Einführung in die Stochastik für Informatiker
Sommersemester 2000
Prof. Mathar

geTEXt von

René Wörzberger
rene@woerzberger.de

Bilder

Thorsten Uthke

Review

Diego Biurrun
diego@pool.informatik.rwth-aachen.de

15. Januar 2001

Inhaltsverzeichnis

Kapitel 1

Einführung

Betrachte „Zufallsexperimente“, z. B.

- Münzwurf, Würfelwurf, Spiele, Roulette, Lotto
- Ankunft von Kunden an Schaltern, Pakete in Netzwerken
- Input für Algorithmen
- Signale, die von einer Quelle ausgesendet werden
- Positionierung von Mobilstationen in Zellnetzen

Diesen Beispielen gemeinsam ist, daß die interessierenden Größen nicht vorhergesagt werden können und zufallsabhängig sind.

Definition 1.1 (Stochastik) *Stochastik* ist die mathematische Behandlung von Zufallsphänomenen ($\hat{o}$ $\sigma\tau o\chi o\varsigma$: das Vermutete)

Die Stochastik umfaßt mehrere Teilgebiete, die sich in etwa so aufteilen lassen

- Wahrscheinlichkeitstheorie
  - theoretisch (stochastische Prozesse, Grenzwertsätze, stochastische Differentialgleichungen)
  - angewandt (stochastische Modellierung, z. B., Warteschlangen, Zuverlässigkeitstheorie, stochastische Signalerkennung)
- Mathematische Statistik (mit vielen Teilgebieten)

Diese Vorlesung legt die Schwerpunkte auf angewandte Wahrscheinlichkeitstheorie (stochastische Modellierung) mit Betonung der Anwendung in der Informatik.

Ziel der Vorlesung:
Bereitstellung von Grundlagen als Basis für weiterführende Veranstaltungen z.B. Warteschlangensysteme I + II, Informationstheorie I + II, Kryptologie, stochastische Simulation, Zufallsgesteuerte Optimierungsverfahren

Historische Entwicklung:

Fermat	(1601-1665)	
Pascal	(1623-1662)	
Bernoulli	(1654-1705)	
Laplace	(1749-1827)	Kombinatorischer Zugang, motiviert durch Spielprobleme, relative Häufigkeiten als Wahrscheinlichkeiten
Kolmogoroff	(1973)	axiomatische Entwicklung

Kapitel 2

σ-Algebren und Wahrscheinlichkeitsverteilungen

Mathematische Beschreibung von Zufallsexperimenten mit Mengen. Betrachte relevante *Ergebnisse* und fasse diese zu einer Menge zusammen. Man nennt Ω (Omega) die *Ergebnismenge* (oft: $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}^n, \{0,1\}^n$). *Ereignisse* A werden beschrieben durch Teilmengen von Ω ($A \subseteq \Omega$). Die Menge aller Teilmengen der Ergebnismenge $\mathfrak{P}(\Omega)$ heißt *Ereignismenge*. Die *Wahrscheinlichkeit von Ereignissen* wird durch eine Funktion $P : \mathfrak{P}(\Omega) \to [0,1]$ beschrieben mit

1. $P(\Omega) = 1,\ P(\emptyset) = 0$ $\quad (*)$
2. $P(A^\complement) = 1 - P(A) \quad \forall A \in \mathfrak{P}(\Omega)$
3. $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathfrak{P}(\Omega)$ *mit* $A \cap B = \emptyset$
4. $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{P}(\Omega),\ A_i \cap A_j = \emptyset \quad \forall i, j \quad i \neq j$ $\quad (*)$

Aus den Eigenschaften $(*)$ lassen sich die anderen herleiten.
Sprechweisen:

- $A^\complement$: „A tritt nicht ein“, $A \in \mathfrak{P}(\Omega)$
- $A \cup B$: „A oder B treten ein“, $A, B \in \mathfrak{P}(\Omega)$
- $A \cap B$: „A und B treten ein“, $A, B \in \mathfrak{P}(\Omega)$

Wie erhält man nun Wahrscheinlichkeiten? Bei endlichen Ω durch Abzählen.

Definition 2.1 (Laplacescher Wahrscheinlichkeitsbegriff) Sei Ω eine endliche Menge. $|A|$ bezeichne die *Mächtigkeit* von $A \subset \Omega$. Durch

$$P(A) = \frac{|A|}{|\Omega|},\ A \in \mathfrak{P}(\Omega)$$

wird eine Wahrscheinlichkeitsverteilung auf $\mathfrak{P}(\Omega)$ mit den Eigenschaften $(*)$ definiert. P heißt *Laplace-Verteilung* oder *diskrete Gleichverteilung* über Ω.

Beispiel 2.2 (binäre Suche) Gegeben sei ein geordnetes Feld von $2^n - 1$ Elementen und ein mit diesen vergleichbares Schlüsselelement y.

Problemstellung: Ist y in einem derartigen Feld vorhanden, und an welcher Stelle?

Zur Lösung dieses Problems bietet sich beispielsweise „binary search" an.

Stochastisches Modell: Sei $\Omega = \{0, 1, ..., 2^n - 1\}$ und $\omega \in \Omega,\ \omega \geq 1$. Definiere: $\omega = 1$ falls das gesuchte Element y sich an der Stelle ω befindet. $\omega = 0$, falls y nicht im Feld vorkommt. Es lassen sich nun Ereignisse A_k bestimmen, wobei k bedeutet, daß y im k-ten Schritt gefunden wird.

$$\begin{aligned} A_1 &= \{2^{n-1}\},\ \ A_2 = \{2^{n-2},\ 3 \cdot 2^{n-2}\} \ldots \\ A_k &= \{(2j-1) \cdot 2^{n-k}\ :\ j = 1, \ldots, 2^{k-1},\ k = 1, \ldots, n\}. \end{aligned}$$

Es wird angenommen, daß jede Platznummer und die 0 gleichwahrscheinlich sind. Dann ist

$$|A_k| = 2^{k-1},\ P(A_k) = \frac{2^{k-1}}{2^n},\ 1 \leq k \leq n$$

wobei $P(A_k)$ die Wahrscheinlichkeit ist, y in genau k Schritten zu finden. Mit zusammangesetzten Ereignissen lassen sich auch andere Fragestellungen modellieren. Sei beispielsweise B_k das Ereignis, daß y in höchstens k Schritten gefunden wird.

$$\begin{aligned} B_k &= \bigcup_{j=1}^{k} A_j,\ \text{alle } A_j \text{ paarweise disjunkt} \\ P(B_k) &= P\left(\bigcup_{j=1}^{k} A_j\right) = \sum_{j=1}^{k} P(A_j) = \sum_{j=1}^{k} \frac{2^{j-1}}{2^n} = \frac{1}{2^n}(2^k - 1) = \frac{2^k - 1}{2^n}. \end{aligned}$$

Beispiel 2.3 (Hashing) Gegeben sei ein Universum U und eine gewisse Teilmenge M $(M \subseteq U)$ davon, mit $|M| = k$. Diese k Werte sollen in einem Hashfeld $a : array\ [0, ..., n-1]\ of\ type$ abgespeichert werden. Definiere dazu eine Hashfunktion $h : U \to \{0, ..., n-1\}$. Ein $x \in M$ wird dann in $a[h(x)]$ abgespeichert, was zu einer Kollision führen kann, falls ein von x verschiedenes $y \in M$ existiert mit $h(x) = h(y)$ (Kollisionsauflösung durch lineare Listen).

