}^n \mathbb{V}\text{ar}X_i.
 \end{aligned}$$

□

Satz 3.27 (Faltungsformel für Wahrscheinlichkeitsfunktionen): Seien X und Y unabhängige, diskrete Zufallsvariablen mit den Wahrscheinlichkeitsfunktionen p_X resp. p_Y . Dann hat ihre Summe $Z = X + Y$ die Wahrscheinlichkeitsfunktion

$$p_Z(z) = \sum_x p_X(x)p_Y(z-x) = \sum_y p_X(z-y)p_Y(y). \quad (3.26)$$

BEWEIS Das Ereignis $\{X + Y = z\}$ schreiben wir unter Verwendung von x und y mit $x + y = z$ als disjunkte Vereinigung der Ereignisse $\{X = x\} \cap \{Y = y\} = \{X = x, Y = y\}$. Dann gilt $y = z - x$ und es folgt mit dem Satz von der totalen Wahrscheinlichkeit und der Definition der Unabhängigkeit, dass

$$\begin{aligned} \mathbb{P}(X + Y = z) &= \sum_x \sum_y \mathbb{P}(X = x, Y = y) = \sum_x \mathbb{P}(X = x, Y = z - x) \\ &= \sum_x \mathbb{P}(X = x)\mathbb{P}(Y = z - x) = \sum_x p_X(x)p_Y(z - x) \end{aligned}$$

Analog erhalten wir die zweite Identität. □

Notation: Für die Faltung zweier Wahrscheinlichkeitsfunktionen schreiben wir $p_X * p_Y$.

Beispiel 3.28 (Binomialverteilung): (a) Seien X und Y unabhängige Zufallsvariablen mit $\text{Bin}(n, \eta)$ bzw. $\text{Bin}(m, \eta)$ -Verteilung. Dann lässt sich die Wahrscheinlichkeitsfunktion von $Z = X + Y$ berechnen

$$\begin{aligned} p_Z(k) &= \sum_{i=0}^k \binom{n}{i} \eta^i (1-\eta)^{n-i} \cdot \binom{m}{k-i} \eta^{k-i} (1-\eta)^{m-(k-i)} \\ &= \eta^k (1-\eta)^{n+m-k} \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i} \stackrel{\text{Lemma 3.20}}{=} \eta^k (1-\eta)^{n+m-k} \binom{n+m}{k} \end{aligned}$$

was die Wahrscheinlichkeitsfunktion der $\text{Bin}(n+m, \eta)$ -Verteilung ist.

(b) Betrachte den fünffachen und den siebenfachen Münzwurf, die unabhängig voneinander stattfinden. Sei X die Anzahl der Erfolge beim fünffachen Wurf, also $X \sim \text{Bin}(5, 1/2)$ und sei Y die Anzahl der Erfolge beim siebenfachen Wurf, also $Y \sim \text{Bin}(7, 1/2)$. Dann gilt für $Z = X + Y$, dass $Z \sim \text{Bin}(12, 1/2)$ und Z beschreibt die Anzahl der Erfolge beim zwölffachen Münzwurf.

Definition 3.29 (Kovarianz und Korrelationskoeffizient): Für zwei Zufallsvariablen X und Y definieren wir die *Kovarianz* durch

$$\mathbb{C}_{\text{ov}}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) \quad (3.27)$$

und den *Korrelationskoeffizienten* durch

$$\rho_{X,Y} = \frac{\mathbb{C}_{\text{ov}}(X, Y)}{\sqrt{\text{Var}X} \sqrt{\text{Var}Y}}. \quad (3.28)$$

Die Zufallsvariablen heißen unkorreliert, falls $\rho_{X,Y} = 0$ ist.

Satz 3.30: Zwei unabhängige Zufallsvariablen X und Y sind unkorreliert.

BEWEIS Sind X und Y unabhängig, so gilt $\mathbb{C}_{\text{ov}}(X, Y) \stackrel{(3.25)}{=} 0$ und deswegen ist $\rho_{X,Y} = 0$. □

Das Gegenteil gilt nicht.

3 Zufallsvariablen und Verteilungen

Beispiel 3.31 (Gegenbeispiel zur Umkehrung): Betrachte die Zufallsvariable X mit $X(\Omega) = \{-1, 0, 1\}$ und der Wahrscheinlichkeitsfunktion $p_X(-1) = 1/4, p_X(0) = 1/2, p_X(1) = 1/4$. Sei weiter $Y := X^2$. Dann gilt $Y(\Omega) = \{0, 1\}$ und Y besitzt die Wahrscheinlichkeitsfunktion $p_Y(0) = 1/2, p_Y(1) = 1/2$. Dann gilt $\mathbb{E}X = 1/4 \cdot (-1) + 1/2 \cdot 0 + 1/4 \cdot 1 = 0$ und $\mathbb{E}Y = 1/2 \cdot 0 + 1/2 \cdot 1 = 1/2$ und die Kovarianz ist

$$\begin{aligned}\mathbb{C}_{\text{ov}}(X, Y) &= \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \mathbb{E}(X(Y - 1/2)) = \mathbb{E}(XY) - 1/2\mathbb{E}X \\ &= \mathbb{E}(X^3) = 1/4 \cdot (-1)^3 + 1/2 \cdot 0^3 + 1/4 \cdot 1^3 = 0,\end{aligned}$$

und damit ist auch $\rho_{X,Y} = 0$, also sind X und Y unkorreliert. Sie sind aber nicht unabhängig, da z. B. $\mathbb{P}(X = 0, Y = 1) = \mathbb{P}(X = 0, X^2 = 1) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1)$.

Satz 3.32: Seien X und Y Zufallsvariablen. Dann gilt

- (a) $\mathbb{C}_{\text{ov}}(X, X) = \text{Var}X$,
- (b) $\mathbb{C}_{\text{ov}}(X, Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$,
- (c) $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2\mathbb{C}_{\text{ov}}(X, Y)$.

BEWEIS Teil (i) folgt direkt aus den Definitionen 3.14 und 3.29, (ii) erhält man analog zum Satz 3.15 (ii) aus der Linearität des Erwartungswertes und (iii) folgt aus dem Beweis zum Satz 3.26. \square

3.3 Stetige Zufallsvariablen

Für viele Zufallsexperimente, die wir modellieren wollen, benötigen wir ein Kontinuum an möglichen Werten für die verwendete Zufallsvariable. Beispielsweise sind dies die Lebensdauer einer Person, eine Zahl aus dem Intervall $[0, 1]$ oder der zufällige Zeitpunkt an dem eine Option auf ein Wertpapier ausgeübt wird usw. Der Wertebereich einer (absolut) stetigen³ Zufallsvariablen ist überabzählbar.

Definition 3.33: Eine integrierbare, nicht-negative Funktion f heißt *Wahrscheinlichkeitsdichte*, *Dichtefunktion* oder *Dichte* der Zufallsvariablen X , falls für alle $a, b \in \mathbb{R}$ mit $a < b$ gilt

$$\mathbb{P}(a < X \leq b) = \mathbb{P}_X((a, b]) = \int_a^b f(x)dx. \quad (3.29)$$

Verteilungen mit einer Dichtefunktion heißen (absolut) stetige Verteilungen. Die Verteilungsfunktion F_X besitzt die Darstellung

$$F_X(y) = \int_{-\infty}^y f(x)dx. \quad (3.30)$$

Anmerkung: (i) Für eine Dichtefunktion f gilt immer $\int f(x)dx = 1$. Umgekehrt wird von einer Funktion f mit dieser Eigenschaft eine Wahrscheinlichkeitsverteilung auf \mathbb{R} definiert. Daher ist jede Funktion f mit $\int f(x)dx = 1$ eine Wahrscheinlichkeitsdichte.

(ii) Zwei Verteilungen \mathbb{P}_X und \mathbb{P}_Y sind gleich, wenn ihre Dichtefunktionen bis auf endlich oder abzählbar unendlich viele Werte übereinstimmen.

³Im allgemeinen Sprachgebrauch wird der Begriff *stetig* verwendet, mathematisch spricht man in dem Fall von *absolut stetig*. Trotz unterschiedlicher Bezeichnungen wird in beiden Kontexten eine Zufallsvariable betrachtet, für deren Verteilung eine Dichte (siehe Definition 3.33) existiert.

Beispiel 3.34: Sei $f_c(x) = cx^3 \cdot \mathbb{1}_{[0,1]}(x)$ mit $c \in \mathbb{R}$ eine Funktionenschar. Für welche Werte von c ist f_c eine Dichtefunktion?

Wir halten fest, dass $f_c(x) \geq 0$ für alle $x \in \mathbb{R}$ und alle $c \in \mathbb{R}$ gilt. Als nächstes bestimmen wir das Integral

$$\int cx^3 \cdot \mathbb{1}_{[0,1]}(x) dx = \int_0^1 cx^3 dx = c \cdot \frac{1}{4} x^4 \Big|_0^1 = \frac{c}{4},$$

d. h. für $c = 4$ haben wir eine Dichte. Sei X eine Zufallsvariable mit der Verteilung mit Dichte f_4 . Dann gilt

$$\mathbb{P}(1/4 < X \leq 1/2) = \int_{1/4}^{1/2} 4x^3 \mathbb{1}_{[0,1]}(x) dx = \int_{1/4}^{1/2} 4x^3 dx = x^4 \Big|_{1/4}^{1/2} = \frac{1}{16} - \frac{1}{256} = \frac{15}{256}$$

und

$$\begin{aligned} \mathbb{P}(X \leq a) &= \int_{-\infty}^a 4x^3 \mathbb{1}_{[0,1]}(x) dx = \int_0^{\min\{\max\{0,a\},1\}} 4x^3 dx = x^4 \Big|_0^{\min\{\max\{0,a\},1\}} \\ &= \begin{cases} 0 & \text{falls } 0 < a, \\ a^4 & \text{falls } 0 \leq a \leq 1, \\ 1 & \text{falls } 1 < a. \end{cases} \end{aligned}$$

Satz 3.35 (Eigenschaften der Verteilungsfunktion einer (absolut) stetigen Zufallsvariablen): Sei X eine (absolut) stetige Zufallsvariable und sei F ihre die Verteilungsfunktion. Dann gilt

- (i) F ist stetig und monoton wachsend mit Werten im Intervall $[0, 1]$.
- (ii) Für die Grenzen gilt

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

- (iii) Für jeden Punkt $x \in \mathbb{R}$ gilt $\mathbb{P}(X = x) = 0$.
- (iv) Für Intervalle erhalten wir

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

und $\mathbb{P}(X \geq a) = 1 - F(a)$.

3.3.1 Charakterisierung stetiger Zufallsvariablen

Wir gehen hier analog zum Abschnitt 3.2.1 für diskrete Zufallsvariablen vor.

Definition 3.36 (Erwartungswert): Sei X eine stetige Zufallsvariable mit Dichte f . Der Erwartungswert existiert, falls $\int |x|f(x)dx < \infty$ und ist dann definiert durch

$$\mathbb{E}X = \int xf(x)dx. \tag{3.31}$$

3 Zufallsvariablen und Verteilungen

Satz 3.37 (Transformationsformel für den Erwartungswert): Sei X eine stetige Zufallsvariable mit Dichte f und sei $u : \mathbb{R} \rightarrow \mathbb{R}$ eine messbare Abbildung. Dann gilt

$$\mathbb{E}(u(X)) = \int u(x)f(x)dx \quad (3.32)$$

falls $\int |u(x)|f(x)dx < \infty$ ist.

Anmerkung (Analogie von Eigenschaften stetiger Zufallsvariablen): Die anderen Sätze zum Erwartungswert aus dem Abschnitt 3.2.1 wie die Dreiecksungleichung (Satz 3.12) und die Linearität (Satz 3.13) gelten analog.

Für die Definition der Varianz (Definition 3.14) als auch ihre Eigenschaften (Satz 3.15) haben wir nur die Definition des Erwartungswertes verwendet und können sie deswegen problemlos übernehmen.

Die Faltung stetiger Zufallsvariablen verhält sich ganz analog zu der diskreter. Wie beim Erwartungswert wird auch hier die Summe durch ein Integral, die Wahrscheinlichkeitsfunktion durch die Dichte ersetzt.

Das Konzept der Unabhängigkeit (Definition 3.23) lässt sich ebenfalls problemlos auf stetige Zufallsvariablen übertragen.

3.3.2 Spezielle/Wichtige stetige Verteilungen

Stetige Gleichverteilung

Die Gleichverteilung auf dem Intervall $[a, b] \subset \mathbb{R}$ ist gegeben durch

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \quad (3.33)$$

für alle $x \in \mathbb{R}$. Dann ist f eine Dichtefunktion da

(i) $f(x) \geq 0$ für alle $x \in \mathbb{R}$ und

(ii) $\int \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a} \int \mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a} \int_a^b 1dx = \frac{1}{b-a} \cdot x \Big|_a^b = \frac{1}{b-a}(b-a) = 1$ gilt.

Die Verteilungsfunktion ergibt sich dann als

$$F(x) = \begin{cases} 0 & \text{falls } x < a, \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b, \\ 1 & \text{falls } b < x. \end{cases} \quad (3.34)$$

Als Symbol verwenden wir $\mathcal{U}(a, b)$ und sagen, dass eine Zufallsvariable mit dieser Verteilung auf $[a, b]$ gleichverteilt ist. Eine auf $[0, 1]$ gleichverteilte Zufallsvariable nennen wir standardgleichverteilt.

Die stetige Gleichverteilung ist ein stetiges Analogon der Laplace-Wahrscheinlichkeit.

Sei nun X eine auf $[a, b]$ gleichverteilte Zufallsvariable. Dann ist $\mathbb{E}X = \frac{a+b}{2}$, da

$$\mathbb{E}X = \int x \cdot \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \cdot \frac{1}{2}x^2 \Big|_a^b = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{b+a}{2}$$

gilt. Für die Varianz berechnen wir zuerst

$$\mathbb{E}X^2 = \int x^2 \cdot \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \cdot \frac{1}{3} x^3 \Big|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

und damit die Varianz

$$\begin{aligned} \mathbb{V}\text{ar}X &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{1}{12}(4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2) = \frac{1}{12}(a^2 - 2ab + b^2) = \frac{(a-b)^2}{12} \end{aligned}$$

Exponentialverteilung

Die Exponentialverteilung ist das stetige Analogon der geometrischen Verteilung. Siehe hierzu Übungsaufgabe 4 auf Blatt 5.

Normalverteilung

Die Normalverteilung mit den Parametern μ und σ^2 , wobei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ sind, wird definiert durch die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2} \quad (3.35)$$

Bereits zu zeigen, dass f eine Dichte ist, ist nicht ganz trivial und deswegen übergehen wir diesen Schritt. Als Symbol verwenden wir $\mathcal{N}(\mu, \sigma^2)$. Es gilt $\mathbb{E}X = \mu$ und $\mathbb{V}\text{ar}X = \sigma^2$, was wir ebenfalls nicht nachrechnen.

Anmerkung: Die Normalverteilung hat eine große Bedeutung, da viele Messgrößen zumindest approximativ normalverteilt sind. Eine Erklärung werden wir mit dem *zentralen Grenzwertsatz* kennenlernen.

Anmerkung: Zuerst wurde die Normalverteilung von de Moivre definiert als Approximation der Binomialverteilung für große n . Gauß hat die Wichtigkeit der Normalverteilung gezeigt, weshalb die Dichtefunktion auch Gaußsche Glockenkurve genannt wird.

Die Standardnormalverteilung ist der Spezialfall mit $\mu = 0$ und $\sigma^2 = 1$, ihre Dichte ist

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3.36)$$

und ihre Verteilungsfunktion $\Phi(x)$, die sich nicht durch elementare Funktionen darstellen lässt.

Aus der (Achsen-)Symmetrie von φ bezüglich $x = 0$, d. h.

$$\varphi(-x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-x)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \varphi(x)$$

folgt die Punktsymmetrie von Φ bezüglich des Punktes $(0, 1/2)$ und es gilt

$$\Phi(-x) = 1 - \Phi(x). \quad (3.37)$$

3 Zufallsvariablen und Verteilungen

Standardisierung Die Verteilungsfunktion F einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen X lässt sich durch Umrechnen aus der Verteilungsfunktion einer Standardnormal-verteilten Zufallsvariablen ausdrücken. Es gilt

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right); \quad (3.38)$$

die Zufallsvariable $Z = \frac{X - \mu}{\sigma}$ ist $\mathcal{N}(0, 1)$ -verteilt.

Additivität

Satz 3.38: Seien X und Y unabhängige Zufallsvariablen, die $\mathcal{N}(\mu_X, \sigma_X^2)$ resp. $\mathcal{N}(\mu_Y, \sigma_Y^2)$ -verteilt sind. Dann ist deren Summe $Z = X + Y$ wieder normalverteilt, und zwar $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Diese Eigenschaft wird *Faltungsstabilität* genannt.

Beispiel 3.39: Ein Café-Besitzer weiß aus Erfahrung, dass der Umsatz am Samstag normalverteilt angenommen werden kann, genauso der am Sonntag. Sei X der Umsatz am Samstag einer festen Woche, sei Y der Umsatz am Sonntag danach. Als Erwartungswert setzt der Besitzer an: $\mathbb{E}X = 1000$, $\mathbb{E}Y = 1200$, für die Standardabweichungen $\sqrt{\text{Var}X} = 100$, $\sqrt{\text{Var}Y} = 200$, jeweils in Euro. Dann ist der Umsatz am Wochenende $Z = X + Y$ normalverteilt mit $\mathbb{E}Z = 2200$ und $\sqrt{\text{Var}Z} = \sqrt{100^2 + 200^2} = 223,6$.

3.4 Mehrdimensionale Verteilungen

Bis jetzt hatten wir nur selten mit mehreren Zufallsvariablen gleichzeitig zu tun. Bei der Linearität des Erwartungswertes und der Faltung hatten wir da etwas vorgegriffen, indem wir für zwei Zufallsvariablen X und Y auch deren Summe als Zufallsvariable angesehen haben.

3.4.1 Gemeinsamen und marginale Verteilungen

Wenn wir Verteilungen von mehreren Zufallsvariablen X_1, \dots, X_n betrachten, so können wir diese als einen Zufallsvektor $X = (X_1, \dots, X_n)$ auffassen.

Eigentlich sollten wir X als Spaltenvektor schreiben und damit konsequenterweise $X = (X_1, \dots, X_n)^T$. Wir lassen das der besseren Lesbarkeit zu Grunde oft ungeschrieben.

Definition 3.40 (Zufallsvektor): Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X_1, \dots, X_n Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$ für $1 \leq i \leq n$. Die Abbildung $X : \Omega \rightarrow \mathbb{R}^n$ heißt dann *n-dimensionaler Zufallsvektor* mit den Komponenten (oder Koordinaten) X_1, \dots, X_n . Die Funktion $F_X : \mathbb{R}^n \rightarrow [0, 1]$ definiert durch

$$F_X(x_1, \dots, x_n) = F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (3.39)$$

heißt *gemeinsame (oder multivariate) Verteilungsfunktion* des Zufallsvektors $X = (X_1, \dots, X_n)$.

Anmerkung: Dann ist $X : \Omega \rightarrow \mathbb{R}^n$ eine messbare Funktion. Dazu müssten wir uns eigentlich die Regularitätsbedingung der Messbarkeit auf \mathbb{R}^n genauer ansehen. Man kann zeigen, dass die Messbarkeit von $X = (X_1, \dots, X_n)$ zur Messbarkeit aller Koordinaten X_1, \dots, X_n ist.

Satz 3.41 (Eigenschaften multivariater Verteilungsfunktionen): Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor mit der Verteilungsfunktion F_X . Dann gilt

(i) *Monotonie:* F_X ist monoton wachsend mit Werten im Intervall $[0, 1]$, formal

$$F_X(x_1, \dots, x_n) \leq F_X(x_1 + h_1, \dots, x_n + h_n)$$

für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $h_1, \dots, h_n \geq 0$.

(ii) *Asymptotisches Verhalten:* Für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $1 \leq i \leq n$ gilt

$$\lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = 0,$$

weiterhin

$$\lim_{x_1, \dots, x_n \rightarrow \infty} F_X(x_1, \dots, x_n) = 1.$$

Seien $1 \leq i_1 < \dots < i_k \leq n$ Indizes und $\{j_1, \dots, j_{n-k}\} := \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$ die komplementäre Indexmenge. Dann ist

$$\lim_{x_{j_1}, \dots, x_{j_{n-k}} \rightarrow \infty} F_X(x_1, \dots, x_n) = \lim_{x_{j_1}, \dots, x_{j_{n-k}} \rightarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k})$$

die Verteilungsfunktion der *k-dimensionalen Marginalverteilung* der gemeinsamen Verteilung.

(iii) *Rechtsstetigkeit:* F_X ist stetig von rechts mit linksseitigen Grenzwerten, formal

$$F_X(x_1, \dots, x_n) = \lim_{h_1, \dots, h_n \rightarrow 0} F_X(x_1 + h_1, \dots, x_n + h_n)$$

für alle $x_1, \dots, x_n \in \mathbb{R}$ und alle $h_1, \dots, h_n \geq 0$.

Anmerkung: (i) Ein Zufallsvektor heißt diskret, wenn jede Koordinate X_i für $a \leq i \leq n$ diskret ist. Dann existiert eine Wahrscheinlichkeitsfunktion $p_X = p_{X_1, \dots, X_n}$ von X .

(ii) Ist jede Koordinate X_i für $a \leq i \leq n$ stetig, so nennen wir auch den Zufallsvektor stetig. Es existiert dann eine integrierbare Funktion $f_X = f_{X_1, \dots, X_n}$ von X .

(iii) Analog zur Verteilungsfunktion einer Marginalverteilung können wir die zugehörige Wahrscheinlichkeitsfunktion oder Dichte durch aufsummieren resp. integrieren erhalten. Seien $\{i_1, \dots, i_k\}$ und $\{j_1, \dots, j_{n-k}\}$ zwei komplementäre Indexmengen mit Indizes aus $\{1, \dots, n\}$. Dann gilt

$$\sum_{j_1, \dots, j_{n-k}} p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k})$$

und

$$\int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_{j_1} \cdots dx_{j_{n-k}} = f_{X_{i_1}, \dots, X_{i_k}}(x_{i_1}, \dots, x_{i_k}).$$

Satz 3.42 (Transformation von Zufallsvektoren): Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor und $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $m \geq 1$. Dann ist $u(X)$ ein m -dimensionaler Zufallsvektor.

Anmerkung: Sind die Zufallsvektoren $X = (X_1, \dots, X_{n_1})$ und $Y = (Y_1, \dots, Y_{n_2})$ unabhängig und $u : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{m_1}$ und $v : \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{m_2}$ mit $m_1, m_2 \geq 1$ zwei messbare Funktionen. Dann sind auch die Vektoren $u(X)$ und $v(Y)$ unabhängig.

Beispiel 3.43: Sei $X = (X_1, X_2)$ ein Zufallsvektor mit der Dichte

$$f_X(x_1, x_2) = 4 x_1 x_2 \mathbb{1}_{[0,1]}(x_1) \mathbb{1}_{[0,1]}(x_2) \quad (3.40)$$

für alle $x_1, x_2 \in \mathbb{R}$. Dann ergibt sich für die Dichten der Marginalverteilung

$$f_{X_1}(x_1) = \int 4 x_1 x_2 \mathbb{1}_{[0,1]}(x_1) \mathbb{1}_{[0,1]}(x_2) dx_2 = 4 x_1 \mathbb{1}_{[0,1]}(x_1) \int_0^1 x_2 dx_2 = 4 x_1 \cdot \frac{1}{2} x_2^2 \Big|_0^1 = 2 x_1$$

3 Zufallsvariablen und Verteilungen

und analog

$$f_{X_2}(x_2) = \int 4 x_1 x_2 \mathbb{1}_{[0,1]}(x_1) \mathbb{1}_{[0,1]}(x_2) dx_1 = 2 x_2$$

Zusätzlich berechnen wir die Wahrscheinlichkeit:

$$\begin{aligned} \mathbb{P}(X_1 \leq 2X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{2x_2} f_X(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{2x_2} 4 x_1 x_2 \mathbb{1}_{[0,1]}(x_1) \mathbb{1}_{[0,1]}(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} 4 x_2 \mathbb{1}_{[0,1]}(x_2) \left(\int_{-\infty}^{2x_2} x_1 \mathbb{1}_{[0,1]}(x_1) dx_1 \right) dx_2 \\ &= \int_{-\infty}^{\infty} 4 x_2 \left(\mathbb{1}_{[0,1/2)}(x_2) \int_{-\infty}^{2x_2} x_1 \mathbb{1}_{[0,1]}(x_1) dx_1 + \mathbb{1}_{[1/2,1]}(x_2) \int_{-\infty}^{2x_2} x_1 \mathbb{1}_{[0,1]}(x_1) dx_1 \right) dx_2 \\ &= \int_{-\infty}^{\infty} 4 x_2 \left(\mathbb{1}_{[0,1/2)}(x_2) \int_0^{2x_2} x_1 dx_1 + \mathbb{1}_{[1/2,1]}(x_2) \int_0^1 x_1 dx_1 \right) dx_2 \\ &= \int_{-\infty}^{\infty} 4 x_2 \left(\mathbb{1}_{[0,1/2)}(x_2) \frac{1}{2} x_1^2 \Big|_0^{2x_2} + \mathbb{1}_{[1/2,1]}(x_2) \frac{1}{2} x_1^2 \Big|_0^1 \right) dx_2 \\ &= \int_0^{1/2} 4 x_2 \frac{1}{2} \cdot 4 x_2^2 dx_2 + \int_{1/2}^1 4 x_2 \frac{1}{2} dx_2 = \int_0^{1/2} 8 x_2^3 dx_2 + \int_{1/2}^1 2 x_2 dx_2 \\ &= 8 \cdot \frac{1}{4} x_2^4 \Big|_0^{1/2} + 2 \cdot \frac{1}{2} x_2^2 \Big|_{1/2}^1 = \frac{1}{8} + \frac{3}{4} = \frac{7}{8} \end{aligned}$$

Anmerkung (Geometrische Interpretation): An Hand des vorangegangenen Beispiels machen uns die geometrische Bedeutung des Ereignisses bewusst. Dazu betrachten wir zunächst die gegebene Dichte in der $x_1 - x_2$ -Ebene. Die geometrische Interpretation des Ereignisses ist die folgende: Es ergibt sich ein Trapez, welches sich in ein Rechteck und ein Dreieck aufteilen lässt. Diese Teilflächen entsprechen den Integrationsbereichen aus der Rechnung. Hier sieht man, dass der Volumenanteil des Ereignisses $3/4$ wäre, die Wahrscheinlichkeit aber größer sein muss, da die Dichte in der rechten oberen Ecke konzentriert ist.

Die Wahrscheinlichkeit des Ereignisses ist das Integral der Dichte, integriert über den Raumbereich, d.h. ein gewichtetes Volumen.

Diese Interpretation war nicht von der gewählten Dichte abhängig. Ein wichtiger Fall ergibt sich bei der Dichte der Gleichverteilung.