Stochastisches Modell: Es wird ein rein zufälliges Ablegen von k Daten in einem Feld der Länge n angenommen. Sei $S = \{1, ..., n\}$ die Menge der Speicherplätze. Die Ergebnismenge Ω mit $|\Omega| = S^k$ beinhaltet alle Arten,

k Daten abzulegen (k-n-Permutationen mit Wiederholung). Es interessiert die Menge A_{kn} aller Arten, k Daten ohne Kollisionen abzulegen. Die Wahrscheinlichkeit, daß es bei rein zufälligem Ablegen der k Daten in einem Feld der Länge n zu keiner Kollision kommt ist

$$\begin{aligned}
P(A_{kn}) &= \frac{|A_{kn}|}{|\Omega|} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\
&= \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right) \leq \exp\left(-\sum_{i=0}^{k-1} \frac{i}{n}\right) \\
&= \exp\left(-\frac{(k-1)k}{2n}\right).
\end{aligned}$$

Beispielsweise ergibt sich mit den Werten $n = 365$, $k = 23$ eine Wahrscheinlichkeit von $P(A_{kn}) \leq 0,499998$. Das bedeutet z. B., daß es „eher unwahrscheinlich“ ist, daß in einer Klasse mit 23 Schülern niemand am gleichen Tag Geburtstag hat.

Der Laplacesche Wahrscheinlichkeitsbegriff reicht nicht aus. Das hat insbesondere folgende Gründe:

- Ω ist oft unendlich oder sogar überabzählbar unendlich.
- Viele experimentelle Studien sind nicht durch diskrete Gleichverteilung beschreibbar.

Beispiel 2.4 (unendlicher Münzwurf) Man werfe eine Münze unendlich oft. Als Ergebnisse werden hier Folgen von Nullen und Einsen betrachtet, wobei die Nullen Zahl und die Einsen Kopf repräsentieren, also

$$\Omega = \{\omega = (x_1, x_2, \dots) \ : \ x_i \in \{0,1\}\} = \{0,1\}^{\mathbb{N}},$$

wobei Ω überabzählbar ist (Diagonalisierungsargument). Das Problem ist die Beschreibung einer Gleichverteilung auf Ω. Als erster Ansatz bietet sich an, jeder Folge die gleiche von Null verschiedene Wahrscheinlichkeit zuzuordnen, also: $P(\{\omega\}) = P(\{\omega'\})\ \forall \omega, \omega' \in \Omega$. Von Null verschieden bedeutet dann: $P(\{\omega\}) = \delta > 0$. Sei nun $A = \{\omega_1, \omega_2, \dots.\} \subset \Omega$ abzählbar unendlich. Dann steht

$$P(A) = \sum_{i=1}^{\infty} P(\{\omega\}) = \sum_{i=1}^{\infty} \delta = \infty$$

im Widerspruch zu $P(\Omega) = 1$. Also muß $P(\{\omega\}) = 0\ \forall \omega \in \Omega$ sein, was aber nicht sehr hilfreich ist.

Beispiel 2.5 (Gleichverteilung über $[0,1]$) Anwendbar z.B. in einem Modell, in dem jeder Zeitpunkt zwischen 0 und 1 gleichwahrscheinlich ist. Die Ergebnismenge ist also $\Omega = [0,1]$. Man definiere probehalber eine Wahrscheinlichkeitsfunktion P mit folgenden Eigenschaften:

- $P([a,b]) = b - a \quad \forall\, 0 \leq a < b \leq 1$
- P ist σ-additiv, d.h.: $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

Eine Funktion mit diesen Eigenschaften existiert nicht. Im $\mathbb{R}^3$ sogar auch dann nicht, wenn man nur endliche Additivität verlangt.

> Es gibt kein endlich additives drehinvariantes Maß auf $\mathfrak{P}(\mathbb{R}^3)$.
> (*Hausdorff* 1914)

Für einen allgemeinen Wahrscheinlichkeitsbegriff (ohne Existenzprobleme) also: Ereignismenge nicht $\mathfrak{P}(\Omega)$, kleineres Mengensystem wählen, das noch alle interessanten Ereignisse enthält.

Definition 2.6 (σ-Algebra) Sei $\Omega \neq \emptyset$ und $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathfrak{A}$ heißt *σ-Algebra* (von Ereignissen) über Ω, wenn

(i) $\Omega \in \mathfrak{A}$

(ii) $A \in \mathfrak{A} \Rightarrow A^{\complement} \in \mathfrak{A}$

(iii) $A_n \in \mathfrak{A}$, $n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

Das Paar $(\Omega, \mathfrak{A})$ heißt *Meßraum*. Mit den deMorgan-Regeln folgt desweiteren:

$$\begin{aligned} A_n \in \mathfrak{A} \quad &\Rightarrow \quad A_n^{\complement} \in \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n^{\complement} \in \mathfrak{A} \Rightarrow \left(\bigcup_{n=1}^{\infty} A_n^{\complement}\right)^{\complement} \in \mathfrak{A} \\ &\Rightarrow \quad \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A} \ \ \forall n \in \mathbb{N} \end{aligned}$$

σ-Algebren enthalten alle Ereignisse, die durch die Verknüpfungen mit „nicht“, „oder“, „und“ entstehen (auch abzählbar unendlich). Dies ist wichtig für die Festlegung von Wahrscheinlichkeitsverteilungen.

Beispiel 2.7 Beispiele und Gegenbeispiele für σ-Algebren:

- $\mathfrak{P}(\Omega)$ ist stets eine σ-Algebra (feinste σ-Algebra).
- $\{\emptyset, \Omega\}$ ist σ-Algebra (gröbste σ-Algebra).

- Sei $\Omega = \mathbb{N}$, $G = \{2, 4, 6, \ldots\}$, $U = \{1, 3, 5, \ldots\}$. Dann ist $\mathfrak{A} = \{\emptyset, G, U, \mathbb{N}\}$ eine σ-Algebra.