Geometrische Verteilungen

Betrachte jetzt die n -dimensionale Gleichverteilung für den Zufallsvektor $X = (X_1, \dots, X_n)$ auf dem Quader $Q := [a_1, b_1] \times \dots \times [a_n, b_n]$, deren Dichte

$$f_X(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{b_i - a_i} \mathbb{1}_{[a_i, b_i]}(x_i) \quad (3.41)$$

ist. Ihre 1-dimensionalen Marginalverteilungen der X_i sind Gleichverteilungen auf den Intervallen $[a_i, b_i]$ für alle $1 \leq i \leq n$.

Ein Ereignis A mit positiver Wahrscheinlichkeit kann bestimmt werden durch eine Bedingung der Form

$$g(X_1, \dots, X_n) \geq 0,$$

wobei $g : \mathbb{R}^n \rightarrow \mathbb{R}$ eine differenzierbare Funktion ist, die außerhalb des Quaders Q verschwindet.

Zum Beispiel betrachten wir für $n = 3$ den Einheitswürfel $Q = [0, 1]^3$ und die Funktion $g(X_1, X_2, X_3) = X_3 - X_1 \geq 0$.

Wir bestimmen die Wahrscheinlichkeit von A als

$$\begin{aligned} \mathbb{P}(g(X_1, \dots, X_n) \geq 0) &= \int_A f_X(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_A \prod_{i=1}^n \frac{1}{b_i - a_i} \mathbb{1}_{[a_i, b_i]}(x_i) dx_1 \dots dx_n \\ &= \prod_{i=1}^n \frac{1}{b_i - a_i} \int_A \prod_{i=1}^n \mathbb{1}_{[a_i, b_i]}(x_i) dx_1 \dots dx_n \\ &= \frac{1}{\text{Vol}(Q)} \int_A \prod_{i=1}^n \mathbb{1}_{[a_i, b_i]}(x_i) dx_1 \dots dx_n \\ &= \frac{1}{\text{Vol}(Q)} \int_A 1 dx_1 \dots dx_n \\ &= \frac{1}{\text{Vol}(Q)} \cdot \text{Vol}(A) \end{aligned}$$

d. h. die Ereignisse können geometrisch als Teilflächen/-volumina/-hyperebenen verstanden werden.

3.4.2 Erwartungswert und Varianz

Definition 3.44 (Erwartungswert): (i) Sei $X = (X_1, \dots, X_n)^T$ ein diskreter Zufallsvektor mit gemeinsamer Wahrscheinlichkeitsfunktion p_X bei dem alle Koordinaten X_i mit $1 \leq i \leq n$ einen Erwartungswert besitzen. Dann ist

$$\begin{aligned} \mathbb{E}X &= \mathbb{E} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \sum_{x_1, \dots, x_n} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} p_X(x_1, \dots, x_n) = \begin{pmatrix} \sum_{x_1, \dots, x_n} x_1 p_X(x_1, \dots, x_n) \\ \vdots \\ \sum_{x_1, \dots, x_n} x_n p_X(x_1, \dots, x_n) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{x_1} x_1 p_{X_1}(x_1) \\ \vdots \\ \sum_{x_n} x_n p_{X_n}(x_n) \end{pmatrix} = \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{pmatrix} \end{aligned} \quad (3.42)$$

der Erwartungswert von X .

3 Zufallsvariablen und Verteilungen

(ii) Sei $Y = (Y_1, \dots, Y_n)^T$ ein stetiger Zufallsvektor mit gemeinsamer Wahrscheinlichkeitsdichte f_Y bei dem alle Koordinaten Y_i mit $1 \leq i \leq n$ integrierbar sind. Dann ist der Erwartungswert von Y gegeben durch

$$\mathbb{E}Y = \mathbb{E} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \int_{y_1, \dots, y_n} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} f_Y(y_1, \dots, y_n) dy_1 \dots dy_n = \begin{pmatrix} \mathbb{E}Y_1 \\ \vdots \\ \mathbb{E}Y_n \end{pmatrix}. \quad (3.43)$$

Korollar 3.45 (Transformationsformel für den Erwartungswert): Aus dem Satz 3.42 und der obigen Definition 3.44 folgt direkt:

Sei $X = (X_1, \dots, X_n)$ ein diskreter Zufallsvektor mit der gemeinsamen Wahrscheinlichkeitsfunktion p und sei $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Funktion. Dann gilt

$$\mathbb{E}(u(X)) = \mathbb{E}(u(X_1, \dots, X_n)) = \sum_{x_1, \dots, x_n} u(x_1, \dots, x_n) p(x_1, \dots, x_n). \quad (3.44)$$

(i) (ii) Sei $Y = (Y_1, \dots, Y_n)$ ein stetiger Zufallsvektor mit der gemeinsamen Wahrscheinlichkeitsdichte f und sei $u : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Funktion. Dann gilt

$$\mathbb{E}(u(Y)) = \mathbb{E}(u(Y_1, \dots, Y_n)) = \int_{y_1, \dots, y_n} u(y_1, \dots, y_n) f(y_1, \dots, y_n). \quad (3.45)$$

Satz 3.46 (Eigenschaften des Erwartungswertes): Seien $X = (X_1, \dots, X_n)^T$ und $Y = (Y_1, \dots, Y_n)^T$ beide n -dimensionale Zufallsvektoren, $a, b \in \mathbb{R}$ und $A \in \mathbb{R}^{k \times n}$ eine Matrix. Dann gilt

- (i) $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$,
- (ii) $\mathbb{E}(AX) = A\mathbb{E}X$,

der Erwartungswert ist also linear. Sind X und Y unabhängig, so gilt

$$\mathbb{E}(XY^T) = \mathbb{E}(X)\mathbb{E}(Y)^T. \quad (3.46)$$

Definition 3.47 (Kovarianzmatrix): Sei $X = (X_1, \dots, X_n)^T$ ein Zufallsvektor und definiere die Kovarianzen $s_{ij} := \mathbb{Cov}(X_i, X_j)$ für $1 \leq i, j \leq n$. Dann ist

$$\Sigma := \mathbb{V}_{\text{ar}}X = \mathbb{E} \left((X - \mathbb{E}X)(X - \mathbb{E}X)^T \right) = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix} \quad (3.47)$$

die Kovarianzmatrix von X .

Anmerkung: Zu beachten ist, dass auf der Diagonalen der Kovarianzmatrix die Varianzen zu finden sind. Wir wissen schon, dass $\mathbb{Cov}(X_i, X_i) = \mathbb{V}_{\text{ar}}(X_i)$ gilt. Deswegen haben wir den Zusammenhang $s_{ii} = \sigma_i^2$ für die Einträge der Kovarianzmatrix mit der Varianz.

Satz 3.48: Die Kovarianzmatrix Σ ist

- (i) symmetrisch, d. h. $s_{ij} = s_{ji}$,
 - (ii) nicht-negativ definit, d. h. $x^T \Sigma x = x \geq 0$ für alle $x \in \mathbb{R}$.
- Gilt sogar $x^T \Sigma x > 0$, so ist Σ positiv definit.

Multinomialverteilung

Die Multinomialverteilung ist das multivariate Analogon der Binomialverteilung. Dabei betrachten wir eine Experiment mit k möglichen Ausgängen $1, \dots, k$ und den Einzelwahrscheinlichkeiten η_1, \dots, η_k mit $\sum_{i=1}^k \eta_i = 1$. Wir ziehen n -mal mit Zurücklegen.

Bevor wir uns den Zufallsvariablen und der Wahrscheinlichkeitsfunktion zuwenden, benötigen wir noch die folgende Definition:

Definition 3.49: Seien $n, n_1, \dots, n_k \in \mathbb{N}_0$ mit $\sum_{i=1}^k n_i = n$. Dann ist

$$\binom{n}{n_1, \dots, n_k} = \frac{n!}{n_1! \dots n_k!} \quad (3.48)$$

der Multinomialkoeffizient. Der Vollständigkeit halber sagen wir $\binom{n}{n_1, \dots, n_k} = 0$, falls $\sum_{i=1}^k n_i \neq n$ oder $n_i < 0$ für ein $1 \leq i \leq k$.

Lemma 3.50: Für den Multinomialkoeffizienten gelten folgende Identitäten

$$\binom{n}{m, n-m} = \frac{n!}{m!(n-m)!} = \binom{n}{m} \quad (3.49)$$

und

$$\binom{n}{n_1, \dots, n_k} = \binom{n}{n_i} \cdot \binom{n-n_i}{n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k} \quad (3.50)$$

für alle $1 \leq i \leq k$.

Bezeichne nun N_i die Anzahl der Experimente mit dem Ausgang i . Die gemeinsame Wahrscheinlichkeitsfunktion ist

$$p_{N_1, \dots, N_k}(n_1, \dots, n_k) = \binom{n}{n_1, \dots, n_k} \eta_1^{n_1} \dots \eta_k^{n_k}.$$

Wir müssen zuerst nachprüfen, dass p_{N_1, \dots, N_k} tatsächlich eine Wahrscheinlichkeitsfunktion ist.

(i) $p_{N_1, \dots, N_k}(n_1, \dots, n_k)$ für alle $n_1, \dots, n_k \in \mathbb{N}_0$ wegen der Definition des Binomialkoeffizienten.

(ii) Die Normiertheit ergibt sich induktiv aus der Gleichung (3.50) durch sukzessive Berechnung der Marginalverteilungen.

Insbesondere folgt, dass die 1-dimensionalen Marginalverteilungen Binomialverteilungen mit dem entsprechenden Parameter η_i sind.

Beispiel 3.51 (Multivariate Normalverteilung): Wir betrachten den Fall $n = 2$. Ein Zufallsvektor $X = (X_1, X_2)^T$ mit der Dichte

$$f_X(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 \cdot \det(\Sigma)}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (3.51)$$

heißt bivariat normalverteilt mit den Parametern $\mu = (\mu_1, \mu_2)^T$ und $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, wobei σ_1^2 und σ_2^2 die Varianzen der zwei Koordinaten sind und $\rho < 0$ deren Korrelationskoeffizient.

Anmerkung: (a) Die Marginalverteilungen von X_1 und X_2 sind $\mathcal{N}(\mu_1, \sigma_1^2)$ resp. $\mathcal{N}(\mu_2, \sigma_2^2)$.

(b) Zu beachten ist, dass X_1 und X_2 im Allgemeinen nicht unabhängig sind. Dies ist den Marginalverteilungen nicht anzusehen.

3 Zufallsvariablen und Verteilungen

(c) Gilt hingegen Unabhängigkeit, so ist Σ eine Diagonalmatrix. Unter der zusätzlichen Bedingung $\sigma_1 = \sigma_2$ ist die Dichte rotationsinvariant.

(d) Wie im eindimensionalen Fall durch μ und σ^2 wird im mehrdimensionalen Fall die Verteilung eines normalverteilten Zufallsvektors eindeutig durch $(\mu_1, \mu_2)^T$ und Σ bestimmt.

(e) Die Definition lässt sich auf beliebige n ausweiten. Die Dimension findet sich im Exponenten von $\sqrt{2\pi}$ wiederfinden.

Beispiel 3.52: Wir beziehen uns wieder auf das Beispiel 3.39, in dem wir das Café betrachten. Seien X_1, \dots, X_7 die Umsätze an den sieben Wochentagen. Definiere $X = (X_1, \dots, X_7)^T$. Der Besitzer erzählt uns, dass die Samstage und Sonntage von den anderen Tagen und ebenfalls voneinander unabhängig sind, die anderen Wochentage jedoch sich oft gemeinsam verhalten. Konkret würden sich die Tage von Montag bis Donnerstag ähnlich entwickeln, der Freitag hingegen genau konträr. Wir formulieren dazu die folgende Kovarianzmatrix

$$\Sigma := \text{Var}X = \mathbb{E} \left((X - \mathbb{E}X)(X - \mathbb{E}X)^T \right) = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} & s_{27} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} & s_{37} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} & s_{47} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} & s_{57} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} & s_{67} \\ s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & s_{77} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & 0 & 0 \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & 0 & 0 \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & 0 & 0 \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & 0 & 0 \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 100^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 200^2 \end{pmatrix};$$

desweiteren wissen wir, dass $s_{ij} = s_{ji}$ und weiterhin, dass $s_{5j} < 0$ mit $j = 1, 2, 3, 4$ und die Einträge $s_{21}, s_{31}, s_{32}, s_{41}, s_{42}, s_{43} > 0$. Über die Varianzen $\sigma_1^2 = s_{11}, \dots, \sigma_5^2 = s_{55}$ wissen wir leider nicht mehr.

3.4.3 Bedingte Verteilungen

Definition 3.53: Seien X und Y diskrete Zufallsvariablen und sei x mit $\mathbb{P}(X = x) > 0$. Die Wahrscheinlichkeitsverteilung auf \mathbb{R} , die jeder messbaren Teilmenge $A \subset \mathbb{R}$ die Wahrscheinlichkeit $\mathbb{P}(Y \in A | X = x)$ zuordnet, wird *bedingte Verteilung* von Y gegeben $X = x$ genannt. Die zugehörige bedingte Wahrscheinlichkeitsfunktion wird definiert durch

$$p_{Y|X}(y|x) := \mathbb{P}(Y = y | X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \quad (3.52)$$

für alle y aus dem Wertebereich von Y .

Notation: Die bedingte Wahrscheinlichkeitsfunktion wird gut abgekürzt mit $p(y|x)$ oder $p_{Y|X}$ oder $p_{Y|X=x}$.

Anmerkung: (i) Die bedingte Wahrscheinlichkeitsfunktion ist eine Wahrscheinlichkeitsfunktion, da

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} > 0$$

und

$$\sum_y p_{Y|X}(y|x) = \sum_y \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{\sum_y p_{X,Y}(x,y)}{p_X(x)} = \frac{p_X(x)}{p_X(x)} = 1$$

gelten.

(ii) Wir haben in der Definition \mathbb{R} und messbare Mengen aus \mathbb{R} verwendet. Die Definition lässt sich auch auf Zufallsvektoren ausdehnen. Dazu betrachten wir die messbare Menge $A \in \mathbb{R}^m$ und $x = (x_1, \dots, x_m)^T \in \mathbb{R}^m$. Dann wird $(Y \in A | X = x)$ die bedingte Verteilung von Y gegeben $X = x$ genannt. Die bedingte Wahrscheinlichkeitsfunktion wird ganz analog definiert durch

$$p_{Y|X}(y|x) = \mathbb{P}(Y = y | X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)} \quad (3.53)$$

für $y \in \mathbb{R}^m$ und $x \in \mathbb{R}^m$, falls $p_X(x) \neq 0$ ist.

Beispiel 3.54: Seien X und Y unabhängige Zufallsvariablen, die $\text{Bin}(n, \eta)$ resp. $\text{Bin}(m, \eta)$ verteilt sind. Dann ist für die bedingte Wahrscheinlichkeitsfunktion von X gegeben $X + Y = k$ gegeben durch

$$\begin{aligned} p_{X|X+Y}(i|k) &= \mathbb{P}(X = i | X + Y = k) = \frac{\mathbb{P}(X = i, X + Y = k)}{\mathbb{P}(X + Y = k)} = \frac{\mathbb{P}(X = i, Y = k - i)}{\mathbb{P}(X + Y = k)} \\ &= \frac{\mathbb{P}(X = i) \mathbb{P}(Y = k - i)}{\mathbb{P}(X + Y = k)} \\ &= \frac{\binom{n}{i} \eta^i (1 - \eta)^{n-i} \cdot \binom{m}{k-i} \eta^{k-i} (1 - \eta)^{m-k+i}}{\binom{n+m}{k} \eta^k (1 - \eta)^{n+m-k}} = \frac{\binom{n}{i} \binom{m}{k-i}}{\binom{n+m}{k}}, \end{aligned}$$

d. h. die Verteilung von X gegeben $X + Y = k$ ist eine hypergeometrische Verteilung $\text{Hyp}(n + m, n, k)$.

Satz 3.55: Die gemeinsame Wahrscheinlichkeitsfunktion unabhängiger diskreter Zufallsvariablen X, Y ist gleich dem Produkt der marginalen Wahrscheinlichkeitsfunktion,

$$p_{X,Y}(x, y) = p_X(x) p_Y(y). \quad (3.54)$$

Umgekehrt folgt aus der Produktgestalt der gemeinsamen Wahrscheinlichkeitsfunktion

$$p_{X,Y}(x, y) = g(x) h(y), \quad (3.55)$$

dass X und Y unabhängige Zufallsvariablen und ihre Wahrscheinlichkeitsfunktionen

$$p_X(x) = \alpha g(x) \quad \text{und} \quad p_Y(y) = \beta h(y) \quad (3.56)$$

sind, wobei $\alpha := (\sum_x g(x))^{-1}$ und $\beta := (\sum_y h(y))^{-1}$ sind.

Satz 3.56: Zwei Zufallsvariablen X und Y sind genau dann unabhängig, wenn die bedingte Verteilung von Y gegeben $X = x$ nicht von x abhängt, d. h. $p_{Y|X}(y|x) = h(y)$ für alle $x \in X(\Omega)$.

BEWEIS Sind X und Y unabhängige Zufallsvariablen, so gilt $p_{X,Y}(x, y) = p_X(x) p_Y(y)$ und somit

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)} = \frac{p_X(x) p_Y(y)}{p_X(x)} = p_Y(y). \quad (3.57)$$

Umgekehrt gilt, wenn $h(y) = p_{Y|X}(y|x)$ nicht von x abhängt, dass $p_{X,Y}(x, y) = p_X(x) h(y)$ und damit sind X und Y unabhängig. \square

Anmerkung: Insbesondere folgt aus dem Satz, dass $p_{Y|X}(y|x) = p_Y(y)$, falls X und Y unabhängig sind.

3 Zufallsvariablen und Verteilungen

Anmerkung: Für stetige Zufallsvariablen können wir in der Definition nicht den Weg $\mathbb{P}(y \in A | X = x) = \frac{\mathbb{P}(y \in A, X=x)}{\mathbb{P}(X=x)}$ gehen, da $\mathbb{P}(X = x) = 0$ gilt.

Definition 3.57: Seien X und Y stetige Zufallsvariablen mit gemeinsamer Dichte $f_{X,Y}(x,y)$. Dann wird die bedingte Dichte $f_{Y|X}(y|x)$ von Y gegeben $X = x$ definiert durch

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{falls } f_X(x) > 0, \\ 0 & \text{falls } f_X(x) = 0. \end{cases} \quad (3.58)$$

Die bedingte Verteilung von X gegeben $X = x$ ist die Verteilung mit der Dichtefunktion $f_{Y|X}(y|x)$, d.h. $\mathbb{P}(Y \in A | X = x) = \int_A f_{Y|X}(y|x) dy$.

Anmerkung: (i) Auch im stetigen Fall können wir die Definition auf mehrdimensionale Vektoren erweitern.

(ii) Der Zusammenhang zwischen bedingten Verteilungen und Unabhängigkeit gilt analog zum diskreten Fall.

3.4.4 Transformation von Dichten

Wir hatten uns Transformationen bei der Berechnung von Erwartungswerten angesehen. In diesem Abschnitt werden wir uns mit Dichten und deren Umrechnungen beschäftigen.

Beispiel 3.58: Sei $X \sim \mathcal{U}(0, 1]$ und $Y := u(X) = \frac{1}{X}$. Dann gilt mit der Ausnutzung der Stetigkeit von X , dass

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(\frac{1}{X} \leq y\right) = \mathbb{P}\left(X \geq \frac{1}{y}\right) = 1 - \mathbb{P}(X \leq 1/y) = 1 - \frac{1}{y} \mathbb{1}_{[1, \infty)}(y).$$

Da die Verteilungsfunktion stetig und außer in $y = 1$ auch stetig differenzierbar ist, haben wir die Dichte

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} \left(1 - \frac{1}{y} \mathbb{1}_{[1, \infty)}(y)\right) = \frac{d}{dy} \left(1 - \frac{1}{y}\right) \mathbb{1}_{[1, \infty)}(y) = \frac{1}{y^2} \mathbb{1}_{[1, \infty)}(y).$$

Beispiel 3.59 (Lineare Transformation): Sei X eine beliebige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$ und seien $a, b \in \mathbb{R}$ mit $a \neq 0$. Dann ist $Y := u(X) = aX + b$ mit der Verteilungsfunktion

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}(aX \leq y - b) \\ &= \begin{cases} \mathbb{P}\left(X \leq \frac{y-b}{a}\right) & \text{falls } a > 0, \\ \mathbb{P}\left(X \geq \frac{y-b}{a}\right) & \text{falls } a < 0, \end{cases} \\ &= \begin{cases} \mathbb{P}\left(X \leq \frac{y-b}{a}\right) & \text{falls } a > 0, \\ 1 - \mathbb{P}\left(X \leq \frac{y-b}{a}\right) + \mathbb{P}\left(X = \frac{y-b}{a}\right) & \text{falls } a < 0, \end{cases} \\ &= \begin{cases} F_X\left(\frac{y-b}{a}\right) & \text{falls } a > 0, \\ 1 - F_X\left(\frac{y-b}{a}\right) + \mathbb{P}\left(X = \frac{y-b}{a}\right) & \text{falls } a < 0. \end{cases} \end{aligned}$$

Falls X (absolut) stetig ist mit der Dichte f_X , so ist auch Y stetig mit der Dichte $f_Y(y) = f_{aX+b}(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

Als nächstes beschäftigen wir uns mit einem direkten Weg, die Dichte zu berechnen, ohne über die Verteilungsfunktion zu gehen.

Satz 3.60 (Transformationsformel für Dichten): Sei X eine stetige Zufallsvariable mit Werten in $I \subset \mathbb{R}$, wobei I ein offenes Intervall ist, und der Dichte f_X . Ist $J \subset \mathbb{R}$ ebenfalls ein offenes Intervall und $u : I \rightarrow J$ bijektiv mit u und u^{-1} stetig diffbar, so besitzt $Y := u(X)$ die Dichte

$$f_Y(y) = f_X(u^{-1}(y)) \left| \frac{d}{dy} u^{-1}(y) \right| \quad (3.59)$$

für $y \in J$ und $f_Y(y) = 0$ für $y \in \mathbb{R} \setminus J$.

BEWEIS Die Funktion u ist entweder streng monoton fallend oder streng monoton steigend. Wir untersuchen den letzteren Fall. Dann gilt für $a, b \in J$, dass

$$\begin{aligned} \mathbb{P}(a < Y \leq b) &= \mathbb{P}(a < u(X) \leq b) = \mathbb{P}(u^{-1}(a) < X \leq u^{-1}(b)) \\ &= \int_{u^{-1}(a)}^{u^{-1}(b)} f_X(x) dx = \int_a^b f_X(u^{-1}(y)) \left| \frac{d}{dy} u^{-1}(y) \right| dy. \end{aligned}$$

Der zweite Fall ist analog. □

Beispiel 3.61: Wir nehmen das gleiche Szenario wie im Beispiel 3.58. Die Funktion $u(x) = \frac{1}{x}$ ist bijektiv und stetig differenzierbar, $u^{-1}(y) = \frac{1}{y}$ ebenso. Dann hat $Y = u(X)$ die Dichte

$$f_Y(y) = f_X\left(\frac{1}{y}\right) \cdot \frac{1}{y^2} = \mathbb{1}_{(0,1]}\left(\frac{1}{y}\right) \cdot \frac{1}{y^2} = \frac{1}{y^2} \mathbb{1}_{[1,\infty)}(y) \quad (3.60)$$

Beispiel 3.62 (Kubische Transformation): Sei X eine stetige Zufallsvariable und $Y := u(X) = X^3$. Dann ist die Funktion $u(x) = x^3$ bijektiv und stetig diffbar, ihre Umkehrung

$$u^{-1}(y) = |y|^{\frac{1}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right)$$

ebenfalls mit

$$\frac{d}{dy} u^{-1}(y) = \frac{1}{3} |y|^{-\frac{2}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right).$$

Die Dichte von Y ist

$$\begin{aligned} f_Y(y) &= f_X\left(|y|^{\frac{1}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right)\right) \cdot \frac{1}{3} |y|^{-\frac{2}{3}} \cdot \left(-\mathbb{1}_{(-\infty,0)}(y) + \mathbb{1}_{[0,\infty)}(y) \right) \\ &= \begin{cases} f_X\left(y^{\frac{1}{3}}\right) \cdot \frac{1}{3} \cdot y^{-\frac{2}{3}} & \text{falls } y \geq 0, \\ -f_X\left(-|y|^{\frac{1}{3}}\right) \cdot \frac{1}{3} \cdot |y|^{-\frac{2}{3}} & \text{falls } y < 0. \end{cases} \end{aligned}$$

Die Funktionen sehen komplizierter aus als sie sind. Wir müssen einfach unterscheiden, ob y negativ ist oder nicht, da Wurzeln aus negativen Zahlen höchstens über die Betragsschreibweise, wie wir sie verwenden, definiert sind.

Beispiel 3.63 (Quadrierung): Sei X eine stetige Zufallsvariable und wir untersuchen die Transformation $Y := u(X) = X^2$. Dann ist die Funktion $u(x) = x^2$ für $u \geq 0$ bijektiv mit $u^{-1}(y) = \sqrt{y}$ und u und u^{-1} sind stetig diffbar. Dann ist

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}.$$

3 Zufallsvariablen und Verteilungen

Bierbei haben wir die Bedingung $X > 0$. Über den Umweg der Verteilungsfunktion

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) = F_X(\sqrt{y}),$$

die stetig und differenzierbar ist, erhalten wir die Dichte

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}.$$

Aber: Für eine beliebige stetige Zufallsvariable X haben wir die Verteilungsfunktion

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(|X| \leq \sqrt{y}) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}), \end{aligned}$$

die stetig und differenzierbar ist, und erhalten die Dichte

$$f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

An diesem Beispiel sehen wir, dass wir nicht ohne Nachdenken einfach die Formel verwenden können, sondern uns genau über die Voraussetzungen klar sein sollen.

Im Weiteren wollen wir uns mit Transformationen beschäftigen, die mehrere Zufallsvariablen einbeziehen. Wir können auch sagen, wir haben einen Zufallsvektor und betrachten die mehrdimensionale Transformation. Die Bildung einer Summe ist die Faltung der Dichten, die wir bereits kennengelernt haben.

Satz 3.64 (Produkt- und Quotienttransformationen): Seien X_1 und X_2 stetige Zufallsvariablen mit den Dichten f_{X_1} resp. f_{X_2} und seien sie unabhängig. Dann sind die Zufallsvariablen $Y := X_1 \cdot X_2$ und $Z := X_1/X_2$ stetig mit den Dichten

$$f_Y(y) = \int \frac{1}{|t|} f_{X_1}(t) f_{X_2}(y/t) dt \quad \text{für alle } y \in \mathbb{R} \quad (3.61)$$

und

$$f_Z(z) = \int |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \quad \text{für alle } z \in \mathbb{R}. \quad (3.62)$$

Anmerkung: Das gedankliche Vorgehen ist hier genauso wie bei der Faltung. Es wird über alle Werte aufsummiert/integriert, die die eine der beteiligten Zufallsvariablen annehmen kann, kombiniert mit dem Wert, der dann der anderen Zufallsvariable zukommt.

Beispiel 3.65: (a) Seien X_1 und X_2 unabhängig und beide $\mathcal{U}(0, 1)$ -verteilt, so haben wir die Dichten $f_{X_1}(x) = f_{X_2}(x) = \mathbb{1}_{(0,1)}(x)$. Die Dichte des Produktes $Y = X_1 \cdot X_2$ ist dann

$$\begin{aligned} f_Y(y) &= \int \frac{1}{|t|} \mathbb{1}_{(0,1)}(t) \mathbb{1}_{(0,1)}(y/t) dt \\ &= \int_0^1 \frac{1}{t} \mathbb{1}_{(0,1)}(y/t) dt \\ &= \int_y^1 \frac{1}{t} dt \quad \text{falls } 0 < y < 1, \\ &= \log t \Big|_y^1 = \log 1 - \log y = -\log y. \end{aligned}$$

(b) Seien X_1 und X_2 unabhängig und beide standardnormalverteilt, d. h. die Dichten sind $f_{X_1}(x) = f_{X_2}(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2x^2}$. Dann ist die Dichte von $Z = X_1/X_2$ gegeben durch

$$\begin{aligned} f_Z(z) &= \int |t| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot (zt)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot t^2} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |t| e^{-\frac{1}{2} \cdot t^2 (z^2+1)} dt \\ &= 2 \cdot \frac{1}{2\pi} \int_0^{\infty} t e^{-\frac{1}{2} \cdot t^2 (z^2+1)} dt \\ &= \frac{1}{\pi(z^2+1)} \int_0^{\infty} t(z^2+1) e^{-\frac{1}{2} \cdot t^2 (z^2+1)} dt \quad (\text{substituiere } v = \frac{1}{2} \cdot t^2 (z^2+1)) \\ &= \frac{1}{\pi(z^2+1)} \int_0^{\infty} e^{-v} dv = \frac{1}{\pi(z^2+1)} \cdot \left(-e^{-v} \Big|_0^{\infty} \right) = \frac{1}{\pi(z^2+1)} \cdot 1 = \frac{1}{\pi(z^2+1)} \end{aligned}$$

Die Transformation von Dichten, die wir im Satz kennengelernt haben, können wir allgemeiner mehrdimensional formulieren.

Satz 3.66 (Transformationsformel für gemeinsame Dichten): Sei $X = (X_1, \dots, X_n)$ ein stetiger Zufallsvektor mit der Dichte f_X . Seien $A, B \subset \mathbb{R}^n$ offene zusammenhängende Teilmengen, wobei X nur Werte in A annimmt. Sei weiter $u : A \rightarrow B$ eine bijektive Abbildung mit u, u^{-1} stetig differenzierbar. Dann besitzt $Y = (Y_1, \dots, Y_n) := u(X)$ die Dichte

$$f_Y(y_1, \dots, y_n) = f_X \left(u^{-1}(y_1, \dots, y_n) \right) \left| \det \left(J_{u^{-1}}(y_1, \dots, y_n) \right) \right| \cdot \mathbb{1}_B(y_1, \dots, y_n), \quad (3.63)$$

wobei $J_{u^{-1}}$ die Jacobi-Matrix von u^{-1} ist.

Anmerkung (Jacobi-Matrix): Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine differenzierbare Funktion mit den Komponenten f_1, \dots, f_m , die dem Wert $x = (x_1, \dots, x_n)$ den Wert $y = (f_1(x), \dots, f_m(x))$ zuordnet. Die Jacobi-Matrix J_f von f ist die $m \times n$ -Matrix aller erster partieller Ableitungen, d. h.

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Beispiel 3.67: Seien X_1 und X_2 Zufallsvariablen für die Wartezeiten zweier Personen an unabhängigen Service-Schaltern, d. h. X_1 und X_2 sind unabhängig und exponentialverteilt mit den Parametern λ_1 resp. λ_2 .

Wir interessieren uns für den Quotienten $\frac{X_1}{X_2}$, wollen den aber mit dem verallgemeinerten Satz 3.66 angehen. Da wir zur Anwendung eine Funktion benötigen, die beim zwei-dimensionalen Zufallsvektor wieder einen zwei-dimensionalen Vektor liefert, betrachten wir die Funktion

$$u \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1(x_1, x_2) \\ y_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2/x_1 \end{pmatrix} \quad \text{mit} \quad u^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \cdot y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2/x_1 \cdot x_1 \end{pmatrix}.$$

3 Zufallsvariablen und Verteilungen

Die entsprechende Jacobi-Matrix ist

$$J_{u^{-1}}(y_1, y_2) = \begin{pmatrix} \frac{\partial y_1}{\partial y_1} & \frac{\partial y_1}{\partial y_2} \\ \frac{\partial(y_2 \cdot y_1)}{\partial y_1} & \frac{\partial(y_2 \cdot y_1)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ y_2 & y_1 \end{pmatrix}$$

mit der Determinante $\det(J_{u^{-1}}(y_1, y_2)) = 1 \cdot y_1 - 0 \cdot y_2 = y_1$. Damit ergibt sich für den Zufallsvektor $Y = (Y_1, Y_2) = u(X_1, X_2)$ die Dichte

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= f_X \begin{pmatrix} y_1 \\ y_2 \cdot y_1 \end{pmatrix} \cdot |y_1| = \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot |y_1| \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_1 y_2) \\ &= \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot y_1 \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_2). \end{aligned}$$

Da wir uns für die Marginaldichte interessieren, integrieren wir y_1 aus:

$$\begin{aligned} f_{Y_2}(y_2) &= \int \lambda_1 \lambda_2 e^{-(\lambda_1 y_1 + \lambda_2 y_1 y_2)} \cdot y_1 \cdot \mathbb{1}_{(0, \infty)}(y_1) \mathbb{1}_{(0, \infty)}(y_2) \\ &= \lambda_1 \lambda_2 \mathbb{1}_{(0, \infty)}(y_2) \int_0^{\infty} y_1 e^{-(\lambda_1 + \lambda_2 y_2) y_1} dy_1, \end{aligned}$$

was wir auch durch die Anwendung von Satz 3.64, Gleichung (3.61) erhalten könnten. Das Integral lösen wir durch partielle Integration, wobei $\alpha := \lambda_1 + \lambda_2 y_2$ ist:

$$\begin{aligned} \int_0^{\infty} y_1 e^{-(\lambda_1 + \lambda_2 y_2) y_1} dy_1 &= \int_0^{\infty} \underbrace{y_1}_{:=v} \underbrace{e^{-\alpha y_1}}_{:=u'} dy_1 = y_1 \left(-\frac{1}{\alpha} \right) e^{-\alpha y_1} \Big|_0^{\infty} - \int_0^{\infty} \left(-\frac{1}{\alpha} \right) e^{-\alpha y_1} dy_1 \\ &= \frac{1}{\alpha} \int_0^{\infty} e^{-\alpha y_1} dy_1 = \frac{1}{\alpha^2} \underbrace{\int_0^{\infty} e^{-z} dz}_{=1} \quad (\text{Substitution } z := \alpha y_1) \\ &= \frac{1}{\alpha^2} = \frac{1}{(\lambda_1 + \lambda_2 y_2)^2}. \end{aligned}$$

Damit haben wir

$$f_{Y_2}(y_2) = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 y_2)^2} \mathbb{1}_{(0, \infty)}(y_2)$$

Als nächstes interessieren wir uns für den Erwartungswert. Wir substituieren $s := \lambda_1 + \lambda_2 y_2$, d. h. $y_2 = \frac{s - \lambda_1}{\lambda_2}$ und $ds = \lambda_2 dy_2$, und haben

$$\begin{aligned} \mathbb{E} \left(\frac{X_2}{X_1} \right) &= \mathbb{E}(Y_2) = \int_0^{\infty} y_2 \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 y_2)^2} dy_2 \\ &= \int_{\lambda_1}^{\infty} \frac{s - \lambda_1}{\lambda_2} \cdot \frac{\lambda_1 \lambda_2}{s^2} \cdot \frac{1}{s^2} ds = \frac{\lambda_1}{\lambda_2} \int_{\lambda_1}^{\infty} \frac{s - \lambda_1}{s^2} ds \\ &= \frac{\lambda_1}{\lambda_2} \left[\int_{\lambda_1}^{\infty} \frac{1}{s} ds - \int_{\lambda_1}^{\infty} \frac{\lambda_1}{s^2} ds \right] = \frac{\lambda_1}{\lambda_2} \log s \Big|_{\lambda_1}^{\infty} + \frac{\lambda_1^2}{\lambda_2} \frac{1}{s} \Big|_{\lambda_1}^{\infty} = \infty \end{aligned}$$

Für $\mathbb{E}(X_1/X_2)$ erhalten wir die exakt analoge Rechnung und damit dasselbe Ergebnis.

Fazit: Zwei Personen mit den exponentialverteilten Wartezeiten X_1 und X_2 haben beide das Gefühl, in der falschen Schlange zu stehen.

4 Abschätzungen und Grenzwertsätze

Oftmals lässt sich die Wahrscheinlichkeit eines bestimmten Ereignisses nicht in geschlossener Form angeben bzw. die Berechnung ist sehr aufwendig. In diesem Fall helfen vielfach Ungleichungen, um Schranken für die Wahrscheinlichkeit zu bestimmen.

4.1 Ungleichungen

Satz 4.1 (Markov-Ungleichung): Sei X eine Zufallsvariable mit $\mathbb{E}|X| < \infty$. Dann gilt

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a} \mathbb{E}|X| \quad (4.1)$$

für alle reellen $a > 0$.

BEWEIS Sei X diskret mit der Wahrscheinlichkeitsfunktion p . Für $x \in \mathbb{R}$ mit $|x| \geq a$ gilt $\frac{|x|}{a} \geq 1$ und damit

$$\mathbb{P}(|X| \geq a) = \sum_{x:|x| \geq a} p(x) \leq \sum_{x:|x| \geq a} \frac{|x|}{a} p(x) \leq \sum_x \frac{|x|}{a} p(x) = \frac{1}{a} \sum_x |x| p(x) = \frac{1}{a} \mathbb{E}|X|.$$

Sei nun X stetig mit der Dichte f_X . Dann folgt mit analogen Argumenten zu oben

$$\begin{aligned} \mathbb{P}(|X| \geq a) &= \int_{-\infty}^{\infty} \mathbb{1}_{\{|x| \geq a\}} f_X(x) dx \leq \int_{-\infty}^{\infty} \frac{|x|}{a} \mathbb{1}_{\{|x| \geq a\}} f_X(x) dx \\ &\leq \int_{-\infty}^{\infty} \frac{|x|}{a} f_X(x) dx = \frac{1}{a} \int_{-\infty}^{\infty} |x| f_X(x) dx \\ &= \frac{1}{a} \mathbb{E}|X|. \quad \square \end{aligned}$$

Beispiel 4.2: Betrachte den n -fachen Wurf mit einer fairen Münze, sei X die Zufallsvariable für die Anzahl der Würfe mit dem Ausgang Kopf. Dann ist $X \sim \text{Bin}(n, 1/2)$. Die möglichen Ausgänge sind $X(\Omega) = \{0, 1, \dots, n\}$. Damit ist $|X| = X$ und wir kennen auch $\mathbb{E}X = n/2$.

Die Wahrscheinlichkeit, in mindestens 75% der Würfe Kopf zu erzielen, ist beschränkt durch

$$\mathbb{P}(X \geq \frac{3}{4} \cdot n) \leq \frac{\mathbb{E}X}{\frac{3}{4} \cdot n} = \frac{n \cdot \frac{1}{2}}{\frac{3}{4} \cdot n} = \frac{2}{3}$$

wobei die Markov-Ungleichung mit $a = \frac{3}{4} \cdot n$ verwendet wurde.

4 Abschätzungen und Grenzwertsätze

Satz 4.3 (Tschebyschev-Ungleichung): Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann gilt

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{1}{a^2} \mathbb{V}\text{ar}X \quad (4.2)$$

für alle reellen $a > 0$.

BEWEIS Wir betrachten die Zufallsvariable $|X - \mathbb{E}X|^2$ und wenden auf diese die Markov-Ungleichung (4.1) an. Dann ist

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2) \leq \frac{1}{a^2} \mathbb{E}(|X - \mathbb{E}X|^2) = \frac{1}{a^2} \mathbb{V}\text{ar}X. \quad \square$$

Anmerkung: Die Tschebyschev-Ungleichung gibt eine einfache Abschätzung für die Wahrscheinlichkeit an, dass die Zufallsvariable vom Erwartungswert abweicht.

Korollar 4.4: Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann gilt $\mathbb{V}\text{ar}X = 0$ genau dann, wenn $\mathbb{P}(X = \mathbb{E}X) = 1$.

BEWEIS Falls $\mathbb{V}\text{ar}X = 0$ gilt, so folgt aus der Tschebyschev-Ungleichung (4.2), dass

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = 0$$

für alle $\varepsilon > 0$ und damit

$$1 - \mathbb{P}(X = \mathbb{E}X) = \mathbb{P}(|X - \mathbb{E}X| > 0) = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = 0,$$

woraus über das Gegenereignis $\mathbb{P}(X = \mathbb{E}X) = 1$ folgt.

Andererseits ergibt sich aus $\mathbb{P}(X = \mathbb{E}X) = 1$, dass X diskret ist. Dann gilt

$$\begin{aligned} (\mathbb{E}X)^2 &= \left(\sum_x x \mathbb{P}(X = x) \right)^2 \\ &= \mathbb{E}X \cdot \mathbb{P}(X = \mathbb{E}X) \cdot \mathbb{E}X \cdot \mathbb{P}(X = \mathbb{E}X) \\ &= (\mathbb{E}X)^2 \mathbb{P}(X = \mathbb{E}X) = \sum_x x^2 \mathbb{P}(X = x) \\ &= \mathbb{E}X^2 \end{aligned}$$

wobei die Transformationsformel für Erwartungswerte (Satz 3.11) verwendet wurde. Nun haben wir

$$\mathbb{V}\text{ar}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0, \quad \square$$

also die Aussage.

Beispiel 4.5: Wir betrachten eine Messmethode, die fehlerhafte Ergebnisse liefert. Dabei sei $\mu \in \mathbb{R}$ zu messen, X_i der Messfehler und $Z_i := \mu + X_i$ das Messergebnis der i -ten Messung für $1 \leq i \leq n$. Wir gehen davon aus, dass X_1, \dots, X_n unabhängig und identisch verteilt sind (i.i.d. Folge).

Wir berechnen die Anzahl der erforderlichen Messungen, die notwendig sind, damit das arithmetische Mittel $Y_n := \frac{1}{n} \sum_{i=1}^n Z_i$ der zufälligen Messwerte mit einer Wahrscheinlichkeit von höchstens 0,1 um mehr als 1 von dem unbekanntem Wert μ abweicht.

(a) Wir wissen über die Verteilung der X_i nur, dass $\mathbb{E}X_i = 0$ und $\text{Var}X_i = 1$. Wegen der Linearität des Erwartungswertes und der Unabhängigkeit wissen wir, dass

$$\mathbb{E}Y_n = \mu \quad \text{und} \quad \text{Var}Y_n = \frac{1}{n}$$

und damit folgt aus der Tschebyschev-Ungleichung (4.2)

$$\mathbb{P}(|Y_n - \mu| > 1) \leq \frac{1}{n},$$

also $\mathbb{P}(|Y_n - \mu| > 1) \leq 0,1$, falls $1/n \leq 0,1$ ist, d. h. $n \geq 10$.

(b) Haben wir hingegen Information über die Verteilung der Fehler, so können wir diese nutzen. Seien hier die Messfehler $X_i \sim \mathcal{N}(0,1)$. Dann gilt

$$Y_n \sim \mathcal{N}(\mu, 1/n) \quad \text{und normiert} \quad \sqrt{n}(Y_n - \mu) \sim \mathcal{N}(0,1).$$

Dann gilt

$$\begin{aligned} \mathbb{P}(|Y_n - \mu| \geq 1) &= \mathbb{P}(\sqrt{n}|Y_n - \mu| \geq \sqrt{n}) \\ &= \mathbb{P}(\sqrt{n}(Y_n - \mu) \leq -\sqrt{n}) + \mathbb{P}(\sqrt{n}(Y_n - \mu) \geq \sqrt{n}) \\ &= \Phi(-\sqrt{n}) + (1 - \Phi(\sqrt{n})) = (1 - \Phi(\sqrt{n})) + (1 - \Phi(\sqrt{n})) \\ &= 2(1 - \Phi(\sqrt{n})), \end{aligned}$$

wobei wir die Punktsymmetrie der Verteilungsfunktion Φ der Normalverteilung verwendet haben. Es gilt

$$2(1 - \Phi(\sqrt{n})) \leq 0,1 \quad \iff \quad \Phi(\sqrt{n}) \geq 0,95 \quad \iff \quad \sqrt{n} \geq 1,645$$

d. h. $n \geq 3$.

Wir sehen hier, dass Verteilungsinformation zu einer viel höheren Genauigkeit führt.

Anmerkung: Sowohl die Markov- als auch die Tschebyschev-Ungleichung kommen ohne Verteilungsannahme aus. Das macht sie zu einem beliebten Instrument. Der Nachteil davon ist allerdings, dass die Abschätzungen, die sie liefern, relativ grob sind. Sind zusätzlich Annahmen über die Verteilung der Zufallsvariablen X möglich, so lassen sich genauere Abschätzungen erreichen.

4.2 Grenzwertsätze

Grenzwertsätze sind Aussagen, die es uns ermöglichen, Annäherungen zur Berechnung von Wahrscheinlichkeiten zu verwenden. Das ist immer dann sinnvoll, wenn die exakte Berechnung zu kompliziert ist. Der erkaufte Nachteil ist, dass wir eine gewisse Ungenauigkeit bekommen.

Betrachte eine $\text{Bin}(n, \eta)$ -verteilte Zufallsvariable. Die Berechnung der Binomialkoeffizienten ist nicht einfach, wenn n groß ist. Die Poisson-Approximation wird für große Werte von n und sehr kleine Werte von p angewendet, also seltene Ereignisse.

Satz 4.6 (Poisson-Grenzwertsatz): Es sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen, wobei $X_n \sim \text{Bin}(n, p_n)$ -verteilt ist. Falls ein $\lambda > 0$ existiert so dass $n p_n \rightarrow \lambda$ für $n \rightarrow \infty$, so gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (4.3)$$

also bekommen wir eine Poisson-Verteilung.

4 Abschätzungen und Grenzwertsätze

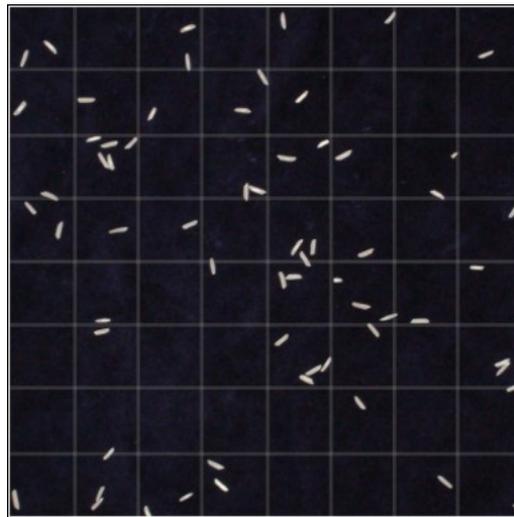


Abbildung 4.1: Reiskörner auf einem Schachbrett zur Darstellung der Poisson-Approximation aus dem Beispiel 4.7 (b).

Anmerkung: Zur Genauigkeit betrachten wir die Vergleichstabelle

Verteilung	Wahrscheinlichkeit				
	0	1	2	3	≥ 4
Bin(5, 0.3)	0.1681	0.3602	0.3087	0.1323	0.0308
Bin(50, 0.03)	0.2181	0.3372	0.2555	0.1264	0.0628
Bin(500, 0.003)	0.2226	0.3349	0.2515	0.1256	0.0654
Bin(5000, 0.0003)	0.2231	0.3347	0.2511	0.1255	0.0656
Poi(1.5)	0.2231	0.3347	0.2510	0.1255	0.0656

Beispiel 4.7 (Gesetz der kleinen Zahlen): (a) Betrachte eine Folge X_1, \dots, X_n identisch verteilter Zufallsvariablen, die voneinander unabhängig sind und n mögliche Ausgänge haben, d. h. $|X_1(\Omega)| = n$, die gleichverteilt angenommen werden. Dabei interessieren wir uns dafür, einen bestimmten Ausgang k -mal zu erzielen, sei Z_n die zugehörige Zufallsvariable, d. h. $Z_n(k) := \sum_{i=1}^n \mathbb{1}_{\{X_i=k\}}$. Dann ist $Z_n \sim \text{Bin}(n, 1/n)$ -verteilt. Mit dem Satz 4.6 gilt, dass Z_n ungefähr Poi(1)-verteilt ist.

(b) Wir beschäftigen wir uns mit einem Schachbrett und 64 Reiskörnern, die wir auf das Schachbrett werfen. Dazu untersuchen wir die Zufallsvariable Z_{64} aus dem Teil (a), die angibt, auf wie vielen Schachbrettfeldern jeweils die Anzahl k der Reiskörner zu liegen kommt. (Siehe hierzu Bild 4.1.)

k	Z_{64}	$Z_{64}/64$	Poi(1)
0	23	0.3594	0.3679
1	25	0.3906	0.3679
2	12	0.1875	0.1839
3	2	0.0312	0.0613
4	1	0.0156	0.0153
5	1	0.0156	0.0031
Σ	64	1	0.9994

(c) Wir gehen (virtuell) ins Casino und spielen Roulette. Hier sind 37 Ausgänge möglich. Betrachte immer einen Satz von 37 Spielen und schaue wie viele Ausgänge 0 mal, 1 mal, usw. vorkommen.

Wir simulieren 10000 solcher Sätze, bezeichne mit \bar{Z} das arithmetische Mittel der Zufallsvariablen Z aus den 10000 Läufen. Dann erhalten wir

k	\bar{Z}	$\bar{Z}/37$	Poi(1)
0	13.4364	0.3631	0.3679
1	13.7757	0.3723	0.3679
2	6.9008	0.1865	0.1839
3	2.3293	0.0630	0.0613
4	0.5310	0.0144	0.0153
5	0.0268	0.0007	0.0031
Σ	37	1	0.9994

Jedes Mal wird im Schnitt ein Drittel der möglichen Ausgänge nicht erreicht. Dies wird von Casino-Gängern auch das *Gesetz des Drittels* genannt. Allerdings ist damit *nicht* gesagt, dass es sich immer um die Gleichen handle. Im Gegensatz: Zählen wir alle Vorkommnisse zusammen, so bekommen wir

k	0	1	2	3	4	5	6
Y	9969	10021	9816	9921	10019	9833	9918

k	7	8	9	10	11	12
Y	9850	9914	10041	10097	10184	9935

k	13	14	15	16	17	18
Y	10026	10068	10283	10165	10169	10216

k	19	20	21	22	23	24
Y	10014	9943	9938	9932	9817	10159

k	25	26	27	28	29	30
Y	9879	10053	9904	10141	9912	10005

k	31	32	33	34	35	36
Y	10001	10049	9945	9830	10011	10022

Beachte dabei, dass wir insgesamt $10000 \cdot 37$ Roulette-Spiele simuliert haben. Da jeder Ausgang gleich wahrscheinlich ist, also $p = 1/37$ folgt für die $\text{Bin}(370000, 1/37)$ -verteilte Zufallsvariable Y , dass $\mathbb{E}Y = 370000 \cdot 1/37 = 10000$ ist.

4.2.1 Gesetze der großen Zahlen

Den Effekt der Mittelung, den wir im letzten Beispiel zum Schluss gesehen haben, wollen wir hier genauer untersuchen. Außerdem lernen wir im folgenden Beispiel eine neue Sichtweise multivariater Zufallsvariablen kennen.

Beispiel 4.8: Seien X_1, \dots, X_n für $n \geq 1$ Bernoulli-verteilte Zufallsvariablen mit dem Parameter η , die unabhängig sind, beispielsweise Münzwurf mit einer fairen Münze. Dann kann eine Folge Ω_n von Wahrscheinlichkeitsräumen definiert werden durch

$$\begin{aligned}\Omega_n &= \Omega^{(1)} \times \Omega^{(2)} \times \dots \times \Omega^{(n)} \\ &= \left\{ \omega : \omega = (\omega_1, \dots, \omega_n) \text{ mit } \omega_i \in \Omega^{(i)} = \{\text{Kopf, Zahl}\} \text{ für alle } 0 \leq i \leq n \right\}\end{aligned}$$

wobei $\Omega^{(i)} = \{\text{Kopf, Zahl}\}$ der Grundraum von X_i ist und somit $X_i(\Omega^{(i)}) = \{0, 1\}$. Desweiteren sei das Wahrscheinlichkeitsmaß \mathbb{P} gegeben durch

$$\mathbb{P} \left(\left\{ \omega : \omega = (\omega_1, \dots, \omega_n) \in \Omega \right\} \right) = \frac{1}{2^n}$$

als die Laplace-Wahrscheinlichkeit auf Ω_n . Die einzelnen Würfe X_1, \dots, X_n können als Projektionen

$$X_i(\omega) = (X(\omega))_i$$

für $1 \leq i \leq n$ aufgefasst werden.

Definiere die Zufallsvariable $Z_n := 1/n \sum_{i=1}^n X_i$ für das arithmetische Mittel von X_1, \dots, X_n beim n -maligen Werfen der Münze und betrachte das Ereignis

$$\left\{ \omega : Z_n(\omega) = (\mathbb{E}X_1) \right\}$$

d. h. die Verbindung des arithmetischen Mittels und des Erwartungswertes. Aus Erfahrung schließen wir, dass diese nah beieinander liegen.

Fragestellung Welcher Zusammenhang besteht allgemein zwischen dem arithmetischen Mittel und dem Erwartungswert?

Definition 4.9 (Stochastische Konvergenz): Seien $X_1, X_2 \dots$ und X Zufallsvariablen auf dem gleichen Wahrscheinlichkeitsraum. Wir sagen, dass X_n gegen X *stochastisch konvergiert*, geschrieben

$$X_n \xrightarrow{st} X,$$

falls

$$\lim_{n \rightarrow \infty} \mathbb{P} (|X_n - X| > \varepsilon) = 0 \tag{4.4}$$

für alle $\varepsilon > 0$.

Satz 4.10 (Schwaches Gesetz der großen Zahlen): Sei $X_1, X_2 \dots$ eine Folge von Zufallsvariablen mit dem gleichen Erwartungswert $\mu := \mathbb{E}X_i$ und mit $\mathbb{E}X_i^2 < \infty$ für alle $i \geq 1$ und sei $Z_n := 1/n \sum_{i=1}^n X_i$. Falls

$$\lim_{n \rightarrow \infty} \text{Var} Z_n = 0,$$

dann konvergiert die Folge $X_1, X_2 \dots$ stochastisch gegen μ , i. e.

$$\lim_{n \rightarrow \infty} \mathbb{P} (|Z_n - \mu| > \varepsilon) = 0 \tag{4.5}$$

für alle $\varepsilon > 0$.