- Sei $\Omega = \mathbb{R}$. Dann ist $\varepsilon = \{(a, b] \, : \, a < b \in \mathbb{R}\}$ keine σ-Algebra, denn sei $a < b < c < d$. Dann ist $(a, b] \in \varepsilon$ und $(c, d] \in \varepsilon$ aber $(a, b] \cup (c, d] \notin \varepsilon$.

Problem: Gibt es eine kleinste (im Sinne der Mengeninklusion) σ-Algebra, die ε enthält?

Lemma 2.8 $\Omega \neq \emptyset$, $\mathfrak{A}_i$, $i \in I$ seien σ-Algebren über Ω. Dann ist $\bigcap_{i \in I} \mathfrak{A}_i$ ebenfalls eine σ-Algebra.
Sei $\varepsilon \subseteq \mathfrak{P}(\Omega)$. Dann heißt $\mathfrak{A}(\varepsilon) = \bigcap_{\varepsilon \subseteq \mathfrak{A}} \mathfrak{A}$ (die kleinste σ-Algebra, die ε enthält), die von ε erzeugte *σ-Algebra*.

Definition 2.9 (Borelsche σ-Algebra) Sei $\Omega = \mathbb{R}$ und $\varepsilon = \{(a, b] \, : \, a < b \in \mathbb{R}\}$. Dann heißt $\mathfrak{B}^1 = \mathfrak{A}(\varepsilon)$ *Borelsche σ-Algebra*. Auf σ-Algebren können Wahrscheinlichkeitsverteilungen mit den Eigenschaften $(*)$ definiert werden.

Definition 2.10 Sei $\Omega \neq \emptyset$, $\mathfrak{A}$ eine σ-Algebra über Ω. Eine Abbildung $P : \mathfrak{A} \to [0, 1]$ mit:

(i) $P(\Omega) = 1$

(ii) $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{A}$, paarweise disjunkt

heißt *Wahrscheinlichkeitsverteilung* oder *Wahrscheinlichkeitsmaß* auf $(\Omega, \mathfrak{A})$. $(\Omega, \mathfrak{A}, P)$ heißt *Wahrscheinlichkeitsraum*.

Der Laplacesche Wahrscheinlichkeitsbegriff ist ein Spezialfall einer σ-Algebra, denn: Wähle $\mathfrak{A} = \mathfrak{P}(\Omega)$

(i) $P(A) = \frac{|A|}{|\Omega|}$ erfüllt I., denn $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$.

(ii)

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= \frac{|\bigcup A_n|}{|\Omega|} = \frac{\sum |A_n|}{|\Omega|} = \sum_{n=1}^{\infty} \frac{|A_n|}{|\Omega|} \\ &= \sum_{n=1}^{\infty} P(A_n) \ \ \forall \, i, j : \ A_i \cap A_j = \emptyset \end{aligned}$$

$(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum. Eine Folge von Ereignissen $\{A_n\}_{n \in \mathbb{N}}$ heißt *aufsteigend* (*absteigend*), wenn $A_n \subseteq A_{n+1}$ $(A_{n+1} \subseteq A_n)$ $\forall n \in \mathbb{N}$. $\bigcup_{n=1}^{\infty} A_n$ $(\bigcap_{n=1}^{\infty} A_n) = lim_{n \to \infty} A_n$ heißt *Limes der Mengenfolge A_n*.

Lemma 2.11 (Eigenschaften von Wahrscheinlichkeitsverteilungen) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B, A_n \in \mathfrak{A}\ n \in \mathbb{N}$. Dann gilt:

a) $P(A^{\complement}) = 1 - P(A)$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

c) $A \subseteq B \Rightarrow P(A) \leq P(B)$

d) $A \subseteq B \Rightarrow P(B \backslash A) = P(B) - P(A)$

e) $\{A_n\}$ aufsteigend $\Rightarrow P\left(\lim_{n\to\infty}(A_n)\right) = \lim_{n\to\infty} P(A_n)$
$\{A_n\}$ absteigend $\Rightarrow P\left(\lim_{n\to\infty}(A_n)\right) = \lim_{n\to\infty} P(A_n)$.

Beweis zu Aussage e): Sei $\{A_n\}$ aufsteigend, d.h. $A_n \subseteq A_{n+1}\ \forall n \in \mathbb{N}$. Setze $B_1 = A_1,\ B_2 = A_2 \backslash A_1, \cdots, B_{n+1} = A_{n+1} \backslash A_n$. Dann sind alle B_n paarweise disjunkt und es gilt

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$$

Setze $A_0 = \emptyset$. Dann ist

$$\begin{aligned} P\left(\bigcup^{\infty} A_n\right) &= P\left(\bigcup^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{k\to\infty} \sum_{n=1}^{k} P(A_n \backslash A_{n-1}) \\ &= \lim_{n\to\infty} \sum_{n=1}^{k} \left[P(A_n) - P(A_{n-1})\right] = \lim_{k\to\infty} P(A_k) = P(\lim_{n\to\infty} A_n) \end{aligned}$$

$\{A_n\}$ absteigend $\Rightarrow \left\{A_n^{\complement}\right\}$ aufsteigend. Es gilt:

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - P\left(\bigcup_{n=1}^{\infty} A_n^{\complement}\right) = 1 - \lim_{n\to\infty} P\left(A_n^{\complement}\right) \\ &= \lim_{n\to\infty} \left(1 - P\left(A_n^{\complement}\right)\right) = \lim_{n\to\infty} P(A_n) \end{aligned}$$

Lemma 2.12 Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathfrak{A}$. Dann gelten:

a) *Siebformel von Poincare-Sylvester* (*inclusion-exclusion formula*)

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1<i_2\leq n} P\left(A_{i_1} \cap A_{i_2}\right) \\ &+ \sum_{1\leq i_1<i_2<i_3\leq n} P\left(A_{i_1} \cap A_{i_2} \cap A_{i_3}\right) - \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right) \end{aligned}$$

b) *Bonferroni-Ungleichung* (*Bonferroni inequality*)

$$\sum_{i=1}^{n} P(A_i) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2}) \le P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$$

Weitere obere bzw. untere Schranken ergeben sich durch Abbruch in a) nach + oder - Zeichen.