BEWEIS Da $\mathbb{E}X_i < \infty$ für alle $i \geq 1$ gilt, haben wir

$$\mathbb{E}Z_n = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

Mit der Tschebyschevschen Ungleichung gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Z_n - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\text{Var}Z_n}{\varepsilon^2} = \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} \text{Var}Z_n = 0. \quad (4.6)$$

für alle $\varepsilon > 0$. □

Anmerkung: Die Bedingung des Satzes 4.10 sind insbesondere erfüllt, wenn die Zufallsvariablen X_1, X_2, \dots unabhängig sind und die gleiche Varianz σ^2 haben, da dann

$$\lim_{n \rightarrow \infty} \text{Var}Z_n = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var}X_i = \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0$$

gilt.

Definition 4.11 (Fast sichere Konvergenz): Seien X_1, X_2, \dots und X Zufallsvariablen auf dem gleichen Wahrscheinlichkeitsraum. Dann konvergiert X_n gegen X fast sicher, geschrieben

$$X_n \xrightarrow{f.s.} X,$$

falls

$$\mathbb{P}\left(\left\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1 \quad (4.7)$$

ist.

Satz 4.12 (Starkes Gesetz der großen Zahlen): Sei X_1, X_2, \dots eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $\mathbb{E}|X_1| < \infty$ und sei $\mu = \mathbb{E}X_1$. Sei weiter $Z_n := 1/n \sum_{i=1}^n X_i$ das arithmetische Mittel. Dann konvergiert die Folge X_1, X_2, \dots fast sicher gegen μ , i. e.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} Z_n = \mu\right) = 1. \quad (4.8)$$

Anmerkung (Unterschied zwischen den Konvergenzarten): Das schwache Gesetz sagt aus, dass für große n das arithmetische Mittel nah an μ ist. Deswegen gilt, dass $|Z_n(\omega) - \mu| > \varepsilon$ für beliebige ω unendlich oft möglich ist, aber in unregelmäßigen Abständen (und auch immer seltener); die Wahrscheinlichkeit wird immer kleiner.

Das starke Gesetz sagt aus, dass wir zu jedem $\varepsilon > 0$ ein N finden, das groß genug ist, so dass für alle $n \geq N$ die Schranke $|Z_n - \mu| < \varepsilon$ eingehalten wird. Dies gilt dann für fast alle ω ¹, so dass wir sagen können, dass es mit Wahrscheinlichkeit 1 gilt.

Anmerkung: Aus der fast sicheren Konvergenz folgt stochastische Konvergenz. Das Gegenteil gilt im Allgemeinen nicht.

Beispiel 4.13: Sei X_1, X_2, \dots eine Folge unabhängiger Zufallsvariablen, wobei $X_n \sim \text{Bin}(1, 1/n)$ -verteilt ist. Dann gilt

$$X_n \xrightarrow{st} 0.$$

¹Mit *fast alle* meinen wir, dass einzelne Werte z aus der Wertemengen von Z_n , für die $\mathbb{P}(Z_n = z) = 0$ gilt, aus der Reihe tanzen dürfen. Gedacht sei dabei z. B. an stetige Zufallsvariablen.

4 Abschätzungen und Grenzwertsätze

Andererseits nehme an, es gäbe ein N so, dass $\mathbb{P}(\{\omega : X_n(\omega) = 0 \text{ für alle } n \geq N\}) = 1$. Dann würde auch $X_n \xrightarrow{f.s.} 0$ gelten. Wir betrachten das obere Ereignis genauer

$$\mathbb{P}(\{\omega : X_n(\omega) = 0 \text{ für alle } n \geq N\}) = \frac{N-1}{N} \cdot \frac{N}{N+1} \cdot \frac{N+1}{N+2} \cdot \dots = 0,$$

d. h. es gibt *immer wieder* ein (zufälliges) i , für welches X_i nicht 0 annimmt.

Beispiel 4.14 (Monte-Carlo-Integration im Buffonschen Nadelexperiment): Hier wird das starke Gesetz der großen Zahlen zur numerischen Berechnung angewandt. Betrachte hierzu die Menge

$$K = \{(x, y) : x \in \mathbb{Z} \text{ und } y \in \mathbb{R}\} \subset \mathbb{R}^2$$

der parallelen vertikalen Geraden mit Abstand 1 in der euklidischen Ebene.

Werfe nun eine Nadel der Länge 1 zufällig in die Ebene, d. h. betrachte zwei Zufallsvariablen S und T zur Beschreibung der Lage der Nadel, wobei

- (a) S der senkrechte Abstand des Nadelmittelpunktes zur nächsten linken Geraden aus K ist,
- (b) T für den Winkel steht, den die Nadel zum Lot auf die Geraden aus K bildet,
- (c) S und T unabhängig sind,
- (d) und $S \sim \mathcal{U}[0, 1]$ und $T \sim \mathcal{U}[-\pi/2, \pi/2]$ gilt, d. h. wir kennen die Dichten $f_S(x) = \mathbb{1}_{[0,1]}(x)$ und $f_T(x) = \frac{1}{\pi} \mathbb{1}_{[-\pi/2, \pi/2]}(x)$.

Wir interessieren uns für die Wahrscheinlichkeit, dass die Nadel eine der Geraden aus K schneidet, also für das Ereignis

$$A = \{0 < S < 1/2 \cdot \cos T\} \cup \{1 - 1/2 \cdot \cos T < S < 1\}.$$

Dann gilt

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(0 < S < 1/2 \cdot \cos T) + \mathbb{P}(1 - 1/2 \cdot \cos T < S < 1) \\ &= \int_{-\pi/2}^{\pi/2} P(0 < S < 1/2 \cdot \cos t) f_T(t) dt + \int_{-\pi/2}^{\pi/2} P(1 - 1/2 \cdot \cos t < S < 1) f_T(t) dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \left(\int_0^{1/2 \cos t} f_S(s) ds \right) dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \left(\int_{1-1/2 \cos t}^1 f_S(s) ds \right) dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt \\ &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t dt = \frac{1}{\pi} \cdot \sin t \Big|_{-\pi/2}^{\pi/2} = \frac{2}{\pi}. \end{aligned}$$

Betrachte die Folge von unabhängigen Zufallsvektoren $(S_1, T_1), (S_2, T_2), \dots$ und definiere die Folge

$$X_n := \mathbb{1}_{\{S_n < 1/2 \cos T_n\}} + \mathbb{1}_{\{1 - 1/2 \cos T_n < S_n\}}$$

dann sind X_1, X_2, \dots unabhängig und identisch verteilt mit $\mathbb{E}X_1 = 2/\pi$.

Nach dem starken Gesetz der großen Zahlen gilt für das arithmetische Mittel $Z_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} 2/\pi$. Damit lässt sich für große n eine Näherung der Zahl π durch Simulation berechnen.

4.2.2 Zentraler Grenzwertsatz

Eine weitere Art der Konvergenz öffnet das Schloss zum zentralen Grenzwertsatz

Definition 4.15 (Konvergenz in Verteilung): Sei X_1, X_2, \dots eine Folge von Zufallsvariablen. Sei X eine weitere Zufallsvariable. Seien F_1, F_2, \dots und F ihre jeweiligen Verteilungsfunktionen. Dann konvergiert X_n in Verteilung gegen X , geschrieben

$$X_n \xrightarrow{D} X,$$

wenn

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

für alle $x \in \mathbb{R}$, an denen F stetig ist, gilt. Dabei steht D für *distribution*.

Anmerkung (Zusammenhang zwischen den Konvergenzarten): Für eine Folge X_1, X_2, \dots von Zufallsvariablen und die Zufallsvariable X gelten die Implikationen

$$\left(X_n \xrightarrow{f.s.} X \right) \implies \left(X_n \xrightarrow{st} X \right) \implies \left(X_n \xrightarrow{D} X \right)$$

Die Umkehrung gilt im Allgemeinen nicht.

Satz 4.16 (Satz von Moivre-Laplace): Sei $p \in (0, 1)$ und sei X_1, X_2, \dots eine Folge von Zufallsvariablen, die unabhängig und $\text{Bin}(1, p)$ -verteilt und auf dem gleichen Wahrscheinlichkeitsraum definiert sind. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x) \quad (4.9)$$

für alle $x \in \mathbb{R}$, d. h. die Folge $(Y_n)_{n \in \mathbb{N}}$ mit $Y_n := \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$ in Verteilung konvergiert; wir normieren Y_n und erhalten Konvergenz gegen die Standardnormalverteilung.

Beispiel 4.17: Binomialverteilungen mit steigendem n siehe R-Datei und -Ausgabe.

Beispiel 4.18: Betrachte Münzwurf mit einer fairen Münze, sagen wir 200-fachen. Wir fragen uns, wie groß die Wahrscheinlichkeit ist, dass wir mehr als 105-mal das Ereignis Kopf finden. Genauso finden wir die Wahrscheinlichkeit für weniger als 95 indem wir die Rolle von Kopf und Zahl vertauschen. Wir basteln eine Zufallsvariable

$$Z := \sum_{i=1}^{200} \mathbb{1}_{\{\text{i.ter Wurf} = \text{Kopf}\}}$$

und wissen $Z \sim \text{Bin}(200, 1/2)$. Damit suchen wir nach der Wahrscheinlichkeit

$$\mathbb{P}(Z > 105) = 1 - \mathbb{P}(Z \leq 105) = 1 - \mathbb{P}(Z - 100 \leq 5) = 1 - \mathbb{P} \left(\frac{Z - 100}{\sqrt{50}} \leq \frac{5}{\sqrt{50}} \right),$$

was der Form im Satz 4.16 entspricht. Somit ist

$$\mathbb{P}(Z > 105) \approx 1 - \Phi(5/\sqrt{50}) \approx 0.2398^2$$

²Zu beachten ist, dass die Verteilungsfunktion der Normalverteilung entweder in einem Tabellenwerk nachzusehen oder implementiert in einer Software (z. B. in R) auszulesen ist.

4 Abschätzungen und Grenzwertsätze

Satz 4.19 (Zentraler Grenzwertsatz): Sei X_1, X_2, \dots eine Folge von Zufallsvariablen, die unabhängig und auf dem gleichen Wahrscheinlichkeitsraum definiert sind. Weiterhin gelte $\mu := \mathbb{E}X_n$ und $\infty > \sigma^2 := \text{Var}X_n > 0$ für alle $n \geq 1$. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x) \quad (4.10)$$

für alle $x \in \mathbb{R}$.

Korollar 4.20: Nehme die Voraussetzung von Satz 4.19. Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < x \right) = \Phi(x)$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a).$$

Beispiel 4.21 (Messungen): Wir kommen zurück auf das Beispiel 4.5 und beschäftigen uns mit einer ungenauen Messmethode. Betrachte für $1 \leq i \leq n$ das Messergebnis der i -ten Messung in Form der Zufallsvariablen $Z_i := \mu + X_i$, wobei μ zu messen ist und X_i der Messfehler ist, σ die Varianz des Messfehlers. Alle Messfehler seien unabhängig und identisch verteilt. Die Wahrscheinlichkeit, dass das arithmetische Mittel Y_n der Messwerte Z_1, \dots, Z_n vom „echten Wert μ “ um mehr als ε abweicht, ergibt sich mit dem Satz 4.19

$$\begin{aligned} \mathbb{P} (|Y_n - \mu| > \varepsilon) &= \mathbb{P} \left(\left| \frac{Z_1 + \dots + Z_n - n\mu}{n} \right| > \varepsilon \right) \\ &= \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} \right| > \varepsilon \right) \\ &= \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \right| > \varepsilon \cdot \frac{\sqrt{n}}{\sigma} \right) \\ &= 1 - \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \right| \leq \varepsilon \cdot \frac{\sqrt{n}}{\sigma} \right) \\ &= 1 - \mathbb{P} \left(-\varepsilon \cdot \frac{\sqrt{n}}{\sigma} \leq \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \leq \varepsilon \cdot \frac{\sqrt{n}}{\sigma} \right) \\ &\approx 1 - \left(\Phi \left(\frac{\varepsilon\sqrt{n}}{\sigma} \right) - \Phi \left(-\frac{\varepsilon\sqrt{n}}{\sigma} \right) \right) \\ &= 2 \left(1 - \Phi \left(\frac{\varepsilon\sqrt{n}}{\sigma} \right) \right). \end{aligned}$$

Für eine genügende Anzahl n von Messungen benötigen wir keine Annahme der Normalverteilung für die Messfehler.

Beispiel 4.22: Exponential-verteilte Zufallsvariablen, siehe R-File und -Ausgabe.

5 Grundkonzepte der Statistik

Beschreibende Statistik Alltäglich wird der Begriff *Statistik* belebt mit der Darstellung bzw. *Beschreibung* von Ergebnissen zusammengestellter Daten und Fakten unterschiedlichster Art, wie z. B. politischen Umfragen, ökonomischen Kenngrößen, Marktforschungsumfragen, klinischen Studien, Einwohnerdaten, usw.

Das hängt mit dem Ursprung des Wortes Statistik zusammen, das vom lateinischen *statisticum*, den Staat betreffend, abstammt. Insofern war die amtliche Statistik die erste beschreibende Statistik.

Mathematische Statistik In der mathematischen Statistik wollen wir Eigenschaften und Gesetzmäßigkeiten von Datensätzen untersuchen.

Dabei werden aus mehr oder weniger großen *Stichproben* Aussagen über die *Grundgesamtheit* gewonnen. Die untersuchten Daten sind dabei einerseits *reale Daten*, wie sie sich aus der Beobachtung von z. B. Vorgängen bzw. Strukturen in der Natur, Technik oder Wirtschaft ergeben, oder andererseits *synthetische Daten*, die bei der Simulation solcher Vorgänge bzw. Strukturen durch Algorithmen erzeugt werden.

Die grundlegende Idee der Statistik ist die *stochastische Modellierung* der Daten und Fragestellungen. Im Gegensatz zur Wahrscheinlichkeitsrechnung, bei der wir den Wahrscheinlichkeitsraum gegeben hatten, z. B. durch Beschreibung und Verteilungen von Zufallsvariablen, wollen wir in der Statistik das zu Grunde liegende Wahrscheinlichkeitsmaß mit Hilfe von Stichproben ermitteln.

Um dieses Ziel effektiv angehen zu können, nutzen wir die Konzepte, welche wir in den vorangehenden Abschnitten kennengelernt haben.

5.1 Stichproben

Die genaueste und einfachste Antwort auf statistische Fragen erhält man, wenn man alle Elemente der Grundgesamtheit auf die interessierende Größe, das *Merkmal* untersucht, also eine *Vollerhebung* durchführt. Das ist jedoch meist nicht praktikabel und auch nicht nötig. Oft reicht es, aus einer Grundgesamtheit eine Stichprobe zu ziehen.

Im Hinblick auf die Aussagekraft der mathematischen Statistik, ist es erforderlich, dass diese Stichprobe hinsichtlich des Untersuchungsmerkmals möglichst repräsentativ ist und keinerlei Tendenzen aufzeigt.

So ist z. B. zur Untersuchung der Lebensverhältnisse von Personen mit einem Haushaltseinkommen von unter 1000 € die einseitige Befragung in einem Studentenwohnheim nicht geeignet.

5 Grundkonzepte der Statistik

Diese Eigenschaft der Repräsentativität ist keine, die die einzelne Stichprobe auszeichnet, denn da müsste man die Grundgesamtheit kennen, sondern eine Eigenschaft des Verfahrens zur Auswahl der Stichprobenelemente.

Zufallsstichproben

Angenommen, wir haben n Daten x_1, x_2, \dots, x_n . Dabei muss x_i nicht eine Zahl sein, sondern kann ein Vektor sein mit beispielsweise den Einträgen für Alter, Einkommen, Wohnort, Haushaltsgröße, usw. bei Einwohnerdaten. Wir behandeln hauptsächlich eindimensionale Merkmale, also $x_i \in \mathbb{R}$ für alle $1 \leq i \leq n$.

Wir nehmen nun an, dass die Daten eine Realisierung eines Zufallsvektors $X = (X_1, \dots, X_n)$ der Länge n sind. Dabei ist X ein Zufallsvektor auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, also

$$X = (X_1, \dots, X_n) : \Omega \mapsto \mathcal{X}.$$

Damit haben wir ein erstes stochastisches Modell gebastelt. Wie wir es untersuchen und eine Aussage über die Daten machen können, werden wir im Weiteren kennenlernen.

- Definition 5.1 (Stichprobe):** (a) $X = (X_1, \dots, X_n)$ heißt *Zufallsstichprobe*.
(b) Der Bildbereich \mathcal{X} heißt *Stichprobenraum*.
(c) Jede Realisierung $X(\omega) = (x_1, \dots, x_n)$ der Zufallsvariable X heißt (*konkrete*) *Stichprobe*.
(d) Die Dimension n von X wird *Stichprobenumfang* genannt.

Anmerkung: Oft sind X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariablen. Um von Varianz sprechen zu können, setzen wir $\mathbb{E}(X_i^2) < \infty$ für $1 \leq i \leq n$ voraus.

Anmerkung: Gegenüber der Wahrscheinlichkeitsrechnung ist jetzt in der Statistik neu, dass das zu Grunde liegende Wahrscheinlichkeitsmaß \mathbb{P} nicht gegeben ist. Vielmehr ist es Aufgabe der Statistik, das Wahrscheinlichkeitsmaß \mathbb{P} bzw. die Verteilung von X oder X_1 (bei identischer Verteilung und Unabhängigkeit) über ihre Verteilungsfunktion F zu bestimmen, *schätzen* genannt.

Anmerkung: Statt einer Wahrscheinlichkeitsverteilung ist meistens eine Familie \mathcal{P} von Wahrscheinlichkeitsverteilungen gegeben. Diese lässt sich oft parametrisieren, d. h. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, wobei $\Theta \subset \mathbb{R}^m$ der Parameterraum ist.

Beispiel 5.2: Wir betrachten die Aufgabe 1 auf dem Übungszettel 8. Ein Biologe geht von einer Bernoulliverteilung aus, dabei ist der Parameter $\eta \in [0, 5; 0, 8]$. Dann ist die Verteilungsfamilie $\mathcal{P} = \{\text{Bin}(1, \eta) : \eta \in [0, 5; 0, 8]\}$. Das Intervall $[0, 5; 0, 8]$ ist der Parameterraum.

Beispiel 5.3: Von einigen Schraubenschachteln haben sich die Aufkleber abgelöst. Man geht davon aus, dass die Schraubenbreite in jeder Schachtel normalverteilt ist und die Varianz bekannt und identisch ist, sagen wir σ^2 . In diesem Fall ist $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\}$ mit dem freien Parameter μ .

Anmerkung: Zur Beschreibung von Verteilungen werden charakteristische Eigenschaften der Lage und Form der Häufigkeitsverteilungen verwendet. Dazu gehören verschiedene Mittel, wie das arithmetische oder das geometrische Mittel, empirische Streuung, oder auch die Schiefe der Häufigkeitsverteilung, um nur einige zu nennen.

Um diesen Aufgaben zu begegnen, benötigen wir Funktionen auf dem Stichprobenraum, die uns die Bewertungen der Stichprobe liefern.

Definition 5.4 (Stichprobenfunktion und Statistik): Eine messbare Abbildung $\varphi : \mathcal{X} \mapsto \mathbb{R}^m$ heißt *Stichprobenfunktion*. Auf die Zufallsstichprobe $X = (X_1, \dots, X_n)$ angewendet, wird die Zufallsvariable $\varphi(X_1, \dots, X_n)$ betrachtet und eine *Statistik* genannt.

Anmerkung: In der Schätztheorie wird eine Statistik *Schätzer* genannt, bei statistischen Tests spricht man von *Teststatistik*.

Im Folgenden wollen wir uns zuerst zwei Beispiele von Stichprobenfunktionen ansehen, das Stichprobenmittel und die Stichprobenvarianz. Seien von nun an (falls nicht anders angegeben) X_1, \dots, X_n unabhängig, identisch verteilte Zufallsvariablen auf \mathbb{R} und $X = (X_1, \dots, X_n)$ die Zufallsstichprobe.

5.2 Beispiele für Stichprobenfunktionen

5.2.1 Stichprobenmittel

Zunächst beschäftigt uns die Frage nach der näherungsweise Bestimmung des Erwartungswertes $\mu = \mathbb{E}X_1$ der Stichprobenvariablen aus der Stichprobe. Sei $\sigma^2 = \mathbb{V}\text{ar}X_1$.

Definition 5.5 (Stichprobenmittel): Sei $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ mit $\varphi(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$ die Stichprobenfunktion, die einer Stichprobe ihr arithmetisches Mittel zuweist. Die Zufallsvariable

$$\bar{X}_n = \varphi(X) = \frac{1}{n} \sum_{i=1}^n X_i \quad (5.1)$$

heißt *Stichprobenmittel* der Zufallsstichprobe (X_1, \dots, X_n) .

Wegen der Unabhängigkeit wissen wir:

Satz 5.6: In Abhängigkeit des Erwartungswertes μ und der Varianz σ^2 der Stichprobenvariablen X_1 lassen sich

$$\mathbb{E}\bar{X}_n = \mu \quad (5.2)$$

und

$$\mathbb{V}\text{ar}\bar{X}_n = \frac{\sigma^2}{n}, \quad (5.3)$$

der Erwartungswert und die Varianz von \bar{X}_n , angeben.

BEWEIS Für die Gleichung (5.2) nutzen wir die Linearität des Erwartungswertes (Satz 3.13), für (5.3) den Satz 3.26. \square

Über die Verteilung von \bar{X}_n lassen sich aus dem zentralen Grenzwertsatz (Satz 4.19) und dem Gesetz der großen Zahlen (Satz 4.12) Schlüsse ziehen.

Satz 5.7: Für alle $z \in \mathbb{R}$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) = \Phi(z) \quad (5.4)$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Außerdem konvergiert \bar{X}_n fast sicher gegen μ .

5 Grundkonzepte der Statistik

Anmerkung: Die vorangehenden zwei Sätze sind der Grund dafür, das Stichprobenmittel als einen Schätzer für den Erwartungswert zu verwenden.

- (a) Wir machen keinen *systematischen Fehler*, z. B. durch Überschätzen oder Unterschätzen.
- (b) Über die *Schätzgenauigkeit* gibt uns die Varianz Auskunft. Wir können an dieser auch ablesen, dass die Schätzgenauigkeit mit wachsendem n verbessert wird.
- (c) Das Stichprobenmittel konvergiert fast sicher gegen den zu schätzenden Wert, d. h. es gibt keine Elemente des Wahrscheinlichkeitsraumes, die zu anderen Ergebnissen führen.
- (d) Wir kennen sogar die asymptotische Verteilung.

Bevor wir uns der Theorie der Schätzer zuwenden, in der wir auch die obigen Eigenschaften formalisieren, beschäftigen wir uns noch mit zwei weiteren Beispielen für Stichprobenfunktionen.

5.2.2 Stichprobenvarianz

Als nächstes widmen wir uns der Frage nach der Näherung der Varianz $\text{Var}X_1$ der Stichprobenvariablen aus der Stichprobe.

Definition 5.8 (Stichprobenvarianz): Wir betrachten die Stichprobenfunktion $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$ gegeben durch $\varphi(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$. Die Zufallsvariable

$$S_n^2 = \varphi(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (5.5)$$

heißt *Stichprobenvarianz* der Zufallsstichprobe (X_1, \dots, X_n) .

Wir gehen analog zum Stichprobenmittel vor und geben den Erwartungswert und die Varianz von S_n^2 an.

Satz 5.9: In Abhängigkeit des Erwartungswertes μ und der Varianz σ^2 der Stichprobenvariablen X_1 errechnen sich der Erwartungswert und die Varianz von S_n^2 als

$$\mathbb{E}S_n^2 = \sigma^2 \quad (5.6)$$

und

$$\text{Var}S_n^2 = \frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \sigma^4 \right) \quad (5.7)$$

wobei $\mu_4 = \mathbb{E}X_1^4$ das vierte Moment und $\sigma^4 = (\sigma^2)^2$ die quadrierte Varianz der Stichprobenvariablen X_1 bezeichnen.

Auch hier machen wir demnach keinen systematischen Fehler. Ebenfalls analog erhalten wir

Satz 5.10: Falls $\mathbb{E}X_1^4 < \infty$ ist, gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4} / \sqrt{n}} \leq z \right) = \Phi(z) \quad (5.8)$$

für alle $z \in \mathbb{R}$. Ebenfalls haben wir hier fast sichere Konvergenz, und zwar gegen σ^2 .

5.2.3 Empirische Verteilungsfunktion

Auch die Verteilungsfunktion F der Stichprobenvariablen X_1 lässt sich aus der Stichprobe schätzen. Der Fall ist nur insofern komplizierter, als wir für alle Werte $y \in \mathbb{R}$ eine Zufallsvariable definieren, d. h. eine Familie von Zufallsvariablen haben.

Definition 5.11 (Empirische Verteilungsfunktion): Wir betrachten für alle $y \in \mathbb{R}$ die Stichprobenfunktion $\varphi : \mathbb{R}^n \times \mathbb{R} \mapsto [0, 1]$ gegeben durch $\varphi(x_1, \dots, x_n)(y) = \frac{1}{n} \left| \{i : 1 \leq i \leq n \text{ mit } x_i \leq y\} \right|$, die die relative Häufigkeit der Stichprobenwerte angibt, welche den betrachteten Wert y nicht überschreiten. Wir erhalten die Familie $\{\hat{F}_n(y) : y \in \mathbb{R}\}$ von Zufallsvariablen $\hat{F}_n(y) : \Omega \mapsto [0, 1]$ mit

$$\hat{F}_n(y) = \varphi(X)(y) = \frac{1}{n} \left| \{i : 1 \leq i \leq n \text{ mit } X_i \leq y\} \right| \quad (5.9)$$

und nennen sie die *empirische Verteilungsfunktion* der Zufallsstichprobe (X_1, \dots, X_n) .

Auch bei der empirischen Verteilungsfunktion können wir Aussagen über Erwartungswert, Varianz und Verteilung treffen.

Satz 5.12: Für alle $y \in \mathbb{R}$ gelten (punktweise) folgende Eigenschaften:

(i) Die Zufallsvariable $n\hat{F}_n(y)$ ist binomialverteilt mit den Parametern n und $\eta = F(y)$, d. h. für $k = 0, 1, \dots, n$ gilt

$$\mathbb{P}(n\hat{F}_n(y) = k) = \binom{n}{k} (F(y))^k (1 - F(y))^{n-k}, \quad (5.10)$$

(ii) insbesondere erhalten wir damit den Erwartungswert

$$\mathbb{E}\hat{F}_n(y) = F(y) \quad (5.11)$$

und die Varianz

$$\text{Var}\hat{F}_n(y) = \frac{1}{n} F(y) (1 - F(y)). \quad (5.12)$$

(iii) Falls $0 < F(y) < 1$, so gilt für alle $z \in \mathbb{R}$ außerdem

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\hat{F}_n(y) - F(y)}{\sqrt{F(y) (1 - F(y))} / \sqrt{n}} \leq z \right) = \Phi(z). \quad (5.13)$$

Ein weiteres wichtiges Ergebnis ist die Konvergenz der ganzen Funktion, genannt *Satz von Glivenko-Cantelli*.

BEWEIS Die Zufallsvariable $n\hat{F}_n$ kann als die Anzahl der Erfolge beim n -maligen Münzwurf mit identischen Erfolgswahrscheinlichkeiten $\eta = F(x)$ aufgefasst werden. Daraus ergeben sich die Eigenschaften (i) und (ii), für (iii) wird dann der zentrale Grenzwertsatz für binomialverteilte Zufallsvariablen benötigt. \square

Wir sehen, dass auch hier kein systematischer Fehler gemacht wird.

5 Grundkonzepte der Statistik

6 Parameterschätzung

In diesem Kapitel wollen wir uns der systematischen Betrachtung der Parameterschätzung widmen. Dabei wollen wir zum Einen untersuchen, wie gut eine Schätzung an den zu schätzenden Wert heranreicht, zum Anderen wie wir geeignete Stichprobenfunktion zur Schätzung bestimmen.

6.1 Parametrisierung

Annahme Wir gehen von einer parametrischen Familie $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ aus, die *identifizierbar* ist, d. h. die Abbildung $\theta \mapsto P_\theta$ ist bijektiv, also aus $\theta_1 \neq \theta_2$ folgt, dass $P_{\theta_1} \neq P_{\theta_2}$, und umgekehrt. Sei dabei $\Theta \subset \mathbb{R}^m$ der Parameterraum und $\theta = (\theta_1, \dots, \theta_m)$ der Parameter(vektor) des Wahrscheinlichkeitsmaßes P_θ auf \mathbb{R} mit der Verteilungsfunktion F_θ .

Zusammenhang zwischen der Zufallsstichprobe und der Parametrisierung Sei $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe und seien X_1, \dots, X_n unabhängig, identisch verteilt (i.i.d.) auf \mathbb{R} mit der Verteilungsfunktion F_θ , die von dem unbekanntem Parameter θ abhängt. Das Wahrscheinlichkeitsmaß \mathbb{P} ist dann gegeben durch

$$\mathbb{P}(\{\omega : X_1(\omega) \leq y_1, \dots, X_n(\omega) \leq y_n\}) = F_\theta(y_1) \cdot \dots \cdot F_\theta(y_n) \quad (6.1)$$

für $y_1, \dots, y_n \in \mathbb{R}$. Damit haben wir auch sofort über Grenzwertbildung in $n - 1$ Dimensionen eine Aussage für die einzelnen Stichprobenvariablen X_i der Zufallsstichprobe

$$\lim_{\substack{y_j \rightarrow \infty \\ \text{für } j \neq i}} \mathbb{P}(\{\omega : X_1(\omega) \leq y_1, \dots, X_n(\omega) \leq y_n\}) = \mathbb{P}(\{\omega : X_i(\omega) \leq y_i\}) = F_\theta(y_i) \quad (6.2)$$

mit $y_i \in \mathbb{R}$ für alle $1 \leq i \leq n$.

Deswegen können wir, um zu betonen, dass \mathbb{P} kanonisch von dem Parameter θ abhängt, und sich des Unterschiedes zwischen P_θ und \mathbb{P} bewusst seiend¹, die Notation $\mathbb{P} = \mathbb{P}_\theta$ einführen, analog für den Erwartungswert und die Varianz bezüglich \mathbb{P}_θ die Bezeichnungen \mathbb{E}_θ resp. Var_θ .

Beispiel 6.1 (Parametrische Familien): (a) Eine wichtige parametrische Familie ist die Familie der Normalverteilungen

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\} = \left\{ \mathcal{N}(\mu, \sigma^2) : \theta = (\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0 \right\}.$$

Hier ist $\Theta = \mathbb{R} \times (0, \infty)$.

¹Den Unterschied zwischen mehrdimensionalen und eindimensionalen Wahrscheinlichkeitsräumen haben wir schon im Beispiel 4.8 kennengelernt.

6 Parameterschätzung

(b) Die Binomialverteilungen bilden ebenfalls eine parametrische Familie.

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\} = \{\text{Bin}(n, \eta) : \theta = (n, \eta), n \in \mathbb{N}, \eta \in (0, 1)\},$$

dann ist $\Theta = \mathbb{N} \times (0, 1)$.

(c) Weitere parametrische Familien sind beispielsweise durch die hypergeometrische Verteilung, die Exponentialverteilung, die Gleichverteilung, die negative Binomialverteilung, oder die Poisson-Verteilung gegeben.

6.2 Punktschätzung

Zur Erinnerung (an die Definition 5.4) Wir betrachten eine Stichprobenfunktion $\hat{\theta} : \mathcal{X} \rightarrow \Theta \subset \mathbb{R}^m$. Die zugehörige Zufallsvariable $\hat{\theta}(X_1, \dots, X_n)$ nennen wir Schätzer des Parametervektors θ .

6.2.1 Eigenschaften von Schätzern

Wir hatten uns am Beispiel des Stichprobenmittels sinnvolle Eigenschaften eines Schätzers überlegt.

Anmerkung: Ein Schätzer sollte demnach

(a) keinen systematischen Fehler machen, wir sagen *erwartungstreu* sein,

(b) eine gute Schätzgenauigkeit aufweisen, *effizient* sein,

(c) einen möglichst schnell kleiner werdenden zufälligen Fehler haben, d. h. Konvergenz gegen den zu schätzenden Wert besitzen, *konsistent* sein,

(d) eine bekannte Verteilung haben, z. B. die Normalverteilung, zumindest asymptotisch.

Diese Eigenschaft treten nicht immer auf, sind aber von Vorteil.

Mathematisch lassen sich die Eigenschaften auf folgende Weise definieren. Wir beschreiben nur den Fall $m = 1$ und weisen für $m > 1$ auf etwaige Unterschiede hin.

Definition 6.2 (Erwartungstreu und Bias): Ein Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für den Parameter θ , für den $\mathbb{E}_\theta |\hat{\theta}(X_1, \dots, X_n)| < \infty$ für alle $\theta \in \Theta$ gilt, heißt *erwartungstreu*, falls

$$\mathbb{E}_\theta \hat{\theta} = \theta \tag{6.3}$$

für alle $\theta \in \Theta$. Er heißt *asymptotisch erwartungstreu*, falls

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta \hat{\theta} = \theta \tag{6.4}$$

für alle $\theta \in \Theta$.

Der *Bias* (oder die *Verzerrung*) des Schätzers ist gegeben durch

$$\text{Bias}(\hat{\theta}) = \mathbb{E}_\theta \hat{\theta} - \theta, \tag{6.5}$$

strenggenommen $\text{Bias}_\theta(\hat{\theta})$.

Ist $\hat{\theta}$ erwartungstreu, so gilt $\text{Bias}(\hat{\theta}) = 0$.

Beispiel 6.3: Seien die Stichprobenvariablen X_i normalverteilt und sei

$$\mathcal{P} = \left\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \right\}$$

die parametrisierte Familie. Dann ist $\theta = (\mu, \sigma^2)$, also $m = 2$, und aus (5.2) und (5.6) folgt, dass der Schätzer $\hat{\theta} = (\bar{X}_n, S_n^2)$ für θ erwartungstreu ist.

Anmerkung: An dieser Stelle verstehen wir auch, wieso für die Varianz die Stichprobenvarianz S_n^2 dem kanonischer erscheinenden Schätzer $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ gegenüber bevorzugt verwendet wird.

Definition 6.4 (Erwartete mittlere quadratische Abweichung): Der *mittlere quadratische Fehler*, englisch *mean squared error*, eines Schätzers $\hat{\theta}(X_1, \dots, X_n)$ für θ , für den $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \dots, X_n)^2 \right) < \infty$ für alle $\theta \in \Theta \subset \mathbb{R}$ gilt, ist gegeben durch

$$\text{MSE}(\hat{\theta}) := \mathbb{E}_\theta \left(|\hat{\theta} - \theta|^2 \right) = \text{Var}_\theta(\hat{\theta}) + \text{Bias}_\theta(\hat{\theta})^2. \quad (6.6)$$

Ist $\hat{\theta}$ erwartungstreu, so gilt

$$\text{MSE}(\hat{\theta}) = \text{Var}_\theta(\hat{\theta}). \quad (6.7)$$

Definition 6.5: Seien $\hat{\theta}_1(X_1, \dots, X_n)$ und $\hat{\theta}_2(X_1, \dots, X_n)$ zwei Schätzer für θ , für die für alle $\theta \in \Theta \subset \mathbb{R}$ gilt $\mathbb{E}_\theta \left(\hat{\theta}_1(X_1, \dots, X_n)^2 \right) < \infty$ und $\mathbb{E}_\theta \left(\hat{\theta}_2(X_1, \dots, X_n)^2 \right) < \infty$. Dann sagt man, dass $\hat{\theta}_1$ *effizienter ist als* $\hat{\theta}_2$, falls

$$\text{MSE}(\hat{\theta}_1) \leq \text{MSE}(\hat{\theta}_2) \quad \text{für alle } \theta \in \Theta \quad (6.8)$$

gilt. Sind die Schätzer erwartungstreu, so ist der Schätzer $\hat{\theta}_1$ *effizienter als* $\hat{\theta}_2$, falls

$$\text{Var}_\theta(\hat{\theta}_1) \leq \text{Var}_\theta(\hat{\theta}_2) \quad \text{für alle } \theta \in \Theta \quad (6.9)$$

gilt.

Statt *effizienter* werden auch die Bezeichnungen *besser* oder *wirksamer* verwendet.

Definition 6.6 (Bester erwartungstreuer Schätzer): Ein erwartungstreuer Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ , für den $\mathbb{E}_\theta \left(\hat{\theta}(X_1, \dots, X_n)^2 \right) < \infty$ für alle $\theta \in \Theta$ gilt, ist der *beste erwartungstreue Schätzer* oder *effizienteste Schätzer*, falls er in der Klasse der erwartungstreuen Schätzer die kleinste erwartete mittlere quadratische Abweichung besitzt, d. h. falls für jeden erwartungstreuen Schätzer $\tilde{\theta}(X_1, \dots, X_n)$ gilt, dass

$$\text{Var}_\theta \hat{\theta} \leq \text{Var}_\theta \tilde{\theta} \quad \text{für alle } \theta \in \Theta \quad (6.10)$$

d. h. $\hat{\theta}$ hat in der Klasse aller erwartungstreuen Schätzer für θ die minimale Varianz.

Die Darstellung von MSE über die Summe aus Varianz und Bias gilt nur für $m = 1$, damit auch die Gleichungen (6.7), (6.9) und (6.10).

Definition 6.7 (Konsistenz): Ein Schätzer $\hat{\theta}(X_1, \dots, X_n)$ für θ heißt

(i) *schwach konsistent*, falls $\hat{\theta} \xrightarrow{\text{st}} \theta$ für $n \rightarrow \infty$, d. h. $\hat{\theta}$ stochastisch gegen θ konvergiert, d. h.

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta \left(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \epsilon \right) = 0 \quad \text{für alle } \theta \in \Theta. \quad (6.11)$$

6 Parameterschätzung

(ii) *stark konsistent*, falls $\hat{\theta} \xrightarrow{\text{f.s.}} \theta$ für $n \rightarrow \infty$, d. h. $\hat{\theta}$ fast sicher gegen θ konvergiert, d. h.

$$\mathbb{P}_{\theta} \left(\lim_{n \rightarrow \infty} \hat{\theta}(X_1, \dots, X_n) = \theta \right) = 1 \quad \text{für alle } \theta \in \Theta. \quad (6.12)$$

Definition 6.8 (Asymptotische Normalverteilung): Sei $\hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für $\theta \in \mathbb{R}$. Falls $0 < \text{Var}_{\theta} \hat{\theta} < \infty$ für alle $\theta \in \Theta$ und

$$\frac{\hat{\theta}(X_1, \dots, X_n) - \mathbb{E}_{\theta} \hat{\theta}(X_1, \dots, X_n)}{\sqrt{\text{Var}_{\theta} \hat{\theta}(X_1, \dots, X_n)}} \xrightarrow[n \rightarrow \infty]{\text{d}} Y \sim \mathcal{N}(0, 1) \quad (6.13)$$

dann ist $\hat{\theta}$ *asymptotisch normalverteilt*.

Da für $m > 1$ die Varianz in Wirklichkeit eine Kovarianzmatrix ist, muss für die Normierung der Nenner durch eine geeignete Matrixmultiplikation ersetzt werden.

Anmerkung: Das Stichprobenmittel und die Stichprobenvarianz sind nach (5.4) und (5.8) asymptotisch normalverteilt. Außerdem sind sie sowohl schwach als auch stark konsistent, was wir nicht gezeigt haben.

6.2.2 Methoden der Punktschätzung

Momentenmethode

Bevor wir uns der Konstruktion des Schätzers widmen, müssen wir noch einige Bezeichnungen einführen.

Definition 6.9 (Moment und Stichprobenmoment): Seien X, X_1, \dots, X_n i. i. d. Zufallsvariablen und gelte $\mathbb{E}|X_i|^k < \infty$ für ein $k \in \mathbb{N}$. Dann ist

$$\mu_k := \mathbb{E}(X^k) \quad (6.14)$$

das k -te Moment von X und

$$\hat{\mu}_k := \frac{1}{n} \sum_{i=1}^n X_i^k \quad (6.15)$$

das k -te Stichprobenmoment oder das k -te empirische Moment der Stichprobe $X = (X_1, \dots, X_n)$.

Anmerkung (Varianz und zweites Moment): Für die Varianz gilt

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

und damit

$$= \mu_2 - \mu_1^2.$$

Wir gehen von einer Zufallsstichprobe $X = (X_1, \dots, X_n)$ mit X_1, \dots, X_n i. i. d. aus.

Definition 6.10 (Momentenmethode-Schätzer): Es gelte $\mathbb{E}_\theta |X_i|^r < \infty$ für ein $r \geq m$ und die Momente μ_1, \dots, μ_r seien als Funktionen des Parametervektors $\theta = (\theta_1, \dots, \theta_m) \in \Theta$ gegeben, d. h. es existieren Borel-messbare Funktionen g_k so dass

$$\mu_k = g_k(\theta_1, \dots, \theta_m) \quad \text{für } 1 \leq k \leq r. \quad (6.16)$$

Falls das *Momenten-Gleichungssystem*

$$\hat{\mu}_k = g_k(\theta), \quad \text{für } 1 \leq k \leq r \quad (6.17)$$

eindeutig lösbar bezüglich θ ist, so heißt die Lösung

$$\hat{\theta}(X_1, \dots, X_n) \quad (6.18)$$

Momentenschätzer oder *M-Schätzer* von θ .

Beispiel 6.11 (Momentenschätzer bei der Normalverteilung): Seien X_1, \dots, X_n normalverteilt mit Erwartungswert μ und Varianz σ^2 , sei $\theta = (\mu, \sigma^2)$ der Parametervektor, für den wir einen M-Schätzer suchen. Für die Momente gelten folgende Gleichungen

$$\begin{aligned} g_1(\mu, \sigma^2) &= \mathbb{E}_\theta X_1 = \mu, \\ g_2(\mu, \sigma^2) &= \mathbb{E}_\theta X_1^2 = \text{Var}_\theta X_1 + (\mathbb{E}_\theta X_1)^2 = \sigma^2 + \mu^2. \end{aligned} \quad (6.19)$$

Damit ergibt sich das Gleichungssystem

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i &= \hat{\mu} \\ \frac{1}{n} \sum_{i=1}^n X_i^2 &= \hat{\sigma}^2 + \hat{\mu}^2 \end{aligned} \quad (6.20)$$

womit wir den Schätzer $(\hat{\mu}, \hat{\sigma}^2)$ mit

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \hat{\mu}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n-1}{n} S_n^2 \end{aligned} \quad (6.21)$$

erhalten. Wir wissen schon, dass $\hat{\mu}$ erwartungstreu ist. Weiterhin gilt

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

weswegen $\hat{\sigma}^2$ nicht erwartungstreu ist.

Anmerkung: (a) Die in Abschnitt 6.2.1 kennengelernten Eigenschaften gelten nicht allgemein, insbesondere sind die M-Schätzer nicht immer erwartungstreu.

(b) Die Momentenmethode liefert oft nicht den besten Schätzer, lässt sich aber im Allgemeinen gut berechnen und ist universell anwendbar.

Maximum-Likelihood Schätzer

Eine weitere Methode der Parameterschätzung ist die Maximum-Likelihood-Methode.

Anmerkung: Ziel ist es, die Parameter so zu bestimmen, dass eine möglichst gute Anpassung des Modells an die beobachteten Daten erreicht wird. Anders gesagt gehen wir davon aus, dass die beobachteten Daten typisch waren und maximieren bei der Maximum-Likelihood-Methode die Wahrscheinlichkeit für das Eintreten des beobachteten Ereignisses. Wir suchen den Parameter des parametrischen Modells, der die höchste Plausibilität für die beobachteten Daten liefert.

Im Folgenden beschränken wir uns auf die Fälle, bei denen die Stichprobenvariablen entweder diskret sind, dann existiert eine Zähldichte p_θ , oder absolut stetig, dann existiert eine Dichtefunktion f_θ .

Definition 6.12 (Likelihood-Funktion): (i) Seien die Stichprobenvariablen X_1, \dots, X_n diskret, sei $p_\theta(x)$ die Zähldichte und sei $x = (x_1, \dots, x_n) \in \mathcal{X}$ gegeben. Dann ist die *Likelihoodfunktion* $L_x : \Theta \rightarrow \mathbb{R}$ gegeben durch

$$L_x(\theta) := \mathbb{P}_\theta(x) = p_\theta(x_1) \cdot \dots \cdot p_\theta(x_n). \quad (6.22)$$

(ii) Seien die Stichprobenvariablen X_1, \dots, X_n absolut stetig, sei $f_\theta(x)$ die Dichtefunktion und sei die Stichprobe $x = (x_1, \dots, x_n) \in \mathcal{X}$ gegeben. Dann ist die *Likelihoodfunktion* $L_x : \Theta \rightarrow \mathbb{R}$ gegeben durch

$$L_x(\theta) := f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n). \quad (6.23)$$

Wir wollen den Wert der Likelihood-Funktion maximieren, um einen Parameter zu erhalten, so dass das Auftreten dieser Stichprobe möglichst plausibel ist.

Definition 6.13 (Maximum-Likelihood-Schätzer): Sei $\hat{\theta} : \mathcal{X} \rightarrow \Theta \subset \mathbb{R}^m$ eine Stichprobenfunktion mit

$$L_x(\hat{\theta}) = \max_{\theta \in \Theta} L_x(\theta) \quad \text{für alle } x \in \mathcal{X}. \quad (6.24)$$

Dann wird $\hat{\theta}(X)$ der *Maximum-Likelihood-Schätzer* oder *ML-Schätzer* von θ genannt.

Wir verdeutlichen das Prinzip an einem Beispiel.

Beispiel 6.14: Wir betrachten eine Warenlieferung von 12 Exemplaren eines Artikels, bei der wir den Hersteller nicht kennen. Wir wissen, dass nur drei potentiell mögliche Hersteller in Frage kommen. Deren Lieferungen weisen normalerweise folgende Ausschussanteile aus: $\theta_1 = 0,05$ (erster Hersteller), $\theta_2 = 0,10$ (zweiter Hersteller) und $\theta_3 = 0,15$ (dritter Hersteller).

Bei der Prüfung der Lieferung wird festgestellt, dass der Ausschuss genau ein Exemplar ist.

Wir wollen jetzt den vermutlichen Hersteller der vorliegenden Warenlieferung ermitteln.

Dazu betrachten wir die Stichprobe $X = (X_1, \dots, X_{12})$ mit

$$X_i(\omega) = \begin{cases} 1, & \text{falls das } i\text{-te Exemplar Ausschuss ist,} \\ 0 & \text{falls das } i\text{-te Exemplar in Ordnung ist.} \end{cases}$$

und die parametrische Familie $\mathcal{P} := \{\text{Bin}(1, \eta) : \eta \in \{\theta_1, \theta_2, \theta_3\}\}$ mit den drei Bernoulli-Verteilungen zu den Parametern θ_1, θ_2 und θ_3 .

Der Schätzer ist die Stichprobenfunktion $\hat{\theta} : \{0, 1\}^{12} \rightarrow \Theta$, die so gewählt ist, dass für alle Stichproben $x = (x_1, \dots, x_{12})$ mit $|\{i : x_i = 1\}| = 1$ die Likelihoodfunktion

$$L_x(\theta) = \mathbb{P}_\theta(\{X = x\}) = \theta(1 - \theta)^{11}$$

maximal ist.

Dazu betrachten wir die Tabelle

θ	$L_x(\theta)$
0,05	0,028
0,10	0,031
0,15	0,025

aus der folgt, dass das Maximum bei 0,031 liegt und somit $\hat{\theta}(x) = \theta_2$ ist, d. h. vermutlich war der zweite Hersteller (mit dem Ausschussanteil 0,10) der Lieferant.

Anmerkung: (a) Bei manchen parametrischen Verteilungsfamilien ist der Maximum-Likelihood-Schätzer nicht eindeutig bestimmt.

(b) Erster Ansatz für das Optimierungsproblem ist Differentiation, die bereits oft zum Erfolg führt.

(c) Häufig gibt es keine analytische Lösung. Dann werden numerische Verfahren, z. B. die Newton-Raphson Methode verwendet. Als Startpunkt der Iteration eignet sich der Momentenmethode-Schätzer.

Oft ist vorteilhafter, sich statt der Likelihoodfunktion die Loglikelihoodfunktion anzusehen.

Definition 6.15: Sei $x = (x_1, \dots, x_n) \in \mathcal{X}$ gegeben und $L_x : \Theta \rightarrow \mathbb{R}$ die Likelihoodfunktion. Dann ist $\log L_x(\theta)$ die Loglikelihoodfunktion der Stichprobe x .

Anmerkung: (a) Da der Logarithmus streng monoton ist, gilt

$$\operatorname{argmax} L_x(\theta) = \operatorname{argmax} \log L_x(\theta) \quad (6.25)$$

die Likelihoodfunktion und die Loglikelihoodfunktion nehmen ihr Maximum an selber Stelle an.

(b) Der Vorteil ist, dass das Produkt in der Likelihoodfunktion in eine Summe in der Loglikelihoodfunktion übergeführt wird, die oft einfacher handhabbar ist. Dies gilt insbesondere bei numerischen Betrachtungen.

Im Folgenden betrachten wir einige parametrische Familien und die entsprechenden ML-Schätzer.

Beispiel 6.16 (Bernoulli-Verteilung): Die parametrische Familie der Bernoulli-Verteilungen ist $\mathcal{P} = \{\operatorname{Bin}(1, \eta) : \eta \in [0, 1]\}$. Für die Zähldichte gilt

$$p_\eta(y) = \eta^y (1 - \eta)^{1-y} \quad \text{für } y \in \{0, 1\}$$

und die Likelihoodfunktion ist definiert als

$$L_x(\eta) = \prod_{i=1}^n \eta^{x_i} (1 - \eta)^{1-x_i} \quad \text{für } x = (x_1, \dots, x_n) \in \mathcal{X} = \{0, 1\}^n$$

Wir wollen jetzt das Maximum bestimmen:

(a) Falls $x = \underbrace{(0, \dots, 0)}_{n \text{ mal}}$, so hat $L_x(\eta)$ bei $\eta = 0$ ein Maximum.

(b) Analog ist bei $x = \underbrace{(1, \dots, 1)}_{n \text{ mal}}$ der Wert $\eta = 1$ das Maximum.

(c) Andernfalls können wir die Loglikelihoodfunktion betrachten,

$$\log L_x(\eta) = \left(\sum_{i=1}^n x_i \right) \log \eta + \left(n - \sum_{i=1}^n x_i \right) \log(1 - \eta)$$

die im Intervall $(0, 1)$ stetig ist mit den Grenzwerten

$$\lim_{\eta \rightarrow 0} \log L_x(\eta) = -\infty \quad \text{und} \quad \lim_{\eta \rightarrow 1} \log L_x(\eta) = -\infty,$$

6 Parameterschätzung

woraus folgt, dass $L_x(\eta)$ in dem Intervall $(0, 1)$ ein Maximum besitzt.

Um das Maximum zu berechnen, differenzieren wir nach η und erhalten die Gleichung

$$\frac{\partial}{\partial \eta} \log L_x(\eta) = \left(\sum_{i=1}^n x_i \right) \frac{1}{\eta} - \left(n - \sum_{i=1}^n x_i \right) \frac{1}{1-\eta} = 0,$$

die umgeformt

$$\eta n = (1 - \eta) \cdot \sum_{i=1}^n x_i + \eta \cdot \sum_{i=1}^n x_i$$

liefert, welche die eindeutig bestimmte Lösung

$$\hat{\eta}(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

besitzt.

Der ML-Schätzer für den Parameter η ist in allen drei Fällen gegeben durch

$$\hat{\eta}(X) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n.$$

Analog gehen wir bei der Binomialverteilung vor.

Beispiel 6.17 (Binomialverteilung): Sei $m \in \mathbb{N}$ gegeben. Die parametrische Familie der Binomialverteilungen mit einem Umfang von m ist $\mathcal{P} = \{\text{Bin}(m, \eta) : \eta \in [0, 1]\}$. Für die Zähldichte gilt

$$p_\eta(y) = \binom{m}{y} \eta^y (1-\eta)^{m-y} \quad \text{für } y \in \{0, 1, \dots, m\}$$

und die Likelihoodfunktion ist definiert als

$$L_x(\eta) = \prod_{i=1}^n \binom{m}{x_i} \eta^{x_i} (1-\eta)^{m-x_i} \quad \text{für } x = (x_1, \dots, x_n) \in \mathcal{X} = \{0, 1, \dots, m\}^n.$$

Nach ähnlichen Vorüberlegungen wie im Beispiel 6.16, differenzieren wir nach η und erhalten die Gleichung

$$\begin{aligned} \frac{\partial}{\partial \eta} \log L_x(\eta) &= \frac{\partial}{\partial \eta} \left(\sum_{i=1}^n \binom{m}{x_i} + \log \eta \cdot \sum_{i=1}^n x_i + \log(1-\eta) \left(mn - \sum_{i=1}^n x_i \right) \right) \\ &= \left(\sum_{i=1}^n x_i \right) \frac{1}{\eta} - \left(n \cdot m - \sum_{i=1}^n x_i \right) \frac{1}{1-\eta} = 0, \end{aligned}$$

und somit den ML-Schätzer

$$\hat{\eta}(X) = \frac{\frac{1}{n} \sum_{i=1}^n X_i}{m} = \frac{\bar{X}_n}{m}.$$

Beispiel 6.18 (Normalverteilung): Wir betrachten eine i.i.d. Stichprobe $X = (X_1, \dots, X_n)$ mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. Für die Dichtefunktion gilt

$$f_{\mu, \sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right) \quad \text{für } y \in \mathbb{R},$$

die Likelihoodfunktion ist definiert als

$$L_x(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \quad \text{für } x \in \mathbb{R}^n$$

und die Loglikelihoodfunktion ist

$$\log L_x(\mu, \sigma^2) = -n \log(\sqrt{2\pi\sigma}) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad \text{für } x \in \mathbb{R}^n$$

(a) Zuerst sei σ^2 bekannt und wir betrachten die parametrische Familie der Normalverteilungen $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\}$.