Beispiel 2.13 (Sortieren (Recontre-Problem)) Betrachte Felder der Länge n von vergleichbaren, verschiedenen Elementen. Alle Anordnungen seien gleichwahrscheinlich. Modelliere Situation wie folgt:

$$\begin{aligned} \Omega &= \{\text{Permutationen von } \{1, \dots, n\}\} \\ \mathfrak{A} &= \mathfrak{P}(\Omega) \\ P(\{\omega\}) &= \frac{1}{n!} \quad \forall \omega \in \Omega. \end{aligned}$$

Teil a: Bestimme die Wahrscheinlichkeit, daß mindestens ein Element an der richtigen Stelle ist (vorsortiert). Definiere dazu Ereignismengen A_j, welche jeweils alle Ereignisse beinhalten, bei denen Element j an der j-ten Stelle ist, wie folgt:

$$A_j = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega \ : \ \omega_j = j\}$$

Gesucht ist dann $P(A_1 \cup \ldots \cup A_n) = P\left(\bigcup_{j=1}^{n} A_j\right)$. Da die A_j sicher nicht paarweise disjunkt sind, erfolgt die Berechnung mit Hilfe der Siebformel: Sei $1 \le i_1 \le \ldots \le i_l \le n$, $l \le n$. Dann ist

$$\bigcap_{j=1}^{l} A_{i_j} = \{\omega \in \Omega \ : \ \omega_{i_j} = i_j, \ j = 1, \dots, l\}$$

die Menge aller Permutationen, bei denen die Elemente $i_1, \dots, i_l$ sich an der richtigen Stelle befinden. Die Mächtigkeit dieser Menge ist

$$\left| \bigcap_{j=1}^{l} A_{i_j} \right| = (n-l)!$$

weil man die gegebenen l Elemente auf die richtigen Positionen verteilen muß, dann aber die verbliebenen $(n-l)$ Elemente beliebig auf die restlichen $(n-l)$ Positionen verteilen darf. Die Wahrscheinlichkeit dafür, daß sich l Elemente $i_1, \dots, i_l$ auf den richtigen Positionen befinden, ist deshalb

$$P\left(\bigcap_{j=1}^{l} A_{i_j}\right) = \frac{(n-l)!}{n!} = \frac{1}{\binom{n}{l} l!}, \quad l = 1, \dots, n.$$

Außerdem ist die Mächtigkeit der Menge aller l-elementigen Teilmengen von n, also die Menge aller Möglichkeiten, zunächst l Elemente aus den vorhandenen n auszuwählen

$$|\{(i_1,\ldots,i_l)\ :\ 1 \le i_1 < \ldots < i_l \le n\}| = \binom{n}{l}.$$

Insgesamt ergibt sich für die Wahrscheinlichkeit, daß mindestens ein Element sich an der richtigen Position befindet

$$\begin{aligned}
P\left(\bigcup_{j=1}^{n} A_j\right) &= \sum_{j=1}^{n} P(A_j) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{n+1} P\left(\bigcap_{j=1}^{n} A_j\right) \\
&= \frac{n}{n} - \binom{n}{2}\frac{1}{\binom{n}{2}2!} + \cdots + (-1)^{n+1}\binom{n}{n}\frac{1}{\binom{n}{n}n!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n+1}}{n!} \\
&= 1 - \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} - \cdots + \frac{(-1)^{n+1}}{n!}\right) \\
&\overset{n\to\infty}{\longrightarrow} 1 - e^{-1} \approx 0{,}6321.
\end{aligned}$$

Erstaunlicherweise konvergiert die Wahrscheinlichkeit gegen einen festen Wert. Das bedeutet, daß die Wahrscheinlichkeit dafür, ob mindestens ein Element in einem Feld der Länge n an der richtigen Position ist, für große n fast unabhängig von n ist.

Teil b: Eine Abschätzung dafür, daß mindestens k Elemente vorsortiert sind wäre

$$\begin{aligned}
P\left(\bigcup_{1\le i_1<\ldots<i_k\le n}\bigcap_{l=1}^{k} A_{i_l}\right) &\overset{2.12.b}{\le} \sum_{1\le i_1<\ldots<i_k\le n} P\left(\bigcap_{l=1}^{k} A_{i_l}\right) \\
&= \binom{n}{k}\cdot\frac{1}{\binom{n}{k}k!} = \frac{1}{k!}.
\end{aligned}$$

Teil c: Die Wahrscheinlichkeit dafür, daß *genau* k Elemente vorsortiert sind. Nach Teil a beträgt die Wahrscheinlichkeit dafür, daß in einem Feld der Länge $n-k$ kein Element vorsortiert ist

$$1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!}.$$

Daher ist die Anzahl der Anordungen, bei denen kein Element vorsortiert ist,

$$(n-k)!\cdot\left(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!}\right).$$

Nun gibt es noch $\binom{n}{k}$ Möglichkeiten, ein Feld der Länge n in eines der Länge k und eines der Länge $n-k$ aufzuteilen. Somit ergibt sich für die Wahrscheinlichkeit, daß k Elemente vorsortiert sind

$$\begin{aligned}\frac{1}{n!}\binom{n}{k}(n-k!)\left(1-\frac{1}{1!}+\frac{1}{2!}-\ldots+\frac{(-1)^{n-k}}{(n-k)!}\right)\\ =\frac{1}{k!}\left(1-\frac{1}{1!}+\frac{1}{2!}-\ldots+\frac{(-1)^{n-k}}{(n-k)!}\right).\end{aligned}$$

Wie werden Wahrscheinlichkeiten bestimmt, wenn schon bekannt ist, daß das Ergebnis in einer bestimmten Teilmenge liegt?
Sei beispielsweise Ω eine Menge von Chips, die von zwei verschiedenen Firmen stammen und $|\Omega| = 5000$. Von Firma A stammen $|A| = 1000$ Chips und von Firma B $|B| = 4000$ Chips. Unter den 5000 Chips sind insgesamt $|D| = 300$ defekt, von denen $|A\cap D| = 100$ von Firma A und $|B\cap D| = 200$ von Firma B stammen. Von Firma A sind also 10% aller Chips defekt und von Firma B 5%. Es werde nun zufällig ein Chip gezogen (Laplace Modell). Die Wahrscheinlichkeit dafür, daß der Chip defekt ist, falls er von Firma A stammt ist

$$P(D|A)=\frac{|D\cap A|}{|A|}=\frac{\frac{|D\cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}}=\frac{P(D\cap A)}{P(A)}=\frac{\frac{100}{5000}}{\frac{1000}{5000}}=\frac{1}{10}.$$

Die Wahrscheinlichkeit, daß ein Chip von Firma A stammt, falls er defekt ist, ist

$$P(A|D)=\frac{|A\cap D|}{|D|}=\frac{\frac{|A\cap D|}{|\Omega|}}{\frac{|D|}{|\Omega|}}=\frac{\frac{100}{5000}}{\frac{300}{5000}}=\frac{1}{3}.$$

Definition 2.14 (bedingte Wahrscheinlichkeit) $(\Omega,\mathfrak{A},P)$ sei ein Wahrscheinlichkeitsraum und $A,B\in\mathfrak{A}$ sowie $P(B)>0$.

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

heißt (elementare) *bedingte Wahrscheinlichkeit* von A unter (der Hypothese) B. Durch

$$P(\bullet|B):\quad \mathfrak{A}\to[0,1]\ :\ A\mapsto P(A|B)$$

wird eine Wahrscheinlichkeitsverteilung auf $\mathfrak{A}$ definiert, die (elementare) *bedingte Verteilung* unter B.