Die Abbildung $\mu \mapsto \log L_x(\mu, \sigma^2)$ ist stetig. Außerdem gilt

$$\lim_{\mu \rightarrow -\infty} \log L_x(\mu, \sigma^2) = -\infty \quad \text{und} \quad \lim_{\mu \rightarrow \infty} \log L_x(\mu, \sigma^2) = -\infty,$$

woraus folgt, dass $\mu \mapsto L_x(\mu, \sigma^2)$ in \mathbb{R} ein Maximum besitzt. Differentiation nach μ liefert die Gleichung

$$\frac{\partial}{\partial \mu} \log L_x(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0,$$

welche die Lösung $\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$ besitzt. Da $\frac{\partial^2}{\partial \mu^2} \log L_x(\mu, \sigma^2) = -\frac{n}{\sigma^2} < 0$ ist, ist die gefundene Lösung auch wirklich das Maximum, und zwar für jedes fest vorgegebene $\sigma^2 > 0$. Damit ist der ML-Schätzer $\hat{\mu} = \bar{X}_n$.

(b) Sei $\mu \in \mathbb{R}$ bekannt. Wir untersuchen die Loglikelihoodfunktion bezüglich der einparametrischen Familie der Normalverteilungen $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \sigma^2 > 0\}$.

Die Abbildung $\sigma^2 \mapsto \log L_x(\mu, \sigma^2)$ ist stetig für $\sigma^2 > 0$. Weiter gilt

$$\lim_{\sigma^2 \rightarrow 0} \log L_x(\mu, \sigma^2) = -\infty \quad \text{und} \quad \lim_{\sigma^2 \rightarrow \infty} \log L_x(\mu, \sigma^2) = -\infty,$$

woraus folgt, dass $\sigma^2 \mapsto L_x(\mu, \sigma^2)$ in $(0, \infty)$ ein Maximum besitzt. Differentiation nach σ^2 liefert die Gleichung

$$\begin{aligned} \frac{\partial}{\partial(\sigma^2)} \log L_x(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0, \\ \iff \sigma^2 \cdot n &= \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

welche die eindeutig bestimmte Lösung $\hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ besitzt und der ML-Schätzer ist

$$\hat{\sigma}^2(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

6 Parameterschätzung

(c) Als letztes betrachten wir die zweiparametrische Familie $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$.

In diesem Fall haben wir das Gleichungssystem

$$\begin{aligned}\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) &= 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 &= 0\end{aligned}$$

zu lösen. Die eindeutig bestimmte Lösung lautet $\hat{\mu} = \bar{x}_n$ und, da für absolut stetige Zufallsvariablen $\mathbb{P}(X_1 = \dots = X_n) = 0$ ist, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$.

Die berechneten ML-Schätzer für die Parameter μ und σ^2 sind

$$\hat{\mu}(X) = \bar{X}_n \quad \text{resp.} \quad \hat{\sigma}^2(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

In den bisher betrachteten Beispielen erhielten wir mittels der Maximum-Likelihood-Methode Schätzer, die wir bereits von früher kannten. Das liegt aber mehr an der Auswahl der Beispiele als an der Methode selbst.

Methode der kleinsten Quadrate

Die Methode der kleinsten Quadrate ist ein wichtiges Verfahren der Parameterschätzung, dem das folgende Szenario zu Grunde liegt:

Eine Zufallsvariable Y setzt sich aus einer Funktion h , welche von einer nicht zufälligen Größe x (z. B. auf \mathbb{R}) abhängig ist, und zufälligen Schwankungen Z zusammen

$$Y(x, \omega) := h(x) + Z(\omega) \quad \text{für alle } x \in \mathbb{R} \text{ und alle } \omega \in \Omega, \quad (6.26)$$

wobei Ω der Grundraum des Wahrscheinlichkeitsraums ist, auf dem Z definiert ist.

Annahmen Wir gehen von folgenden Bedingungen aus:

$$\begin{aligned}\mathbb{E}Z &= 0, & \text{d. h. } \mathbb{E}Y &= h, & (6.27) \\ \text{Var}Z &= \sigma^2 & (\text{unabhängig von } x) & & (6.28)\end{aligned}$$

Zu den Werten x_1, \dots, x_n werden die Werte $y_1 := Y_1(x_1, \omega_1), \dots, y_n := Y_n(x_n, \omega_n)$ beobachtet, die bereits die Schwankungen $Z(\omega_1), \dots, Z(\omega_n)$ enthalten. Wir definieren die Vektoren $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$.

Dabei ist zu beachten, dass im Gegensatz zu den bisherigen Modellannahmen, hier Y_i sowohl unterschiedliche Verteilungen haben (können) und im Allgemeinen auch nicht unabhängig sind.

Zu schätzen ist die Funktion h . Diese wird häufig als eine aus der Funktionenschar $\mathcal{F}_\Theta = \{f_\theta : \theta \in \Theta\}$ modelliert, weswegen wir es erneut mit einer Parameterschätzung zu tun haben. Deswegen betrachte h_θ .

Beispiel 6.19: Wir untersuchen die Umsatzentwicklung nach einer Werbemaßnahme. Das Unternehmen besitzt n Filialen und sei nun x der Betrag, der zur Werbung genutzt wurde, Y dann der Umsatz in einem bestimmten (nachfolgenden) Zeitabschnitt. Dann ist h_θ die Funktion für den mittleren Umsatz in Abhängigkeit von der Höhe der eingesetzten Werbung und F_θ ist z. B. die Menge aller Polynome mit Grad $\leq m$.

Vorgehen: Definiere die normierte Summe der quadratischen Abweichungen

$$Q(\tilde{h}) := \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{h}(x_i))^2 \quad (6.29)$$

für $\tilde{h} \in \mathcal{F}_\Theta$.

Kleinste-Quadrate Schätzer ist zu einer Stichprobe (x, y) die Funktion $\hat{h} := h_{\hat{\theta}} \in \mathcal{F}_\Theta$, für die $Q(\hat{h})$ minimal ist, also

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} Q(h_\theta) = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^n (y_i - h_\theta(x_i))^2 \quad (6.30)$$

Anmerkung: (a) Oft ist h linear. Dann ist der Kleinste-Quadrate-Schätzer der beste unter allen linearen erwartungstreuen Schätzern.

(b) Häufig werden die Z_i als normalverteilt modelliert, somit sind auch Y_i normalverteilt. Dann stimmen die Kleinste-Quadrate-Schätzungen mit der ML-Schätzung überein.

6.2.3 Einschub: Grundlagen der linearen Regression

An dieser Stelle betrachten wir die Idee, die der linearen Regression zu Grunde liegt. Obwohl wir uns dabei einerseits von der Parameterschätzung entfernen, ist der Exkurs andererseits passend, da wir die Kleinste-Quadrate-Schätzer hier direkt anwenden.

Die einfachste Art des Zusammenhangs zwischen zwei Zufallsvariablen ist durch eine Gerade $h(x) = \alpha + \beta x$ gegeben. Dann ist

$$Y = h(X) + Z = \alpha + \beta X + Z \quad (6.31)$$

und wir nennen Y die erklärte Variable, X die erklärende Variable.

Die Parameter α und β der Geraden sollen mit Hilfe der Methode der kleinsten Quadrate geschätzt werden.

Annahmen Wir gehen von einer verbundenen Stichprobe

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \left\{ \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \right\}$$

6 Parameterschätzung

aus, d. h. wir notieren uns die Paare (X_i, Y_i) also solche. Weiter ist $\binom{x}{y} = \left\{ \binom{x_1}{y_1}, \dots, \binom{x_1}{y_1} \right\}$ die konkrete verbundene Stichprobe.

Die Klasse der Geraden ist $\mathcal{F}_\Theta = \left\{ f_\theta : f_\theta(x) = \alpha x + \beta \text{ für } x \in \mathbb{R}, \theta = (\alpha, \beta) \in \Theta \subset \mathbb{R}^2 \right\}$

Vorgehen Die Summe der quadratischen Abweichungen ist dann

$$Q(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n \left(y_i - f_{(\alpha, \beta)}(x_i) \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - (\alpha + \beta x_i) \right)^2$$

Zur Lösung der Optimierungsaufgabe des Minimierens wird differenziert und wir erhalten die Gleichungen

$$\begin{aligned} \frac{\partial Q(\alpha, \beta)}{\partial \alpha} &= -\frac{2}{n} \sum_{i=1}^n (y_i - (\alpha + \beta x_i)) = 0 \\ \frac{\partial Q(\alpha, \beta)}{\partial \beta} &= -\frac{2}{n} \sum_{i=1}^n (y_i - (\alpha + \beta x_i)) x_i = 0 \end{aligned}$$

und damit

$$\begin{aligned} \hat{\alpha} &= \bar{y}_n - \hat{\beta} \bar{x}_n \\ \hat{\beta} &= \frac{\sum_{i=1}^n y_i x_i - \bar{y}_n \bar{x}_n}{\sum_{i=1}^n x_i^2 - \bar{x}_n^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \end{aligned}$$

Die Schätzer sind

$$\begin{aligned} \hat{\alpha}(X, Y) &= \bar{Y}_n - \hat{\beta} \bar{X}_n \\ \hat{\beta}(X, Y) &= \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \end{aligned}$$

Tatsächlich sehen wir hier einen Zusammenhang mit dem Korrelationskoeffizienten. Aus diesem Grund definieren wir

Definition 6.20: Sei $\binom{X_1}{Y_1}, \dots, \binom{X_1}{Y_1}$ eine verbundene Stichproben. Dann ist

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$$

der empirische oder Bravais-Pearson-Korrelationskoeffizient. Die empirische Kovarianz ist durch

$$\tilde{S}_{X,Y} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$$

und die empirischen Standardabweichungen sind durch

$$\tilde{S}_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} \quad \text{resp.} \quad \tilde{S}_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}$$

gegeben.

Dann gilt auch $r_{X,Y} = \frac{\tilde{S}_{X,Y}}{\tilde{S}_X \tilde{S}_Y}$ und für unsere Schätzer haben wir

$$\hat{\alpha}(X, Y) = \bar{Y}_n - \hat{\beta} \bar{X}_n, \quad (6.32)$$

$$\hat{\beta}(X, Y) = r_{X,Y} \cdot \frac{\tilde{S}_Y}{\tilde{S}_X}. \quad (6.33)$$

Beispiel 6.21: Aus dem Beispiel 6.19 betrachten wir einige Filialen.

Um die Güte eines Regressionsmodells zu beurteilen, bezeichnen wir den Fehler

$$\hat{Z}_i = Y_i - \hat{Y}_i$$

als *Residuum*, welches wir weiter untersuchen. Damit können wir die Streuung zerlegen. Wir sagen

$$\underbrace{SQT}_{\text{total sum of squares}} = \underbrace{SQE}_{\text{explained sum of squares}} + \underbrace{SQR}_{\text{residual sum of squares}} \quad (6.34)$$

wobei

$$SQT = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 = n \cdot \tilde{S}_{Y,n}$$

$$SQE = \sum_{i=1}^n (\hat{Y}_i - \bar{Y}_n)^2$$

$$SQR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Das *Bestimmtheitsmaß*

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}$$

beschreibt den Anteil der durch das Modell erklärten Varianz.

Lemma 6.22: Im Standardmodell ist

$$\hat{\sigma}^2(X, Y) := \frac{SQR}{n-2}$$

ein erwartungstreuer Schätzer für σ^2 .

Die KQ-Schätzer sind linear in Z_1, \dots, Z_n und damit können wir auch eine Aussage über ihre Verteilung machen. Dazu benötigen wir noch die t - und die F -Verteilungen, die wir im nächsten Abschnitt kennenlernen werden. Hier sei das Ergebnis schon einmal festgehalten.

Verteilung der standardisierten Schätzer Unter der Normalverteilungsannahme gilt weiter

$$\frac{\hat{\alpha}(X, Y) - \alpha}{\hat{\sigma}_{\hat{\alpha}}(X, Y)} \sim t_{n-2} \qquad \frac{\hat{\beta}(X, Y) - \beta}{\hat{\sigma}_{\hat{\beta}}(X, Y)} \sim t_{n-2}$$

6 Parameterschätzung

mit

$$\hat{\sigma}_{\hat{\alpha}}(X, Y) = \hat{\sigma}(X, Y) \frac{\sqrt{\sum_{i=1}^n X_i^2}}{\sqrt{n \sum_{i=1}^n (X_i - \bar{X}_n)^2}} \quad \hat{\sigma}_{\hat{\beta}}(X, Y) = \hat{\sigma}(X, Y) \frac{1}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}}$$

und es ist

$$\hat{F} = \frac{SQE}{SQR/n-2} \sim F_{k, n-(k+1)}. \quad (6.35)$$

6.3 Prüfverteilungen

Aus der Normalverteilung lassen sich weitere wichtige Verteilungen gewinnen. Da diese sowohl für die Intervallschätzung als auch später für die Testverfahren benötigt werden, seien sie hier eingefügt.

Bevor wir beginnen, definieren wir den Begriff der Quantilfunktion.

Definition 6.23: Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion. Dann ist die Umkehrabbildung $F^{-1} : [0, 1] \rightarrow \mathbb{R}$ mit

$$F^{-1}(y) = \inf \{x : F(x) \geq y\} \quad (6.36)$$

die *Quantilfunktion* von F . Sei weiter $\alpha \in (0, 1)$. Dann bezeichnet man die Zahl $F^{-1}(\alpha)$ als das α -Quantil von F .

Anmerkung: Die Quantilfunktion existiert, da die Verteilungsfunktion F monoton wachsend und rechtsstetig ist.

Beispiel 6.24 (Quantilfunktion der Normalverteilung): Bezeichne Φ die Verteilungsfunktion der Standardnormalverteilung. Dann wird das α -Quantil mit z_α für alle $\alpha \in (0, 1)$ notiert. Es gilt $\Phi(z_\alpha) = \alpha$.

Wegen der Symmetrieeigenschaft der Standardnormalverteilung ist

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

für alle $x \in \mathbb{R}$ und somit

$$z_\alpha = -z_{1-\alpha} \quad (6.37)$$

für alle $\alpha \in (0, 1)$.

6.3.1 χ^2 -Verteilung

Definition 6.25: Seien X_1, \dots, X_n unabhängige und identisch $\mathcal{N}(0, 1)$ verteilte Zufallsvariablen. Dann wird die Verteilung der Summe ihrer Quadrate

$$Z_n = \sum_{i=1}^n X_i^2$$

als *Chi-Quadrat-verteilt mit n Freiheitsgraden* bezeichnet, kurz χ_n^2 -verteilt, $Z_n \sim \chi_n^2$.

Anmerkung: Eine χ^2 -verteilte Zufallsvariable ist nicht symmetrisch. Für kleine n ist die Dichte deutlich linksteil. Für wachsendes n nähert sich die χ^2 -Verteilung als Folge des zentralen Grenzwertsatzes der Normalverteilung an.

Satz 6.26: Sei Z_n eine χ^2 -verteilte Zufallsvariable mit n Freiheitsgraden, sei F_{Z_n} ihre Verteilungsfunktion. Dann ist

$$\begin{aligned}\mathbb{E}Z_n &= n, \\ \text{Var}Z_n &= 2n,\end{aligned}$$

außerdem gilt für große $n > 30$ und alle $\alpha \in (0, 1)$ die folgende Approximation der Quantilfunktion

$$F_{Z_n}^{-1}(\alpha) = \frac{1}{2} \left(z_\alpha + \sqrt{2n-1} \right)^2,$$

wobei z_α die Quantilfunktion $\mathcal{N}(0, 1)$ -Verteilung ist.

Satz 6.27 (Additionssatz): Die Summe zweier unabhängiger Zufallsvariablen, die χ_n^2 -verteilt resp. χ_m^2 -verteilt sind, ist χ_{n+m}^2 -verteilt.

6.3.2 t-Verteilung

Definition 6.28: Seien X eine $\mathcal{N}(0, 1)$ -verteilte und Z_n eine von X unabhängige χ_n^2 -verteilte Zufallsvariable. Dann wird die Verteilung der Zufallsvariablen

$$T := \frac{X}{\sqrt{\frac{1}{n}Z_n}}$$

als *t-Verteilung mit n Freiheitsgraden* bezeichnet, $T \sim t_n$.

Satz 6.29: Sei T eine *t-verteilte* Zufallsvariable mit n Freiheitsgraden, sei F_T ihre Verteilungsfunktion. Dann ist

$$\begin{aligned}\mathbb{E}T &= 0 \\ \text{Var}T &= \frac{n}{n-2}\end{aligned}$$

für $n \geq 3$ und es gilt für alle $\alpha \in (0, 1)$

$$F_T^{-1}(\alpha) = 1 - F_T^{-1}(1 - \alpha),$$

was der Symmetrie der Verteilung geschuldet ist.

6.3.3 F-Verteilung

Definition 6.30: Seien X und Y zwei unabhängige Zufallsvariablen, die χ_m^2 -verteilt resp. χ_n^2 -verteilt sind. Dann wird die Verteilung der Zufallsvariablen

$$Z := \frac{\frac{1}{m}X}{\frac{1}{n}Y}$$

als *F-Verteilung mit den Freiheitsgraden m und n* bezeichnet, $Z \sim F_{m,n}$.

Satz 6.31: Sei Z eine F-verteilte Zufallsvariable mit den Freiheitsgraden m und n . Dann gilt

$$\mathbb{E}Z = \frac{n}{n-2},$$

$$\text{Var}Z = \frac{2n^2(n+m-2)}{m(n-4)(n-2)^2}.$$

Falls $F_{m,n}$ die Verteilungsfunktion von Z ist und $F_{n,m}$ die Verteilungsfunktion einer F-verteilten Zufallsvariablen mit den Freiheitsgraden n und m , so lassen sich die Quantilfunktionen über

$$F_{m,n}^{-1}(\alpha) = \frac{1}{F_{n,m}^{-1}(1-\alpha)}$$

für alle $0 < \alpha < 1$ miteinander in Bezug setzen.

Das wichtigste Beispiel zur Verwendung der Prüfverteilungen sei an dieser Stelle als Satz formuliert.

Satz 6.32: Seien X_1, \dots, X_n unabhängige und identisch $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariablen, z. B. eine Zufallsstichprobe. Dann sind

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

voneinander stochastisch unabhängig und es gilt

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

BEWEIS Wir zeigen nur die Verteilungsaussage. Der erste Teil gilt wegen der Linearität des Erwartungswertes und der Unabhängigkeit der Zufallsvariablen, da

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

Der zweite Teil folgt aus folgender Überlegung: Es gilt

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 + \left(\frac{1}{\sigma/\sqrt{n}}(\bar{X}_n - \mu)\right)^2 \quad (6.38)$$

nach Teilung der Identität

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n \left((X_i - \bar{X}_n) + (\bar{X}_n - \mu)\right)^2 \\ &= \sum_{i=1}^n \left((X_i - \bar{X}_n)^2 + 2(X_i - \bar{X}_n)(\bar{X}_n - \mu) + (\bar{X}_n - \mu)^2\right) \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + 2(\bar{X}_n - \mu) \sum_{i=1}^n (X_i - \bar{X}_n) + n(\bar{X}_n - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2 \end{aligned}$$

durch σ^2 . Der Term auf der linken Seite von (6.38) ist als Summe von n Quadraten $\mathcal{N}(0, 1)$ verteilter Zufallsvariablen χ_n^2 -verteilt. Der zweite Summand des rechten Terms ist aus dem gleichen Grund χ_1^2 -verteilt. Nun gilt, dass $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ wegen des Additionssatzes 6.27 χ_{n-1}^2 -verteilt ist. \square

6.4 Intervallschätzung

Der Punktschätzer $\hat{\theta}$ eines Parameters, ist normalerweise nicht mit diesem identisch. Zu dem Schätzer selbst ist es notwendig zu wissen, welche Präzision das Schätzverfahren mitbringt. Bei erwartungstreuen Schätzern ist die Varianz des Schätzers ein Maß für die Präzision.

Einen anderen Ansatz verfolgt die Intervallschätzung. Statt eines einzelnen Wertes ist das Ergebnis der Schätzung ein Intervall. Dabei wird die Wahrscheinlichkeit, dass das Intervall den Parameter nicht enthält kontrolliert.

Anmerkung (Annahmen in diesem Abschnitt): Wir gehen von einer parametrischen Verteilungsfamilie $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ mit den Verteilungsfunktionen F_θ aus. Wir wollen vorerst nur die j -te Komponente des Parameters $\theta = (\theta_1, \dots, \theta_m)$ bestimmen.

Für die Intervallschätzung benötigen wir zwei Stichprobenfunktionen, $\theta_u(X_1, \dots, X_n)$ für die untere und $\theta_o(X_1, \dots, X_n)$ für die obere Intervallgrenze.

Definition 6.33 (Konfidenzintervall): Sei $\alpha \in (0, 1)$ beliebig, aber fest gewählt. Dann heißt das zufällige Intervall $(\theta_u(X_1, \dots, X_n), \theta_o(X_1, \dots, X_n))$ ein *Konfidenzintervall zum Niveau $(1 - \alpha)$* für die Parameterkomponente θ_j , falls

$$\mathbb{P}_\theta (\theta_u(X_1, \dots, X_n) \leq \theta_o(X_1, \dots, X_n)) = 1, \quad (6.39)$$

$$\mathbb{P}_\theta (\theta_u(X_1, \dots, X_n) \leq \theta_j \leq \theta_o(X_1, \dots, X_n)) \geq 1 - \alpha \quad (6.40)$$

gelten. Alternativ zu (6.40) kann die Bedingung

$$\inf_{\theta \in \Theta} \mathbb{P}_\theta (\theta_u(X_1, \dots, X_n) \leq \theta_j \leq \theta_o(X_1, \dots, X_n)) = 1 - \alpha \quad (6.41)$$

oder

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (\theta_u(X_1, \dots, X_n) \leq \theta_j \leq \theta_o(X_1, \dots, X_n)) \geq 1 - \alpha \quad (6.42)$$

betrachtet werden, wobei man im letzteren Fall vom *asymptotischen Konfidenzintervall* spricht.

Damit haben wir ein zweiseitiges Konfidenzintervall.

Anmerkung (Irrtumswahrscheinlichkeit): Das Konfidenzniveau $1 - \alpha$ bestimmt, mit welcher Wahrscheinlichkeit ein Intervall entsteht, das den unbekanntem Parameter θ_j enthält. Umgekehrt ist α die Wahrscheinlichkeit ein Intervall zu enthalten, das θ_j nicht enthält. Deswegen wird α auch *Irrtumswahrscheinlichkeit* genannt.

Anmerkung (Vorgehen): Zur Bestimmung des Konfidenzintervalls

- (i) gibt man sich ein α vor,
- (ii) wählt man zwei Stichprobenfunktionen θ_u und θ_o , die die Bedingung (6.39) und eine der Bedingungen (6.40)–(6.42) erfüllen,
- (iii) berechnet die Werte $\theta_u(X_1, \dots, X_n)$ und $\theta_o(X_1, \dots, X_n)$

Anmerkung: Wir werden uns im folgenden auf symmetrische Konfidenzintervalle beschränken. Damit ist die Wahrscheinlichkeit, dass θ unterhalb der unteren Grenze θ_u liegt, dieselbe wie die Wahrscheinlichkeit, dass θ oberhalb der oberen Grenze θ_o liegt.

6 Parameterschätzung

Anmerkung: Will man die Abweichung der Schätzung vom wahren Wert nur in eine Richtung kontrollieren, so kann man eine der Grenzen durch $-\infty$ bzw. ∞ ersetzen. Dann erhält man ein einseitiges Konfidenzintervall. Das weitere Vorgehen ist analog.

Ob ein zweiseitiges oder ein einseitiges Konfidenzintervall angebracht ist, hängt von der Fragestellung ab.

Beispiel 6.34: (a) These: Frauen sind intelligenter als Männer. Hier ist ein einseitiges Konfidenzintervall zu verwenden.

(b) These: Frauen sind genauso intelligent wie Männer. Hier wird ein zweiseitiges Konfidenzintervall berechnet.

6.4.1 Normalverteilungsannahme

Konfidenzintervall für den Erwartungswert μ bei bekannter Varianz σ^2

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ unbekannt und σ^2 bekannt ist. Dann besitzt $\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$ eine $\mathcal{N}(0, 1)$ -Verteilung.

Sei α die vorgegebene Irrtumswahrscheinlichkeit. Dann ergibt sich, dass

$$\mathbb{P} \left(z_{\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq z_{1-\alpha/2} \right) = 1 - \alpha. \quad (6.43)$$

und mit (6.37) folgt

$$\mathbb{P} \left(-z_{1-\alpha/2} \leq \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \leq z_{1-\alpha/2} \right) = 1 - \alpha,$$

was umgeformt wird

$$\mathbb{P} \left(-z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha,$$

und weiter zu

$$\mathbb{P} \left(\bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha.$$

Damit sind mit den Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{und} \quad \theta_o(X_1, \dots, X_n) = \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (6.44)$$

die Gleichungen (6.41) und (6.43) äquivalent. Da die Standardabweichung bekannt ist, sind θ_u und θ_o konkret berechenbar.

Die Länge des Zufallsintervalls $[\theta_u, \theta_o]$ ist

$$L(X_1, \dots, X_n) = \theta_o(X_1, \dots, X_n) - \theta_u(X_1, \dots, X_n) = 2 z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (6.45)$$

und damit nicht zufällig. Zu einer vorgegebenen Maximallänge l ist dann

$$n = \left\lceil \left(\frac{2\sigma z_{1-\alpha/2}}{l} \right)^2 \right\rceil \quad (6.46)$$

der benötigte Stichprobenumfang bei festem σ^2 und fester Irrtumswahrscheinlichkeit α .

Beispiel 6.35: Mittels einer Füllmaschine werden Flaschen mit Getränken befüllt, nehmen wir an, zuerst mit Sirup, wobei das vorgegebene Volumen μ beträgt, welches stufenlos eingestellt werden kann. Dabei ist die Mengengenauigkeit unabhängig von dem eingestellten Volumen. Die eingefüllte Flüssigkeitsmenge wird als normalverteilt mit dem Erwartungswert μ und der Standardabweichung $\sigma = 0,24$ (in ml) angesehen.

Es werden $n = 9$ Flaschen der laufenden Produktion entnommen und das Volumen ihres Inhaltes nachgemessen. Die Werte sind

$$18,42 \quad 18,26 \quad 18,53 \quad 18,45 \quad 18,62 \quad 18,39 \quad 18,50 \quad 18,71 \quad 18,49$$

Zur Irrtumswahrscheinlichkeit $\alpha = 0,01$ soll eine Intervall-Schätzung durchgeführt werden. Aus den Gleichungen (6.44) ergibt sich die Realisierung $[18,27; 18,69]$ des Zufallsintervalls. Dieses hat die Länge 0,41.

Um eine Länge von höchstens 0,30 zu erhalten, benötigen wir nach (6.46) eine Stichprobe der Länge mindestens 17.

Konfidenzintervall für den Erwartungswert μ bei unbekannter Varianz σ^2

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 beide unbekannt sind. Da wir jetzt σ nicht kennen, können wir das Konfidenzintervall aus (6.44) nicht verwenden.

Wir nutzen deswegen aus, dass

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

Bezeichne mit $t_{n-1,c}$ das c -Quantil der t -Verteilung. Dann gilt

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

und wegen der Symmetrie

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

was umgeformt

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

6 Parameterschätzung

ergibt. Die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \bar{X}_n - t_{n-1, 1-\alpha/2} \frac{S_n}{\sqrt{n}} \quad \text{und} \quad \theta_o(X_1, \dots, X_n) = \bar{X}_n + t_{n-1, 1-\alpha/2} \frac{S_n}{\sqrt{n}} \quad (6.47)$$

liefern ein symmetrisches Konfidenzintervall zum Niveau $1 - \alpha$.

Im Gegensatz zu (6.45) gilt hier für die Länge

$$L(X_1, \dots, X_n) = \theta_o(X_1, \dots, X_n) - \theta_u(X_1, \dots, X_n) = 2 t_{n-1, 1-\alpha/2} \frac{S_n}{\sqrt{n}}$$

nicht, dass sie vom Zufall unabhängig wäre. Deswegen haben die Realisierungen des Konfidenzintervalls unterschiedliche Längen.

Konfidenzintervall für die Varianz σ^2 bei bekanntem Erwartungswert μ

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ bekannt und σ^2 diesmal unbekannt sind.

Wir wollen σ^2 schätzen. Dazu verwenden wir die Zufallsvariable

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

mit der Eigenschaft

$$\frac{n}{\sigma^2} \bar{S}_n^2 \sim \chi_n^2$$

und können das asymmetrische Konfidenzintervall mit $\alpha_1 + \alpha_2 = \alpha \in (0, 1)$ ansetzen mit

$$\mathbb{P} \left(\chi_{n, \alpha_2}^2 \leq \frac{n}{\sigma^2} \bar{S}_n^2 \leq \chi_{n, 1-\alpha_1}^2 \right) = 1 - \alpha \quad (6.48)$$

und erhalten die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \frac{n \bar{S}_n^2}{\chi_{n, 1-\alpha_1}^2} \quad \text{und} \quad \theta_o(X_1, \dots, X_n) = \frac{n \bar{S}_n^2}{\chi_{n, \alpha_2}^2}. \quad (6.49)$$

Konfidenzintervall für die Varianz σ^2 bei unbekanntem Erwartungswert μ

Sei $X = (X_1, \dots, X_n)$ die Zufallsstichprobe mit iid Stichprobenvariablen und sei $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind.

Wir wollen wieder σ^2 schätzen. Wir wissen, dass

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

ist und deswegen können wir für das asymmetrische Konfidenzintervall mit $\alpha_1 + \alpha_2 = \alpha \in (0, 1)$ die Gleichung

$$\mathbb{P} \left(\chi_{n-1, \alpha_2}^2 \leq \frac{n-1}{\sigma^2} S_n^2 \leq \chi_{n-1, 1-\alpha_1}^2 \right) = 1 - \alpha \quad (6.50)$$

ansetzen, wobei χ_{n-1, α_2}^2 das α_2 -Quantil der χ^2 -Verteilung mit $n-1$ Freiheitsgraden bezeichnet. Damit ergeben sich die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) = \frac{(n-1)S_n^2}{\chi_{n-1, 1-\alpha_1}^2} \quad \text{und} \quad \theta_o(X_1, \dots, X_n) = \frac{(n-1)S_n^2}{\chi_{n-1, \alpha_2}^2}. \quad (6.51)$$

Wählen wir $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, so erhalten wir wieder ein symmetrisches Konfidenzintervall.

Zwei-Stichproben-Probleme

Wir können das Szenario verallgemeinern, wenn wir gleichzeitig zwei Stichproben $X_1 = (X_{11}, \dots, X_{1n_1})$ und $X_2 = (X_{21}, \dots, X_{2n_2})$ betrachten.

Annahmen Generell gibt es da zwei mögliche Fälle

(a) Die Stichproben X_1 und X_2 sind voneinander unabhängig. Dann müssen die Längen n_1, n_2 nicht gleich sein.

(b) Die Stichproben hängen voneinander ab. Dann gilt $n_1 = n_2$ und wir definieren $X_i := (X_{1i}, X_{2i})$, wobei X_i i. i. d. sind.

Es stellt sich die Frage, wie wir die Definition 6.33 für das neue Szenario erweitern können: Wir gehen davon aus, dass X_{11}, \dots, X_{1n_1} unabhängig und identisch verteilt sind, das Gleiche gilt für X_{21}, \dots, X_{2n_2} .

Zu schätzen ist ein Funktionswert $g(\theta)$ des Parameters $\theta = (\theta_1, \dots, \theta_m)$, wobei $g : \Theta \rightarrow \mathbb{R}$ eine eindimensionale messbare Abbildung des Parameterwertes ist.

Anmerkung: Im letzten Abschnitt hatten wir den Fall betrachtet, bei dem g die Projektion auf eine Koordinate ist, also $g(\theta) = \theta_j$ für ein $1 \leq j \leq n$.

Dazu betrachten wir die Stichprobenfunktionen $\theta_u : \mathbb{R}^{n_1+n_2} \rightarrow \mathbb{R}$ und $\theta_o : \mathbb{R}^{n_1+n_2} \rightarrow \mathbb{R}$.

Konfidenzintervall für die Differenz zweier Erwartungswerte bei bekannten Varianzen

Annahmen

(a) Seien n_1, n_2 beliebig, jedoch fest vorgegeben, d. h. wir betrachten die i. i. d. Zufallsstichproben $X_1 = (X_{11}, \dots, X_{1n_1})$ und $X_2 = (X_{21}, \dots, X_{2n_2})$, wobei X_1 und X_2 unabhängig sind.

(b) Gelte $X_{11} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_{21} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, wobei $\mu_1, \mu_2 \in \mathbb{R}$ unbekannt sind, $\sigma_1^2, \sigma_2^2 > 0$ bekannt.

(c) Daraus folgt insbesondere, dass $\bar{X}_{1n_1} := \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}$ und $\bar{X}_{2n_2} := \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2i}$ unabhängig sind.

(d) Damit ist der unbekannte Parametervektor $\theta = (\mu_1, \mu_2)$ und $g(\theta) = \mu_1 - \mu_2$.

6 Parameterschätzung

Wir wissen, dass

$$\bar{X}_{1n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2/n_1) \quad \text{und} \quad \bar{X}_{2n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2/n_2),$$

woraus wir mit der Unabhängigkeit auf

$$\bar{X}_{1n_1} - \bar{X}_{2n_2} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

schließen. Damit können wir für das Konfidenzintervall für jedes $\alpha \in (0, 1)$ nach Normierung

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{(\bar{X}_{1n_1} - \bar{X}_{2n_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq z_{1-\alpha/2}\right) = 1 - \alpha \quad (6.52)$$

ansetzen. Nach Umformung haben wir

$$\mathbb{P}\left(\bar{X}_{1n_1} - \bar{X}_{2n_2} - z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_{1n_1} - \bar{X}_{2n_2} + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

bzw. die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) := \bar{X}_{1n_1} - \bar{X}_{2n_2} - z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

und

$$\theta_o(X_1, \dots, X_n) := \bar{X}_{1n_1} - \bar{X}_{2n_2} + z_{1-\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

und damit ein Konfidenzintervall für $\mu_1 - \mu_2$ zum Niveau $1 - \alpha$.

Beispiel 6.36: Zwei Branchen sollen auf ähnliche Entwicklung nach der Wirtschaftskrise überprüft werden. Wir gehen davon aus, dass die Unternehmensgrößen in beiden Branchen etwa gleich sind. Betrachtet wird die Umsatzentwicklung der vergangenen 6 Monate. Diese kann als normalverteilt angenommen werden, da sich der Umsatz in diesem Zeitraum aus vielen Einzelbeträgen zusammensetzt und der Zentrale Grenzwertsatz bemüht werden kann.

Konfidenzintervall für den Quotienten zweier Varianzen bei unbekanntem Erwartungswerten

Annahmen Wir gehen analog zu oben vor.

(a) Seien n_1, n_2 wieder beliebig, jedoch fest vorgegeben, d.h. wir betrachten die Zufallsstichproben $(X_{11}, \dots, X_{1n_1})$ und $(X_{21}, \dots, X_{2n_2})$.

(b) Gelte $X_{11} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_{21} \sim \mathcal{N}(\mu_2, \sigma_2^2)$, wobei $\mu_1, \mu_2 \in \mathbb{R}$ wie auch $\sigma_1^2, \sigma_2^2 > 0$ unbekannt sind.

(c) Daraus folgt in diesem Fall, dass $S_{1n_1}^2 := \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_{1n_1})^2$ und $S_{2n_2}^2 := \frac{1}{n_2-1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_{2n_2})^2$ unabhängig sind.

(d) Damit ist der unbekannte Parametervektor $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ und $g(\theta) = \sigma_1^2/\sigma_2^2$.

Auch hier können wir beim Bekannten ansetzen, und zwar ist

$$\frac{(n_1 - 1)S_{1n_1}^2}{\sigma_1^2} \sim \chi_{n_1-1}^2 \quad \text{und} \quad \frac{(n_2 - 1)S_{2n_2}^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$$

weswegen mit der Unabhängigkeit

$$\frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2} \sim F_{n_2-1, n_1-1}$$

gilt. Für jedes $\alpha \in (0, 1)$ gilt dann

$$\mathbb{P} \left(F_{n_2-1, n_1-1, \alpha/2} \leq \frac{S_{2n_2}^2/\sigma_2^2}{S_{1n_1}^2/\sigma_1^2} \leq F_{n_2-1, n_1-1, 1-\alpha/2} \right) = 1 - \alpha,$$

und nach Umformung

$$\mathbb{P} \left(\frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, \alpha/2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, 1-\alpha/2} \right) = 1 - \alpha,$$

woraus wir die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) := \frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, \alpha/2} \quad \text{und} \quad \theta_o(X_1, \dots, X_n) := \frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1, n_1-1, 1-\alpha/2} \quad (6.53)$$

ablesen und damit ein Konfidenzintervall für $g(\theta) = \sigma_1^2/\sigma_2^2$ zum Niveau $1 - \alpha$ erhalten.

Konfidenzintervall für die Differenz der Erwartungswerte bei verbundenen Stichproben

Problemstellung Hier betrachten wir den Fall, bei dem die Zufallsstichproben $X_1 = (X_{11}, \dots, X_{1n_1})$ und $X_2 = (X_{21}, \dots, X_{2n_2})$ nicht voneinander unabhängig sind. Das ist immer dann der Fall, wenn wir eine Vorher-Nachher-Observation haben. Allgemeiner immer dann, wenn X_{1i} und X_{2i} auch stochastisch abhängig sind.

Annahmen

- (a) Im Gegensatz zu den vorherigen Beispielen ist hier $n_1 = n_2$.
- (b) Sei n beliebig, jedoch fest vorgegeben, d. h. wir betrachten die Zufallsstichproben $(X_{11}, \dots, X_{1n_1})$ und $(X_{21}, \dots, X_{2n_2})$.
- (c) Die Komponenten X_{i1} und X_{i2} müssen nicht unabhängig sein.
- (d) Sei $\mathbb{E}X_{11} = \mu_1$ und $\mathbb{E}X_{21} = \mu_2$, wobei $\mu_1, \mu_2 \in \mathbb{R}$ unbekannt sind.
- (e) Des weiteren seien die Differenzen $Y_i := X_{1i} - X_{2i}$ unabhängig und identisch normalverteilt mit $Y_1 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma^2)$ mit $\sigma^2 > 0$ (unbekannt).

Mit den Eigenschaften des Satzes 6.32 für die Zufallsstichprobe (Y_1, \dots, Y_n) wissen wir, dass

$$\bar{Y}_n \sim \mathcal{N}(\mu_1 - \mu_2, \sigma^2/n) \quad \text{und} \quad \frac{n-1}{\sigma^2} S_{Y,n}^2 \sim \chi_{n-1}^2 \quad \text{und} \quad \frac{\sqrt{n}(\bar{Y}_n - (\mu_1 - \mu_2))}{S_{Y,n}} \sim t_{n-1},$$

6 Parameterschätzung

und können für das Konfidenzintervall zum Niveau $\alpha \in (0, 1)$ die Gleichung

$$\mathbb{P} \left(t_{n-1, \alpha/2} \leq \frac{\sqrt{n} (\bar{Y}_n - (\mu_1 - \mu_2))}{S_{Y,n}} \leq t_{n-1, 1-\alpha/2} \right) = 1 - \alpha$$

ansetzen, die umgeformt

$$\mathbb{P} \left(\bar{Y}_n - t_{n-1, 1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \leq \mu_1 - \mu_2 \leq \bar{Y}_n + t_{n-1, 1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \right) = 1 - \alpha$$

ist und die Stichprobenfunktionen

$$\theta_u(Y_1, \dots, Y_n) = \bar{Y}_n - t_{n-1, 1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \quad \text{und} \quad \theta_o(Y_1, \dots, Y_n) = \bar{Y}_n + t_{n-1, 1-\alpha/2} \frac{S_{Y,n}}{\sqrt{n}} \quad (6.54)$$

liefert.

Beispiel 6.37: Wir kehren zurück zu unserem Kaffeehausbeispiel und wollen untersuchen, wie sich die durchschnittlichen Umsätze zueinander verhalten. Wir wählen den Montag und Dienstag beispielhaft aus und gehen davon aus, dass es Einflüsse gibt, die wochenweise auftreten, z. B. Ferien, besonderes Wetter usw. Aus diesem Grund betrachten wir die wochenweise Differenz der Umsatzzahlen.

Beispiel 6.38: Wir wollen feststellen, ob sich der Luftdruck mit der Höhe ändert. Dazu müssen wir mehrmals Messungen durchführen. Wir messen einen Monat lang jeden Tag am Münsterplatz in Ulm wie auch auf der Besucherplattform des Münsterturms. Da sich das Wetter in dieser Zeit ändert, was wir nicht beeinflussen können, und dies den Luftdruck verändert, haben wir es mit verbundenen Stichproben zu tun.

Anmerkung: Das Konzept der verbundenen Stichprobe wird dann eingesetzt, wenn die Grundbedingungen nicht beeinflussbar sind und wir somit für die Zufallsstichprobe keine identische Verteilung annehmen können. Die Idee ist, eine kontrollierte Veränderung eines anderen Einfluß nehmenden Aspektes zu erzeugen, um einen Effekt festzustellen.

6.4.2 Asymptotische Konfidenzintervalle

Wenn wir die Normalverteilungsannahme nicht treffen können, müssen wir einen anderen Weg finden. In diesem Abschnitt betrachten wir eine mögliche Näherung.

Wir gehen von einer Zufallsstichprobe aus und zeigen, wie mittels des zentralen Grenzwertsatzes und des starken Gesetzes der großen Zahlen asymptotische Konfidenzintervalle aufgestellt werden können.

Konfidenzintervall für den Parameter η bei der Bernoulli-Verteilung

Seien $X_1, \dots, X_n \sim \text{Bin}(1, \eta)$, wobei $\eta \in (0, 1)$ unbekannt und zu schätzen ist. Wir wissen, dass $\mathbb{E}X_1 = \eta$ und $\text{Var}X_1 = \eta(1 - \eta)$. Mit dem zentralen Grenzwertsatz folgt daraus für alle $\alpha \in (0, 1)$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(z_{\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \eta}{\sqrt{\eta(1 - \eta)}} \leq z_{1-\alpha/2} \right) = 1 - \alpha.$$

Wir betrachten die doppelte Ungleichung genauer und erhalten mit $z := z_{1-\alpha/2}$ die äquivalente Bedingung

$$\begin{aligned} \left| \frac{\bar{X}_n - \eta}{\sqrt{\eta(1-\eta)}} \right| &\leq \frac{z}{\sqrt{n}} \\ (\bar{X}_n - \eta)^2 &\leq \frac{z^2}{n} (\eta(1-\eta)) \\ \eta^2 - 2\eta \bar{X}_n + \bar{X}_n^2 &\leq \frac{z^2}{n} \eta - \frac{z^2}{n} \eta^2 \\ \eta^2 \left(1 + \frac{z^2}{n}\right) - \eta \left(2\bar{X}_n + \frac{z^2}{n}\right) + \bar{X}_n^2 &\leq 0 \\ \left(\frac{n+z^2}{n}\right) \left[\eta^2 - 2\eta \frac{\bar{X}_n + \frac{z^2}{2n}}{\frac{n+z^2}{n}} + \frac{n\bar{X}_n^2}{n+z^2} \right] &\leq 0 \\ \eta^2 - 2\eta \frac{n\bar{X}_n + z^2/2}{n+z^2} + \frac{n\bar{X}_n^2}{n+z^2} &\leq 0 \\ \left(\eta - \frac{n\bar{X}_n + z^2/2}{n+z^2}\right)^2 &\leq \left(\frac{n\bar{X}_n + z^2/2}{n+z^2}\right)^2 - \frac{n\bar{X}_n^2}{n+z^2} \\ \left(\eta - \frac{n\bar{X}_n + z^2/2}{n+z^2}\right)^2 &\leq \frac{n^2\bar{X}_n^2 + z^4/4 + n\bar{X}_nz^2}{(n+z^2)^2} - \frac{n^2\bar{X}_n^2 + n\bar{X}_nz^2}{(n+z^2)^2} \\ \left| \eta - \frac{n\bar{X}_n + z^2/2}{n+z^2} \right| &\leq \frac{z\sqrt{z^2/4 + n\bar{X}_n(1-\bar{X}_n)}}{n+z^2}, \end{aligned}$$

aus welcher die Stichprobenfunktionen

$$\theta_u(X_1, \dots, X_n) := \frac{1}{n + z_{1-\alpha/2}^2} \left(n\bar{X}_n + \frac{z_{1-\alpha/2}^2}{2} - z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + n\bar{X}_n(1-\bar{X}_n)} \right) \quad (6.55)$$

$$\theta_o(X_1, \dots, X_n) := \frac{1}{n + z_{1-\alpha/2}^2} \left(n\bar{X}_n + \frac{z_{1-\alpha/2}^2}{2} + z_{1-\alpha/2} \sqrt{\frac{z_{1-\alpha/2}^2}{4} + n\bar{X}_n(1-\bar{X}_n)} \right) \quad (6.56)$$

folgen, die ein asymptotisches Konfidenzintervall $(\theta_u(X_1, \dots, X_n), \theta_o(X_1, \dots, X_n))$ für η zum Niveau $1 - \alpha$ ergeben.

6 Parameterschätzung

7 Statistische Tests

7.1 Prinzipien des Testens

Ein statistischer Test ist ein Verfahren, mit dem wir an Hand von Stichproben Hypothesen über Eigenschaften der unbekanntem Verteilungsfunktion F der Stichprobenvariablen X_1, \dots, X_n prüfen.

Vorgehen

- Wir konstruieren eine Entscheidungsregel, nach der eine Hypothese verworfen oder beibehalten wird.¹
- Die Entscheidungsregel wird so gewählt, dass die Wahrscheinlichkeiten möglicher Fehlentscheidungen minimiert werden.
- Analog zu den Konfidenzintervallen geben wir uns ein Signifikanzniveau α vor.
- Die Entscheidung treffen wir dann abhängig von den beobachteten Daten, d. h. der Realisierung der Zufallsstichprobe.

Beispiel 7.1: Wir haben zwei äußerlich nicht unterscheidbare Münzen, deren Wahrscheinlichkeit η Kopf zu werfen bei 0,4 bzw. 0,5 liegt. Wir haben eine ausgesucht und vergessen welche es war. Nun werfen wir diese Münze 1001 mal zur Bestimmung von η . Dazu stellen wir die Hypothese $\eta = 0.5$ auf. Wir wollen diese ablehnen, falls

$$Z = \{\text{Anzahl von Würfeln mit dem Ausgang Kopf}\} \leq 450$$

und behalten sie weiter bei, falls $Z > 450$.

Falls $Z \sim \text{Bin}(1001, 0,5)$, dann ist $\mathbb{P}_{0,5}(Z \leq 450) = 7,8 \cdot 10^{-4}$; und falls $Z \sim \text{Bin}(1001, 0,4)$, dann ist $\mathbb{P}_{0,4}(Z > 450) = 1 - \mathbb{P}_{0,4}(Z \leq 450) = 6,5 \cdot 10^{-4}$.

Wir haben damit die Fehler der Fehlentscheidung. Diese sind nicht gleich, da unter anderem die Varianzen unterschiedlich sind, 250,25 resp. 240,24.

Es ist möglich, einen Fehler zu reduzieren. Sagen wir, die Hypothese wird beibehalten, falls $Z < 430$ und sonst verworfen. Dann ist $\mathbb{P}_{0,5}(Z < 430) = 4,0 \cdot 10^{-6}$ und $\mathbb{P}_{0,5}(Z \geq 430) = 0,029$.

7.2 Aufbau eines statistischen Tests

Sei \mathcal{P} eine Verteilungsfamilie der Verteilungen P (mit der Verteilungsfunktion F), die den Stichprobenvariablen X_1, \dots, X_n , die wir als unabhängig und identisch verteilt annehmen, zu Grunde liegen.²

¹Zu beachten ist, dass eine Hypothese nicht angenommen werden kann.

²Obwohl formal nicht korrekt, werden wir manchmal die Notation $F \in \mathcal{P}$ verwenden.

Hypothese Wir bilden eine Zweier-Partition von \mathcal{P} , also $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$, $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$ und betrachten die Hypothesen

$$\begin{aligned} H_0 &: P \in \mathcal{P}_0 \\ H_1 &: P \in \mathcal{P}_1 \end{aligned}$$

dafür dass die (unbekannte) Verteilung P in \mathcal{P}_0 bzw. \mathcal{P}_1 liegt. Wir nennen H_0 auch die *Nullhypothese* und H_1 die *Alternativhypothese*.

Die Hypothesen heißen *einfach*, falls $|\mathcal{P}_0| = 1$ bzw. $|\mathcal{P}_1| = 1$, ansonsten *zusammengesetzt* (Im Beispiel 7.1 waren sowohl die Null- als auch die Alternativhypothese einfach.)

Entscheidungsregel Der Stichprobenraum wird deshalb ebenfalls in zwei (messbare) Mengen zerlegt, K und $K^c \subset \mathbb{R}^n$. Dabei wird K der *kritische Bereich* genannt und die Nullhypothese wird verworfen, falls $(x_1, \dots, x_n) \in K$.

Dazu definieren wir uns eine Stichprobenfunktion $\varphi : \mathcal{X} \rightarrow \{0, 1\}$ durch

$$\varphi(x_1, \dots, x_n) = \mathbb{1}_K(x_1, \dots, x_n) \quad (7.1)$$

und sagen, dass H_0 verworfen wird, falls $\varphi(x_1, \dots, x_n) = 1$ ist; falls die Stichprobe also in der Menge K landet.

Definition 7.2: Sei $\alpha \in (0, 1)$ fest und sei \mathcal{P} eine Verteilungsfamilie. Dann ist die Funktion φ aus 7.1 ein *Test zum Niveau α* , falls

$$\mathbb{P}_F(\varphi(X_1, \dots, X_n) = 1) \leq \alpha$$

für alle $F \in \mathcal{P}_0$. Gilt zusätzlich

$$\mathbb{P}_F(\varphi(X_1, \dots, X_n) = 1) \geq \alpha$$

für alle $F \in \mathcal{P}_1$, so heißt φ *unverfälschter Test zum Niveau α* .

Anmerkung: (a) Der *Fehler erster Art* ist es, die Nullhypothese abzulehnen, obwohl sie zutrifft, und seine Wahrscheinlichkeit ist

$$\alpha_n(F) = \mathbb{P}_F((X_1, \dots, X_n) \in K) = \mathbb{P}_F(\varphi(X_1, \dots, X_n) = 1) \quad (7.2)$$

für alle $F \in \mathcal{P}_0$.

(b) Der *Fehler zweiter Art* ist es, die Nullhypothese nicht abzulehnen, obwohl sie nicht zutrifft, und seine Wahrscheinlichkeit ist

$$\beta_n(F) = \mathbb{P}_F((X_1, \dots, X_n) \notin K) = \mathbb{P}_F(\varphi(X_1, \dots, X_n) = 0) \quad (7.3)$$

für alle $F \in \mathcal{P}_1$.

(c) Wir interessieren uns für Tests, für die zu einem vorgegebenen α die Wahrscheinlichkeit $\beta_n(F)$ für den Fehler zweiter Art möglichst klein ist.

Parametrische Verteilungsfamilie Ist die Verteilungsfamilie \mathcal{P} parametrisch, können wir die Struktur und Entscheidung des Tests über die Parameter der zugehörigen Verteilungen angeben. Sei $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ gegeben, so definieren wir zu \mathcal{P}_0 und \mathcal{P}_1 eine Partition $\{\Theta_0, \Theta_1\}$ von Θ so dass

$$\mathcal{P}_0 = \{P_\theta \in \mathcal{P} : \theta \in \Theta_0\} \quad \text{und} \quad \mathcal{P}_1 = \{P_\theta \in \mathcal{P} : \theta \in \Theta_1\} \quad (7.4)$$

gilt und formulieren die Hypothesen

$$\begin{aligned} H_0 &: \theta \in \Theta_0, \\ H_1 &: \theta \in \Theta_1. \end{aligned}$$

Beispiel 7.3: Sei $\mathcal{P} = \left\{ \mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R} \right\}$ für ein vorgegebenes σ^2 . Betrachte die Hypothesen

$$\begin{aligned} H_0 &: \mu = 0 \\ H_1 &: \mu \neq 0 \end{aligned}$$

wobei H_0 einfach und H_1 zusammengesetzt sind.

Definition 7.4: Sei $\alpha \in (0, 1)$ beliebig aber fest und sei \mathcal{P} eine parametrische Verteilungsfamilie. Dann ist die Funktion φ aus 7.1 ein *Parametertest zum Niveau α* , falls

$$\mathbb{P}_\theta (\varphi(X_1, \dots, X_n) = 1) \leq \alpha$$

für jedes $\theta \in \Theta_0$. Gilt zusätzlich

$$\mathbb{P}_\theta (\varphi(X_1, \dots, X_n) = 1) \geq \alpha$$

für alle $\theta \in \Theta_1$, so heißt φ *unverfälschter Parametertest zum Niveau α* .

Bei parameterischen Verteilungsfamilien können wir zusätzlich definieren:

Definition 7.5: (i) Die Funktion

$$\begin{aligned} \alpha_n &: \Theta \rightarrow [0, 1] \\ \alpha_n(\theta) &= \mathbb{P}_\theta (\varphi(X_1, \dots, X_n) = 1) \end{aligned}$$

heißt *Gütefunktion* des Parametertests φ .

(ii) Die Funktion $\alpha_n : \Theta_1 \rightarrow [0, 1]$ (eingeschränkt auf Θ_1) heißt *Macht* des Parametertests φ .

(iii) Der Test zum Niveau α heißt *konsistent*, falls $\lim_{n \rightarrow \infty} \alpha_n(\theta) = 1$ für jedes $\theta \in \Theta_1$.