Satz 2.15 $(\Omega,\mathfrak{A},P)$ sei ein Wahrscheinlichkeitsraum und $B_n\in\mathfrak{A},\ n\in\mathbb{N}$ eine Partition von Ω, d.h. $\bigcup_{n=1}^{\infty}B_n=\Omega$ und alle B_n paarweise disjunkt.

a) Satz von der totalen Wahrscheinlichkeit

$$\forall A \in \mathfrak{A} : \quad P(A) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)$$

Beweis :

$$P(A) = P\left(\biguplus_{n=1}^{\infty} A \cap B_n\right) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n)$$

b) Bayes-Formel
Falls $P(A) > 0$, so gilt $\forall n \in \mathbb{N}$

$$P(B_n|A) = \frac{P(A|B_n) \cdot P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}$$

Beweis :

$$P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} = \frac{P(A|B_n) \cdot P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}$$

Wichtiger Spezialfall: Gelte

$$P(A|B) = P\left(A \middle| B^{\complement}\right), \text{ falls } P(B), P\left(B^{\complement}\right) > 0$$

Die Wahrscheinlichkeit für das Eintreten von A hängt also nicht vom Eintreten von B ab. Dann gilt

$$\begin{aligned} P(A) &= P(A \cap B) + P\left(A \cap B^{\complement}\right) = P(A|B)\, P(B) + P\left(A \middle| B^{\complement}\right) \, P\left(B^{\complement}\right) \\ &= P(A|B) \cdot \left(P(B) + P\left(B^{\complement}\right)\right) = P(A|B) = \frac{P(A \cap B)}{P(B)} \\ \iff & P(A)\, P(B) = P(A \cap B). \end{aligned}$$

Diese Definition wird auf n Ereignisse $A_1, \ldots, A_n \in \mathfrak{A}$, bzw. auf Folgen von Ereignissen, erweitert.

Definition 2.16 (stochastische Unabhängigkeit) $(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathfrak{A}$ seien Ereignisse. Dann heißen $A_1, \ldots, A_n$ (gemeinsam) *stochastisch unabhängig* (s.u.), wenn sich die Wahrscheinlichkeit eines beliebigen Durchschnitts bestimmen läßt als das Produkt der Einzelwahrscheinlichkeiten:

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_k}) \quad \forall 1 \leq i_1 < \cdots < i_k \leq n \quad \forall k.$$

Eine Folge $\{A_n\}_{n \in \mathbb{N}}$ von Ereignissen heißt stochastisch unabhängig, wenn $\forall n \in \mathbb{N} : \; A_1, \cdots, A_n$ stochastisch unabhängig sind (wenn jede beliebige Auswahl stochastisch unabhängig ist).
Beachte: Aus paarweiser stochastischer Unabhängigkeit folgt nicht die stochastische Unabhängigkeit.

Lemma 2.17 $(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitraum und $A_1, \ldots, A_n \in \mathfrak{A}$. Dann gilt

$$A_1, ..., A_n \textit{ s.u.} \iff \forall B_i \in \left\{A_i, A_i^{\complement}\right\},\ i = 1, \ldots, n:\ P\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} P(B_i)$$

$$A_1, ..., A_n \textit{ s.u.} \iff \forall B_i \in \left\{A_i, A_i^{\complement}\right\},\ i = 1, \ldots, n:\ B_1, ..., B_n \textit{ s.u.}$$

Beispiel 2.18 Man betrachte ein Netzwerk wie es in Abb. 2.1 gegeben ist. Dieses Netzwerk hat 5 neuralgische Stellen $K_1, \ldots, K_5$ die mit den Wahrscheinlichkeiten $P(K_1) = 0,9,\ P(K_2) = 0,8,\ P(K_3) = 0,9,\ P(K_4) = 0,7$, $P(K_5) = 0,7$ intakt sind. Das gesamte System ist intakt, wenn mindestens

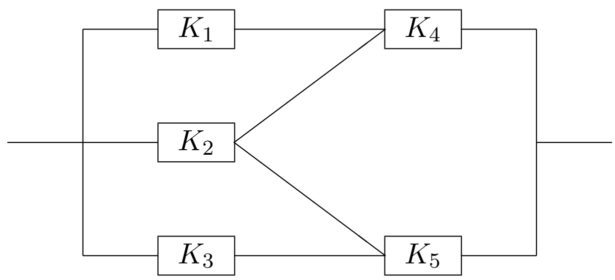

Abbildung 2.1: Netzwerk

ein Pfad intakt ist. Gesucht ist nun die Wahrscheinlichkeit, daß das System intakt ist.

Modellbildung: Alle $p_i,\ i = 1, \ldots, 5$ seien Wahrscheinlichkeiten für stochastisch unabhängige Ereignisse (Komponente i ist intakt). Dann sei die Ergebnismenge

$$\Omega = \{(x_1, \ldots, x_5)\ :\ x_i \in \{0, 1\}\} \qquad (x_i = 1 \Leftrightarrow \text{Komponente i ist intakt})$$

und die untereinander stochastisch unabhängigen Ereignisse A_i dafür, daß Komponente i intakt ist seien

$$A_i = \{(x_i, \ldots, x_5)\ :\ x_i = 1\}, \quad P(A_i) = p_i$$

Das Ereignis dafür, daß das gesamte System intakt ist, ist folglich

$$\begin{aligned} S &= (A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5) \\ P(S) &= P((A_1 \cap A_4) \cup \cdots \cup (A_3 \cap A_5)) \end{aligned}$$

S ist also die Vereinigung nicht disjunkter Mengen. Ein Möglichkeit, $P(S)$ auszurechnen, wäre die Zuhilfenahme der Sylvester-Formel, was aber sehr aufwendig ist. Mit Satz 2.15.a folgt

$$
\begin{aligned}
P(S) &= P(S|A_2)\,P(A_2) + P\left(S\,\middle|\,A_2^{\complement}\right)\,P\left(A_2^{\complement}\right) \quad \text{und} \\
P(S|A_2) &= P(A_4 \cup A_5) = 1 - P\left(A_4^{\complement} \cap A_5^{\complement}\right) = 1 - P\left(A_4^{\complement}\right)\,P\left(A_5^{\complement}\right) \\
&= 1 - (1-p_4)(1-p_5) \\
P\left(S\,\middle|\,A_2^{\complement}\right) &= P\left((A_1 \cap A_4) \cup (A_3 \cap A_5)\right) = 1 - P\left((A_1 \cap A_4)^{\complement} \cap (A_3 \cap A_5)^{\complement}\right) \\
&= 1 - (1-p_1p_4)(1-p_3p_5) \quad \text{und somit} \\
P(S) &= (1-(1-p_4)(1-p_5))p_2 + (1-(1-p_1p_4)(1-p_3p_5)) \cdot (1-p_2) \\
&= 1 - p_2(1-p_4)(1-p_5) - (1-p_2)(1-p_1p_4)(1-p_3p_5) \\
&= 0,90062
\end{aligned}
$$

Das System ist also zu ca. 90% intakt. Mit Satz 2.15.b folgt desweiteren

$$P(A_2|S) = \frac{P(S|A_2) \cdot P(A_2)}{P(S)} = \frac{0,91 \cdot 0,8}{0,90062} = 0,80833$$

die Wahrscheinlichkeit dafür, daß Komponente 2 intakt ist, falls das System intakt ist.