(iv) Seien φ_1 und φ_2 zwei Parametertests zum Niveau α . Falls die Macht von φ_1 größer ist als die Macht von φ_2 für alle $\theta \in \Theta_1$, d. h.

$$\mathbb{P}_\theta (\varphi_1(X_1, \dots, X_n) = 1) \geq \mathbb{P}_\theta (\varphi_2(X_1, \dots, X_n) = 1) \quad \text{für alle } \theta \in \Theta_1,$$

so sagen wir, dass φ_1 *besser* ist als φ_2 .

Anmerkung: Ein Test ist unverfälscht, wenn für die Gütefunktion $\alpha_n(\theta)$ für alle $\theta \in \Theta_1$ die Ungleichung $\alpha_n(\theta) \geq \alpha$ gilt.

Vorgehen zur Bestimmung des kritischen Bereichs K eines Parametertests zum Niveau α .

(a) Definiere die Teststatistik $T : \mathcal{X} \rightarrow \mathbb{R}$ so dass

$$T(X_1, \dots, X_n | \theta) \sim P$$

für alle $\theta \in \Theta_0$ und eine Verteilung P , die nicht von θ abhängt.

(b) Bestimme $c > 0$ so, dass

$$\mathbb{P}_\theta \left(|T(X_1, \dots, X_n)| > c \right) < \alpha$$

für alle $\theta \in \Theta_0$ gilt.

7 Statistische Tests

(c) Damit ist die Menge

$$K = \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| > c\}$$

der kritische Bereich eines Parametertests zum Niveau α .

Manchmal ist ein asymmetrisch definierter Bereich sinnvoller, da er die Macht des Tests erhöhen kann.

Ersetze dann (b) und (c) durch

(b') Bestimme $c_1, c_2 \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ mit $c_1 < c_2$ so, dass

$$\mathbb{P}_\theta (c_1 \leq T(X_1, \dots, X_n) \leq c_2) \geq 1 - \alpha$$

für alle $\theta \in \Theta_0$ gilt.

(c') Dann ist die Menge

$$K = \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : c_1 \leq T(x_1, \dots, x_n) \leq c_2\}$$

der kritische Bereich des asymmetrischen Parametertests zum Niveau α .

Anmerkung (Statistische Tests und Konfidenzintervalle): Die Komplementärmenge K^c des kritischen Bereichs

$$K = \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| \leq c\}$$

bzw.

$$K = \{(x_1, \dots, x_n) : c_1 \leq T(x_1, \dots, x_n) \leq c_2\}$$

führte zur Konstruktion des Konfidenzintervalls. Während wir dabei ein zufälliges Intervall für eine feste Größe, den Parameterwert $g(\theta)$ erstellen, betrachten wir bei den Tests ein festes Intervall (denn wir haben eine Hypothese) für eine zufällige Größe.

7.3 Parametertests bei Normalverteilung

Wir nehmen im ganzen Abschnitt an, dass die Stichprobenvariablen X_1, \dots, X_n i.i.d. und normalverteilt sind, d. h. $\mathcal{P} \subset \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$.

Test des Erwartungswertes μ (bei bekannter Varianz σ^2)

Annahme Gelte $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ unbekannt, $\sigma^2 > 0$ bekannt und sei $X = (X_1, \dots, X_n)$ die Stichprobe.

Wir geben uns ein $\mu_0 \in \mathbb{R}$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \mu &= \mu_0 & \Theta_0 &= \{\mu_0\} \\ H_1 : \mu &\neq \mu_0 & \Theta_1 &= \mathbb{R} \setminus \{\mu_0\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \rightarrow (0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma} \quad \text{und es gilt} \quad T(X) = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \sim \mathcal{N}(0, 1) \quad (7.5)$$

unter der Nullhypothese, wobei $c = z_{1-\alpha/2}$ ist und die Schwellenwerte $-c_1 = c_2 = c$ gewählt werden. Dann gilt für alle $\mu \in \Theta_0$, also für μ_0 , dass

$$\mathbb{P}_{\mu_0}(c_1 \leq T(X) \leq c_2) = 1 - \mathbb{P}_{\mu_0}(|T(X)| > z_{1-\alpha/2}) = 1 - \alpha. \quad (7.6)$$

Der kritische Bereich ist dann die Menge

$$\begin{aligned} K &= \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| > c\} \\ &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| \leq z_{1-\alpha/2}\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : -z_{1-\alpha/2} \leq \frac{\sqrt{n}}{\sigma} (\bar{x}_n - \mu_0) \leq z_{1-\alpha/2} \right\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \bar{x}_n \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\}, \end{aligned}$$

wobei die Menge K^c ebenfalls die Bedingung des symmetrischen Konfidenzintervalls für μ bei bekanntem σ^2 zum Niveau $1 - \alpha$ bei Normalverteilung darstellt.

Beispiel 7.6: Wir erinnern uns an die Füllmaschine aus Beispiel 6.35. Zu $\alpha = 0,01$ wollen wir die Hypothesen

$$\begin{aligned} H_0 : \mu &= \mu_0 = 18,2 & \Theta_0 &= \{18,2\} \\ H_1 : \mu &\neq 18,2 & \Theta_1 &= \mathbb{R} \setminus \{18,2\} \end{aligned}$$

testen. Dann ist $|T(x_1, \dots, x_9)| = 3,5 > 2,576 = z_{0,995}$. Diese Nullhypothese wird abgelehnt.

Anmerkung (Gütefunktion): Wir untersuchen die Gütefunktion $\alpha_n : \rightarrow [0, 1]$ dieses Tests

$$\begin{aligned} \alpha_n(\mu) &= \mathbb{P}_\mu(|T(X_1, \dots, X_n)| > z_{1-\alpha/2}) \\ &= 1 - \mathbb{P}_\mu(-z_{1-\alpha/2} \leq T(X_1, \dots, X_n) \leq z_{1-\alpha/2}) \\ &= 1 - \left(\Phi\left(z_{1-\alpha/2} - \sqrt{n} \frac{\mu - \mu_0}{\sigma}\right) - \Phi\left(-z_{1-\alpha/2} - \sqrt{n} \frac{\mu - \mu_0}{\sigma}\right) \right) \end{aligned}$$

Hieraus folgt

$$\begin{aligned} \alpha_n(\mu_0) &= \alpha \\ \alpha_n(\mu) &> \alpha \quad \text{für } \mu \neq \mu_0 \end{aligned}$$

und

$$\begin{aligned} \mu < \mu_0 : & \lim_{n \rightarrow \infty} \alpha_n(\mu) = 1 \\ \mu > \mu_0 : & \lim_{n \rightarrow \infty} \alpha_n(\mu) = 1 \end{aligned}$$

Der Test ist deshalb unverfälscht und konsistent.

Einseitiger Test Die Alternativhypothese kann auch nur nach einer Seite betrachtet werden. Die Hypothesen für den linksseitigen resp. rechtsseitigem Test sind

$$\begin{aligned} H_0 : \mu &= \mu_0 & \Theta_0 &= \{\mu_0\} \\ H_1 : \mu &> \mu_0 & \Theta_1 &= \{\mu \in \mathbb{R} : \mu > \mu_0\} \\ \text{oder } H_1 : \mu &< \mu_0 & \Theta_1 &= \{\mu \in \mathbb{R} : \mu < \mu_0\} \end{aligned}$$

Mit der gleichen Teststatistik haben wir zwei Möglichkeiten zur Bestimmung des kritischen Bereichs, entweder

$$K = \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| > z_{1-\alpha/2}\}$$

wie bisher oder

$$\tilde{K} = \{(x_1, \dots, x_n) : T(x_1, \dots, x_n) > z_{1-\alpha}\}$$

als einseitigen kritischen Bereich.

Die zugehörigen Gütefunktionen α_n und $\tilde{\alpha}_n$ sind

$$\begin{aligned} \alpha_n &= 1 - \Phi\left(\sqrt{n}\frac{\mu_0 - \mu}{\sigma} + z_{1-\alpha/2}\right) + \Phi\left(\sqrt{n}\frac{\mu_0 - \mu}{\sigma} - z_{1-\alpha/2}\right) \\ \tilde{\alpha}_n &= 1 - \Phi\left(\sqrt{n}\frac{\mu_0 - \mu}{\sigma} + z_{1-\alpha}\right) \end{aligned}$$

und es gilt $\alpha_n(\mu) < \tilde{\alpha}_n(\mu)$ für alle $\mu \in \Theta_1$. Deswegen bevorzugen wir den einseitigen kritischen Bereich \tilde{K} .

Test des Erwartungswertes μ (bei unbekannter Varianz σ^2)

Annahme Gelte $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}, \sigma^2 > 0$ unbekannt und sei $X = (X_1, \dots, X_n)$ die Stichprobe.

Wir geben uns ein $\mu_0 \in \mathbb{R}$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \mu &= \mu_0 & \Theta_0 &= \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\} \\ H_1 : \mu &\neq \mu_0 & \Theta_1 &= \{(\mu, \sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \rightarrow (0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sqrt{n} \frac{\bar{x}_n - \mu_0}{s_n} \quad \text{und es gilt} \quad T(X) = \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} \sim t_{n-1} \quad (7.7)$$

unter der Nullhypothese. Seien $c = t_{n-1, 1-\alpha/2}$ und die Schwellenwerte $-c_1 = c_2 = c$. Dann gilt für alle $(\mu, \sigma^2) \in \Theta_0$, dass

$$\mathbb{P}_{(\mu_0, \sigma^2)}(c_1 \leq T(X) \leq c_2) = 1 - \mathbb{P}_{(\mu_0, \sigma^2)}(|T(X)| > t_{n-1, 1-\alpha/2}) = 1 - \alpha. \quad (7.8)$$

Der kritische Bereich ist dann die Menge

$$\begin{aligned} K &= \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| > c\} \\ &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : |T(x_1, \dots, x_n)| \leq t_{n-1, 1-\alpha/2}\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : -t_{n-1, 1-\alpha/2} \leq \frac{\sqrt{n}}{s_n}(\bar{x}_n - \mu_0) \leq t_{n-1, 1-\alpha/2} \right\} \end{aligned}$$

Beispiel 7.7: Wir beschäftigen uns immer noch mit der Füllmaschine. Sei $\alpha = 0,05$ gewählt und die Hypothese

$$H_0 : \mu = \mu_0 = 18,2$$

$$H_1 : \mu \neq \mu_0$$

Dann ist $|T(x_1, \dots, x_9)| = 6,396 > 2,306 = t_{8, 0,975}$. Die Null-Hypothese wird abgelehnt.

Anmerkung: Die Entscheidungen in den Tests des Erwartungswertes bei bekanntem σ^2 bzw. bei unbekanntem σ^2 können sich auch bei identischem α unterscheiden.

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \mu = \mu_0 & \quad \Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\} \\ H_1 : \mu > \mu_0 & \quad \Theta_1 = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\} \\ \text{oder } H_1 : \mu < \mu_0 & \quad \Theta_1 = \{(\mu, \sigma^2) : \mu < \mu_0, \sigma^2 > 0\} \end{aligned}$$

mit gleicher Teststatistik, dem Schwellenwert $c = t_{n-1, 1-\alpha}$ und dem resultierenden kritischen Bereich $K = \{(x_1, \dots, x_n) : T(x_1, \dots, x_n) > t_{n-1, 1-\alpha}\}$.

Test der Varianz σ^2 (bei bekanntem Erwartungswert μ)

Annahme Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabhängig mit $\mu \in \mathbb{R}$ bekannt, $\sigma^2 > 0$ unbekannt und sei $X = (X_1, \dots, X_n)$ die Stichprobe.

Wir geben uns ein $\sigma_0 > 0$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \sigma^2 = \sigma_0^2 & \quad \Theta_0 = \{\sigma_0^2\} \\ H_1 : \sigma^2 \neq \sigma_0^2 & \quad \Theta_1 = (0, \infty) \setminus \{\sigma_0^2\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \rightarrow (0, \infty)$ mit

$$T(x_1, \dots, x_n) = \frac{n}{\sigma_0^2} \bar{s}_n^2 \quad \text{und es gilt} \quad T(X) = \frac{n}{\sigma_0^2} \bar{S}_n^2 \sim \chi_n^2 \quad (7.9)$$

unter der Nullhypothese, wobei

$$\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad \text{und} \quad \bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (7.10)$$

7 Statistische Tests

ist, mit den Schwellenwerten $c_1 = \chi_{n,\alpha/2}^2$ und $c_2 = \chi_{n,1-\alpha/2}^2$.

Es gilt für alle $\sigma^2 \in \Theta_0$, also für σ_0^2 , dass

$$\mathbb{P}_{\sigma_0^2} (c_1 \leq T(X) \leq c_2) = 1 - \alpha, \quad (7.11)$$

und der kritische Bereich ist

$$\begin{aligned} K &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : c_1 \leq T(x_1, \dots, x_n) \leq c_2\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : \chi_{n,\alpha/2}^2 \leq \frac{n}{\sigma_0^2} s_n^2 \leq \chi_{n,1-\alpha/2}^2 \right\} \end{aligned}$$

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 & \Theta_0 &= \{\sigma_0^2\} \\ H_1 : \sigma^2 &> \sigma_0^2 & \Theta_1 &= \{\sigma^2 : \sigma^2 > \sigma_0^2\} \\ \text{oder } H_1 : \sigma^2 &< \sigma_0^2 & \Theta_1 &= \{\sigma^2 : \sigma^2 < \sigma_0^2\} \end{aligned}$$

Mit der gleichen Teststatistik und den Schwellenwerten $c_1 = -\infty$ und $c := c_2 = \chi_{n,1-\alpha}^2$ gilt (7.11) analog und wir haben den kritischen Bereich

$$\begin{aligned} \tilde{K} &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : T(x_1, \dots, x_n) \leq c\} \\ &= \left\{ (x_1, \dots, x_n) : \frac{n}{\sigma_0^2} s_n^2 > \chi_{n,1-\alpha}^2 \right\}. \end{aligned}$$

Test der Varianz σ^2 (bei unbekanntem Erwartungswert μ)

Annahme Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ unabhängig mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt und sei $X = (X_1, \dots, X_n)$ die Stichprobe.

Wir geben uns ein $\sigma_0 > 0$ vor, auf dem die Hypothese beruht.

Zweiseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 & \Theta_0 &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \\ H_1 : \sigma^2 &\neq \sigma_0^2 & \Theta_1 &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \neq \sigma_0^2\} \end{aligned}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \rightarrow (0, \infty)$ mit

$$T(x_1, \dots, x_n) = \frac{n-1}{\sigma_0^2} s_n^2 \quad \text{und es gilt} \quad T(X) = \frac{n-1}{\sigma_0^2} S_n^2 \sim \chi_{n-1}^2 \quad (7.12)$$

unter der Nullhypothese mit den Schwellenwerten $c_1 = \chi_{n-1,\alpha/2}^2$ und $c_2 = \chi_{n-1,1-\alpha/2}^2$. Es gilt für alle $(\mu, \sigma^2) \in \Theta_0$, dass

$$\mathbb{P}_{(\mu, \sigma_0^2)} (c_1 \leq T(X) \leq c_2) = 1 - \alpha, \quad (7.13)$$

unabhängig von μ und der kritische Bereich ist

$$\begin{aligned} K &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : c_1 \leq T(x_1, \dots, x_n) \leq c_2\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : \chi_{n-1, \alpha/2}^2 \leq \frac{n-1}{\sigma_0^2} s_n^2 \leq \chi_{n-1, 1-\alpha/2}^2 \right\} \\ &= \mathbb{R}^n \setminus \left\{ (x_1, \dots, x_n) : \frac{\sigma_0^2}{\sigma^2} \chi_{n-1, \alpha/2}^2 \leq \frac{n-1}{\sigma^2} s_n^2 \leq \frac{\sigma_0^2}{\sigma^2} \chi_{n-1, 1-\alpha/2}^2 \right\} \end{aligned}$$

Einseitiger Test Die Hypothesen sind

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 & \Theta_0 &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \\ H_1 : \sigma^2 &> \sigma_0^2 & \Theta_1 &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > \sigma_0^2\} \\ \text{oder } H_1 : \sigma^2 &< \sigma_0^2 & \Theta_1 &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 < \sigma_0^2\} \end{aligned}$$

Mit der gleichen Teststatistik und den Schwellenwerten $c_1 = -\infty$ und $c := c_2 = \chi_{n-1, 1-\alpha}^2$ gilt (7.13) analog und wir haben den kritischen Bereich

$$\begin{aligned} \tilde{K} &= \mathbb{R}^n \setminus \{(x_1, \dots, x_n) : T(x_1, \dots, x_n) \leq c\} \\ &= \left\{ (x_1, \dots, x_n) : \frac{n-1}{\sigma_0^2} s_n^2 > \chi_{n-1, 1-\alpha}^2 \right\}. \end{aligned}$$

7.4 p-Wert versus Konfidenzniveau

Der p-Wert einer Teststatistik einer Stichprobe liefert die Wahrscheinlichkeit, dass unter der Nullhypothese eine Stichprobe mindestens so extrem ist wie die untersuchte. Er gewann erst mit der Verwendung von Computern und Statistiksoftware an Bedeutung, als es möglich wurde, diese Wahrscheinlichkeit zu berechnen.

Sei $z = T(x_1, \dots, x_n)$ die ausgewertete konkrete Stichprobe. Dann ist der p-Wert bei rechtsseitigem Test

$$p_r = p_r(x_1, \dots, x_n) = \mathbb{P}_{H_0}(T(X_1, \dots, X_n) \geq z),$$

bei linksseitigem Test

$$p_l = p_l(x_1, \dots, x_n) = \mathbb{P}_{H_0}(T(X_1, \dots, X_n) \leq z),$$

und bei beidseitigem Test

$$p = 2 \cdot \min\{p_l, p_r\}.$$

Vorgehen Zuerst wird ein Signifikanzniveau α festgelegt und wir sagen

$$(x_1, \dots, x_n) \in K, \quad \text{falls} \quad p(x_1, \dots, x_n) \leq \alpha,$$

was zur Entscheidung führt.

7.5 Asymptotische Tests

Ähnlich wie bei den Überlegungen zu Konfidenzintervallen, können wir ein Szenario aufbauen, bei dem für die Teststatistik selbst eine bestimmte Verteilungsaussage nicht gilt, dies jedoch für den Grenzwert der Fall ist. Eine formale Definition überspringen wir an dieser Stelle.

7.5.1 χ^2 -Anpassungstest

Annahme Seien X, X_1, \dots, X_n i. i. d. mit $X \sim F$ für eine Verteilung F .

Wir geben uns Intervalle $A_1 = (a_1, b_1], \dots, A_r = (a_r, b_r]$ vor, die paarweise disjunkt sind.

Desweiteren betrachten wir Vektoren (p_1, \dots, p_r) mit $0 < p_i < 1$ und $\sum_{i=1}^r p_i = 1$.

Hypothesen

$$H_0 : \mathbb{P}(X \in A_i) = \tilde{p}_i \text{ für alle } 1 \leq i \leq r \quad \Theta_0 = \{(\tilde{p}_1, \dots, \tilde{p}_r)\}$$

$$H_1 : \mathbb{P}(X \in A_i) \neq \tilde{p}_i \text{ für ein } 1 \leq i \leq r \quad \Theta_1 = \{(p_1, \dots, p_r) \in (0, 1)^r : \sum_{i=1}^r p_i = 1, \exists i \text{ mit } p_i \neq \tilde{p}_i\}$$

Wir untersuchen die Teststatistik $T : \mathcal{X} \rightarrow [0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sum_{j=1}^r \frac{(y_j(x_1, \dots, x_n) - n \cdot \tilde{p}_j)^2}{n \cdot \tilde{p}_j},$$

wobei $y_j(x_1, \dots, x_n) = \left| \left\{ i : 1 \leq i \leq n : a_j < x_i \leq b_j \right\} \right|$ die Anzahl der Stichprobenwerte x_1, \dots, x_n angibt, die im j -ten Intervall liegen. Die Funktionen T und y lassen sich auf die Zufallsstichprobe anwenden, und wir haben eine asymptotische Verteilungsaussage für die Zufallsvariable $T(X_1, \dots, X_n)$ mit

$$\lim_{n \rightarrow \infty} T(X_1, \dots, X_n) \sim \chi_{r-1}^2 \quad (7.14)$$

und mit dem Schwellenwert $c = \chi_{r-1, 1-\alpha}^2$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_{(\tilde{p}_1, \dots, \tilde{p}_r)} (T(X_1, \dots, X_n) > \chi_{r-1, 1-\alpha}^2) = \alpha. \quad (7.15)$$

Anmerkung: Die asymptotische Verteilung ergibt sich aus der näheren Untersuchung des Zufallsvektors (Y_1, \dots, Y_r) . Dieser ist multinomialverteilt mit den Parametern $n \in \mathbb{N}$ und $(\tilde{p}_1, \dots, \tilde{p}_r)$, d. h. für $k_1, \dots, k_r \in \mathbb{N}$ mit $\sum_{i=1}^r k_i = n$ gilt

$$\mathbb{P}(Y_1 = k_1, \dots, Y_r = k_r) = \frac{n!}{k_1! \cdot \dots \cdot k_r!} p_1^{k_1} \cdot \dots \cdot p_r^{k_r}$$

wobei $p_j = \mathbb{P}(X \in A_j) = F(b_j) - F(a_j)$ ist, da

$$\mathbb{P}(X_1 \in A_{i_1}, \dots, X_n \in A_{i_n}) = \prod_{j=1}^n (F(b_{i_j}) - F(a_{i_j}))$$

wegen der Unabhängigkeit der X_i gilt. Dann ist jede der r Zufallsvariablen Y_1, \dots, Y_r binomialverteilt, und damit asymptotisch normalverteilt, woraus (7.14) folgt.

Beispiel 7.8 (Wahlprognose): Anlässlich der bald stattfindenden Landtagswahl soll überprüft werden, ob sich das Wahlverhalten seit 2006 geändert hat. Die Verteilung, gegen die wir testen, basiert auf den Ergebnissen von 2006:

CDU	44,15 %
SPD	25,15 %
Grüne	11,69 %
FDP	10,65 %
andere	8,36 %

Emnid befragte in der zweiten Dezemberwoche (2010) 1001 Personen³ und wir haben

CDU	407
SPD	191
Grüne	295
FDP	37
andere	71

Wir wenden hier den χ^2 -Anpassungstest an, siehe R-Datei.

7.5.2 χ^2 -Unabhängigkeitstest

Annahme Wir haben eine verbundene i. i. d. Stichprobe X_1, \dots, X_n , wobei $X_i = (X_{i_1}, X_{i_2})$, deren Verteilung der von $X = (\tilde{X}_1, \tilde{X}_2)$ folgt, also $X_1 \stackrel{d}{=} X$. Die zwei Einträge \tilde{X}_1 und \tilde{X}_2 untereinander müssen nicht unabhängig sein, sondern werden gerade auf Unabhängigkeit getestet.

Wir geben uns Intervalle A_1, \dots, A_I und B_1, \dots, B_J vor, die jeweils paarweise disjunkt sind.

Hypothesen

$$H_0 : p_{ij} = \mathbb{P} \left((\tilde{X}_1, \tilde{X}_2) \in A_i \times B_j \right) = \mathbb{P}(\tilde{X}_1 \in A_i) \mathbb{P}(\tilde{X}_2 \in B_j) \quad \text{für alle } i, j$$

$$H_1 : p_{ij} = \mathbb{P} \left((\tilde{X}_1, \tilde{X}_2) \in A_i \times B_j \right) \neq \mathbb{P}(\tilde{X}_1 \in A_i) \mathbb{P}(\tilde{X}_2 \in B_j) \quad \text{für mindestens ein Paar } (i, j)$$

Wir definieren

$$N_{ij}(X_1, \dots, X_n) = \left| \left\{ k : 1 \leq k \leq n : X_{k_1} \in A_i, X_{k_2} \in B_j \right\} \right| \quad \text{für } 1 \leq i \leq I, 1 \leq j \leq J$$

die Anzahl der Stichprobenzufallsvariablen X_1, \dots, X_n , die im (i, j) Rechteck landen, und n_{ij} die Anzahl der (konkreten) Stichprobenwerte x_1, \dots, x_n entsprechend. Diese Werte kann man tabellarisch anordnen;

³Es wurden die relativen Ergebnisse veröffentlicht. Die Zahlen hier sind ein möglicher Ausgang der Umfrage.

die resultierende Tabelle wird Kontingenztafel genannt. In dieser definieren wir Spalten-

$$n_{\bullet j} := \sum_{i=1}^I n_{ij} \quad \text{und Zeilensummen} \quad n_{i\bullet} := \sum_{j=1}^J n_{ij}$$

und die Wahrscheinlichkeiten, in dem (i, j) Rechteck (bzw. Zelle der Tabelle) zu landen führen zu den Marginalwahrscheinlichkeiten

$$p_{\bullet j} := \sum_{i=1}^I p_{ij} \quad \text{und} \quad p_{i\bullet} := \sum_{j=1}^J p_{ij},$$

weiter bezeichnen wir mit π diese zwei-dimensionale Verteilung. Für diese Wahrscheinlichkeiten haben wir die Schätzer⁴

$$\hat{p}_{ij} \stackrel{\text{unter } H_0}{=} \hat{p}_{i\bullet} \cdot \hat{p}_{\bullet j} \stackrel{ML}{=} \frac{n_{i\bullet}}{n} \cdot \frac{n_{\bullet j}}{n}. \quad (7.16)$$

Dann gilt unter der Nullhypothese für die geschätzte Anzahl in einem Rechteck

$$\hat{m}_{ij} = n\hat{p}_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}. \quad (7.17)$$

Die Teststatistik $T : \mathcal{X} \rightarrow [0, \infty)$ mit

$$T(x_1, \dots, x_n) = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}};$$

es gilt $T(X_1, \dots, X_n) \sim \chi_{(I-1)(J-1)}^2$, und die Anzahl der Freiheitsgrade ergibt sich aus $(I \cdot J - 1) - (I - 1) - (J - 1) = (I - 1)(J - 1)$.

Damit haben wir den Schwellenwert $c = \chi_{(I-1)(J-1), 1-\alpha}^2$ und

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\pi} \left(T(X_1, \dots, X_n) > \chi_{(I-1)(J-1), 1-\alpha}^2 \right) = \alpha. \quad (7.18)$$

Beispiel 7.9: Wir nehmen folgende Umfrage an: n Wahlberechtigte in Baden-Württemberg werden befragt, welche Partei sie in der kommenden Landtagswahl wählen wollen und welche sie 2006 gewählt haben. Dann haben wir $I = 5$ (CDU, SPD, Grüne, FDP, andere) und auch $J = 5$ und die einzelnen Befragten werden im jeweiligen Rechteck zusammengezählt.

⁴Diese sind die ML-Schätzer

Abbildungsverzeichnis

1.1	Stochastiker in der Zeit. (Die Bilder entstammen Wikimedia-Commons und unterliegen der CC-Lizenz.)	6
2.1	Venn-Diagramme	8
4.1	Reiskorn Beispiel zur Poisson-Approximation (Das Bild entstammt Wikimedia-Commons und unterliegt der CC-Lizenz.)	56