Betrachte Limites von Mengenfolgen, die nicht notwendig auf- oder absteigend sind.

Beispiel 2.19 (unendlicher Münzwurf) Es werde eine Münze unendlich oft geworfen. Dann ist $\Omega = \{\omega = (x_1, x_2, \ldots) \ : \ x_i \in \{0,1\}\}$. Das Ereignis, daß im n-ten Wurf Kopf fällt, ist dann $A_n = \{\omega = (x_1, x_2, \ldots) \ : \ x_n = 1\}$. Das Ereignis A sei: Es fällt unendlich oft Kopf, es treten also unendlich viele A_n ein.

$$
\begin{aligned}
A &= \{\omega \ : \ \omega \in A_n \text{ für unendlich viele n}\} \\
&= \{\omega \ : \ \forall k \exists n \geq k : \omega \in A_n\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n
\end{aligned}
$$

Analoges gilt für Ereignis B: Fast alle (alle bis auf endlich viele) A_n treten ein

$$B = \{\omega \ : \ \exists k \forall n \geq k : \omega \in A_n\} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

Definition 2.20 $(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitraum und $A_n \in \mathfrak{A}$, $n \in \mathbb{N}$.

$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \text{ heißt } \textit{Limes superior} \text{ der Mengenfolge } \{A_n\}$$

$$\liminf_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \text{ heißt } \textit{Limes inferior} \text{ der Mengenfolge } \{A_n\}$$

Satz 2.21 (Borel-Cantelli-Lemma) $(\Omega, \mathfrak{A}, p)$ sei ein Wahrscheinlichkeitsraum und $\{A_n\}_{n\in\mathbb{N}}$ eine Ereignisfolge mit $A \in \mathfrak{A},\ n \in \mathbb{N}$. Dann gelten

a)
$$\sum_{n=1}^{\infty} P(A_n) < \infty \Longrightarrow P(\limsup_{n\to\infty} A_n) = 0.$$

Beweis: Wegen der Konvergenz der Reihe gilt

$$P\left(\bigcup_{n=k}^{\infty} A_n\right) \le \sum_{n=k}^{\infty} P(A_n) \overset{k\to\infty}{\longrightarrow} 0,$$

so daß mit Lemma 2.11.e

$$P(\limsup_{n\to\infty} A_n) = P\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n\right) = \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty} A_n\right) = 0.$$

b) Ist $\{A_n\}_{n\in\mathbb{N}}$ stochastisch unabhängig, so gilt

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Longrightarrow P\left(\limsup_{n\to\infty} A_n\right) = 1$$

Beweis: Wegen Lemma 2.17 ist die Folge $\left\{A_n^{\complement}\right\}$ ebenfalls stochastisch unabhängig. Es folgt

$$\begin{aligned} P\left(\limsup_{n\to\infty} A_n\right) &= 1 - P\left(\liminf_{n\to\infty} A_n^{\complement}\right) = 1 - \lim_{k\to\infty} P\left(\bigcap_{n=k}^{\infty} A_n^{\complement}\right) \\ &= 1 - \lim_{k\to\infty}\left(\prod_{n=k}^{\infty}(1 - P(A_n))\right) = 1. \end{aligned}$$

Sei $p_n := P(A_n)$. Ist $p_n = 1$ für ein n, so gilt die letzte Gleichheit trivialerweise. Sei also $p_n < 1$ für alle $n \in \mathbb{N}$. Wir schließen mit $\ln x \le x - 1$ für alle $x > 0$

$$\prod_{n=k}^{\infty}(1 - p_n) = \exp\left(\sum_{n=k}^{\infty}\ln(1 - p_n)\right) \le \exp\left(-\sum_{n=k}^{\infty} p_n\right) \overset{k\to\infty}{\longrightarrow} 0,$$

da nach Voraussetzung $\sum_{n=k}^{\infty} p_n = \infty$ für alle $k \in \mathbb{N}$.

Beispiel 2.22 Betrachte einen unendlichen Würfelwurf, wobei die Ergebnisse der einzelnen Würfe unabhängig voneinander sind. Sei $\Omega = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_i \in \{1, 2, 3, 4, 5, 6\}\}$. Gesucht ist nun die Wahrscheinlichkeit, daß unendlich oft die Sequenz $(1, 2, 3, 4, 5, 6)$ fällt. Setze $A_n = \{\omega :$

$\omega_n = 1, \ldots, \omega_{n+5} = 6\}$, also Ereignis dafür, daß ab dem n-ten Wurf die Sequenz fällt. Die Folge $\{A_{6n}\}$ ist stochastisch unabhängig, da Überlappungen ausgeschlossen sind. Mit dem Borel-Cantelli-Lemma folgt

$$P\left(\limsup_{n\to\infty} A_n\right) \geq P\left(\limsup_{n\to\infty} A_{6n}\right) = 1,$$

$$\text{da } \sum_{n=1}^{\infty} P\left(A_{6n}\right) = \sum_{n=1}^{\infty} \frac{1}{6^6} = \infty.$$

Kapitel 3

Zufallsvariablen und ihre Verteilung

Der bisher behandelte Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ dient zur Beschreibung von Zufallsexperimenten und zur Modellierung von Zufallseinflüssen. Oft interessiert jedoch nicht das gesamte Modell, sondern nur gewisse Teilgrößen, wie z.B. 5000 Stichproben mit dem Ausgang „gut“ oder „schlecht“, wobei nur die Anzahl der „schlechten“ Stücke von Interesse ist. So ist die ursprüngliche Ergebnismenge $\Omega = \{(x_1, \ldots, x_{5000}) \ : \ x_i \in \{g, s\}\}$, von Interesse ist aber nur die Teilmenge $T = \{0, \ldots, 5000\}$.
Die Ergebnisse von Zufallsexperimenten sind oft Zahlen oder Vektoren. Dabei sind arithmetische Operationen oft sehr hilfreich. Sei beispielsweise analog zum vorhergehenden Beispiel das Experiment ein 5000-facher Münzwurf mit dem jeweiligen Ergebnis 0 oder 1 ($x_1, \ldots, x_{5000} \in \{0, 1\}$). Man interessiert sich nun für die Anzahl der Einsen, welche $\sum_{i=1}^{5000} x_i$ beträgt oder man interessiert sich für die mittlere Verzögerungszeit $\overline{x}$ in einem Switch, wobei n Verzögerungen gemessen werden. Dann gilt

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Definition 3.1 (Zufallsvariable) Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X : \ \Omega \to \mathbb{R}$ heißt *Zufallsvariable* (ZV) oder *Zufallsgröße* (engl.: *random variable (r.v.)*), wenn

$$X^{-1}(B) := \{\omega \in \Omega \ : \ X(\omega) \in B\} =: \{X \in B\} \in \mathfrak{A} \quad \forall B \in \mathfrak{B}^1.$$

Diese Bedingung[1] heißt *Meßbarkeit von X*.
Schreibweise: $X : \ (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ (X ist eine Abbildung und meßbar).
Der Begriff „Zufallsvariable“ hat sich eingebürgert, obwohl es sich eigentlich um „Funktionen“ handelt. Betont wird die Modellierung des Zufalls, — wichtig ist die *Verteilung* von Zufallsvariablen.

[1] Existenz des Urbildes

Lemma 3.2 Sei $X: \ (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ eine Zufallsvariable. Durch

$$P^X(B) := P\left(X^{-1}(B)\right) = P\left(\{w \ : \ X(\omega) \in B\}\right) =: P(X \in B), \ B \in \mathfrak{B}^1$$

wird eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}, \mathfrak{B}^1)$ definiert. P^X heißt *Verteilung* der Zufallsvariablen X.
Beweis: Es ist Definition 2.10 zu überprüfen:

i. $P^X(\mathbb{R}) = P\left(X^{-1}(\mathbb{R})\right) = P(\Omega) = 1$

ii. Seien B_n paarweise disjunkt (p.d.), $B_n \in \mathfrak{B}^1$, $n \in \mathbb{N}$. Dann gilt

$$\begin{aligned} P^X\left(\bigcup_{n=1}^{\infty} B_n\right) &= P\underbrace{\left(X^{-1}\underbrace{\left(\bigcup_{n=1}^{\infty} B_n\right)}_{\in\mathfrak{B}^1}\right)}_{\in\mathfrak{A}} = P\left(\bigcup_{n=1}^{\infty} \underbrace{X^{-1}(B_n)}_{p.d.}\right) \\ &= \sum_{n=1}^{\infty} P(X^{-1}(B_n)) = \sum_{n=1}^{\infty} P^X(B_n). \end{aligned}$$

Wesentlich in Definition 3.1 ist die „Meßbarkeit“, d.h. die Zufallsvariablen induzieren auf $(\mathbb{R}, \mathfrak{B}^1)$ die Verteilung P^X. Ein mathematisches Modell für Zufallsexperimente ist häufig

$$(\Omega, \mathfrak{A}, P) \xrightarrow{X} \left(\mathbb{R}^1, \mathfrak{B}^1, P^X\right)$$

Hierin modelliert der Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ alle Zufallseinflüsse, wobei das Wissen um seine Existenz ausreicht und genauere Kenntnisse oft nicht erforderlich sind. $(\mathbb{R}^1, \mathfrak{B}^1, P^X)$ ist ebenfalls ein Wahrscheinlichkeitsraum (in dem die bekannten Regeln gelten), der die interessierenden, beobachteten Größen modelliert. P^X ist oft bis auf Parameter bekannt.

Beispiel 3.3 (Binomialverteilung) Betrachtet werde wiederum ein n -facher Münzwurf, in dem Kopf der Eins und Zahl der Null entspreche. Die Würfe sind unabhängig voneinander, die Wahrscheinlichkeit für Kopf sei p und die für Zahl $(1-p)$ in jedem Wurf.
mathematisches Modell: $\Omega = \{\omega = (x_1, \ldots, x_n) \ : \ x_i \in \{0,1\}\}, \ \mathfrak{A} = \mathfrak{P}(\Omega)$ sei die Ergebnismenge und die Wahrscheinlichkeit, eine bestimmte Folge von Ausfällen zu erhalten ist

$$\begin{aligned} P\left(\{(x_1, \ldots, x_n)\}\right) &= \underbrace{p \cdot \ldots \cdot p}_{Anzahl\ Einsen} \cdot \underbrace{(1-p) \cdot \ldots \cdot (1-p)}_{Anzahl\ Nullen\ in\ (x_1, \cdots, x_n)} . \\ &= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i} \end{aligned}$$

Die Zufallsvariable X sei die Anzahl der Einsen, also

$$X(\omega) = X\left((x_1, \ldots, x_n)\right) = \sum_{i=0}^{n} x_i$$

mit einem Wertebereich von X: $T = \{0, 1, 2, \ldots, n\}$. Die Verteilung von X ist dann

$$\begin{aligned}
P^X(\{k\}) &= P\left(X^{-1}(\{k\})\right) = P\left(\{\omega \ : \ X(\omega) = k\}\right) = P(X = k) \\
&= P\left(\left\{(x_1, \ldots, x_n) \ : \ \sum_{i=1}^{n} x_i = k\right\}\right) \\
&= \sum_{(x_1,\ldots,x_n):\sum_{i=1}^n x_i = k} p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n-\sum_{i=1}^n x_i} \\
&= \sum_{(x_1,\ldots,x_n):\sum_{i=1}^n x_i = k} p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, \ldots, n.
\end{aligned}$$

Eine Zufallsvariable heißt *binomialverteilt* mit Parametern $n \in \mathbb{N}, \ p \in [0, 1]$, wenn

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ k = 0, 1, \ldots, n.$$

Notation: $X \sim Bin(n, p)$. $Bin(n, p)$ ist die Verteilung der Anzahl der „Treffer“ in einer Bernoulli[2]-Serie der Länge n mit Trefferwahrscheinlichkeit p.

Im Folgenden werden allgemeine Methoden zur Beschreibung von Wahrscheinlichkeitsverteilungen angegeben, — direkt formuliert für Zufallsvariablen. Mit $X = id_X$ lassen sich die folgenden Überlegungen auch direkt auf Wahrscheinlichkeitsmaße P anwenden.

3.1 Diskrete Verteilungen, Zufallsvariablen

Definition 3.4 Eine Zufallsvariable X (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$) bzw. deren Verteilung P^X heißt *diskret*, wenn eine höchstens abzählbare Menge $T = \{t_1, t_2, \ldots\}$ mit $P^X(T) = P(X \in T) = 1$ existiert. T heißt *Träger* (engl.: *support*) von X bzw. P^X.

Sei X eine diskrete Zufallsvariable mit Träger $T = \{t_1, t_2, \ldots\}, \ A \in \mathfrak{B}^1$. Dann gilt

$$P^X(A \cap T) = P^X(A) + \underbrace{P^X(T)}_{=1} - \underbrace{P^X(A \cup T)}_{=1} = P^X(A).$$

[2]Münzwurf der Länge n mit unabhängigen Würfen und einer Trefferwahrscheinlichkeit von p

Also gilt

$$\begin{aligned} P(X \in A) &= P^X(A) = P^X(A \cap T) = P^X\left(\biguplus_{i=1}^{\infty}(A \cap \{t_i\})\right) \\ &= \sum_{i:\ t_i \in A} P^X(\{t_i\}) = \sum_{i:\ t_i \in A} P(X = t_i) \end{aligned}$$

d.h., P^X, die Verteilung von X, ist eindeutig festgelegt durch $P(X = t_i),\ i = 1, 2, \ldots$.

Definition 3.5 Sei X eine diskrete Zufallsvariable mit Träger $T = \{t_1, t_2, \ldots\}$. Die Abbildung $f_X:\ T \to [0,1]$ mit $f_X(t_i) = P(X = t_i),\ i = 1, 2, \ldots$ heißt *Zähldichte* (engl.: *discrete density function*) der Zufallsvariablen X.

$X \sim Bin(n,p)$ von Beispiel 3.3 ist ein Beispiel für eine diskrete Zufallsvariable mit

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n \ (0 \le p \le 1 \text{ ist ein Parameter}).$$

Beispiel 3.6 (geometrische Verteilung) Betrachtet werde ein unendlicher, unabhängiger Münzwurf wie in Beispiel 3.3. Die Zufallsvariable sei nun aber die „Wartezeit“ bis zum ersten Auftreten der Eins, bis also das erste Mal Kopf fällt. Der Träger ist demnach wiederum $T = \mathbb{N}_0$. Die Wahrscheinlichkeit, daß erst beim k-ten Wurf die Eins fällt, ist

$$\begin{aligned} P(X = k) &= P(\{\omega = (x_1, x_2, \ldots)\ :\ x_1 = x_2 = \ldots = x_k = 0,\ x_{k+1} = 1\} \\ &= (1-p)^k p, \quad k = 0, 1, 2, \ldots \end{aligned}$$

Allgemein heißt die Verteilung mit Zähldichte

$$f_X(k) = (1-p)^k p, \quad k = 0, 1, 2, \ldots \qquad (0 < p\ \le 1 \text{ ein Parameter})$$

geometrische Verteilung. Bezeichnung:

$$X \sim Geo(p), \quad 0 < p \le 1.$$

Beispiel 3.7 (Poissonverteilung, Gesetz seltener Ereignisse) Sei $p_n \in (0,1)$ mit $np_n \overset{n\to\infty}{\longrightarrow} \lambda,\ \lambda > 0$. Dann gilt

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!} \qquad \forall k \in \mathbb{N}_0.$$

Es gilt

$$e^{-\lambda}\frac{\lambda^k}{k!} \ge 0 \text{ und } e^{-\lambda}\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}} = 1.$$

Eine diskrete Zufallsvariable X mit Zähldichte

$$f_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

heißt *Poisson-verteilt*, wobei $\lambda > 0$ ein Parameter ist. Bezeichnung:

$$X \sim Poi(\lambda), \quad \lambda > 0.$$

Als Interpretation stelle man sich das Intervall $[0, 1]$ in n Stücke gleicher Länge aufgeteilt vor. Dann ist die Wahrscheinlichkeit für das Auftreten eines Ereignisses in jedem Teilstück $\frac{\lambda}{n}$. Das Auftreten eines Ereignisses ist unabhängig.

3.2 Verteilungsfunktionen

Definition 3.8 Eine Funktion $F:\ \mathbb{R} \to [0, 1]$ mit den Eigenschaften

i. F ist monoton steigend (nicht notwendig streng monoton).

ii.
$$\lim_{x\to+\infty} F(x) = 1 \text{ und } \lim_{x\to-\infty} F(x) = 0$$

iii. F ist rechtsseitig stetig (d.h. $\forall x_0,\ x_n \downarrow x_0:\ F(x_n) \to F(x_0)$)

heißt *Verteilungsfunktion* (VF) (engl.: (cumulative) *distribution function* (cdf)).

Satz 3.9 Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $X:\ (\Omega, \mathfrak{A}) \to (\mathbb{R}^1, \mathfrak{B}^1)$ eine Zufallsvariable. Durch

$$\begin{aligned} F_X(x) = P(X \le x) &= P(\{\omega \in \Omega\ :\ X(\omega) \le x\}) \\ &= P^X((-\infty, x]),\ x \in \mathbb{R} \end{aligned}$$

wird eine Verteilung definiert, die *Verteilungsfunktion* der Zufallsvariablen X bzw. der Verteilung P^X.

Verteilungen auf $(\mathbb{R}^1, \mathfrak{B}^1)$ werden eindeutig durch Verteilungsfunktionen beschrieben.

Satz 3.10 (Eindeutigkeitssatz für Verteilungsfunktionen) Zwei Zufallsvariablen X und Y besitzen dieselbe Verteilung (auf $(\mathbb{R}^1, \mathfrak{B}^1)$) genau dann, wenn

$$F_X(x) = F_Y(x) \qquad \forall x \in \mathbb{R}.$$

Beweis: Die Hinrichtung ist einfach, die Rückrichtung benutzt den Fortsetzungssatz und den Eindeutigkeitssatz der Maßtheorie.

Beispiel 3.11 (Verteilungsfunktionen)

a) X heißt *gleichverteilt* (*rechteckverteilt*) auf [0,1] (engl.: *uniform* (*rectangular*) *distributed*), wenn

$$F_X(x) = \begin{cases} 0 & : \quad x \le 0 \\ x & : \quad 0 \le x \le 1 \\ 1 & : \quad x \ge 1 \end{cases} .$$

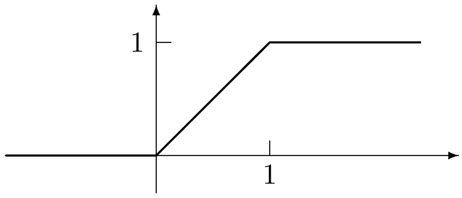

Abbildung 3.1: Rechteckverteilung

Bezeichnung:

$$X \sim R(0,1).$$

Sei nun $0 \le a < b \le 1$. Dann gilt

$$\begin{aligned} P(a < X \le b) &= P(\{\omega \ : \ a < X(\omega) \le b\}) \\ &= P(\{\omega \ : \ X(\omega) \le b\} \setminus \{\omega \ : \ X(\omega) \le a\}) \\ &= P(\{\omega \ : \ X(\omega) \le b\}) - P(\{\omega \ : \ X(\omega) \le a\}) \\ &= P(X \le b) - P(X \le a) = F_X(b) - F_X(a) \\ &= b - a. \end{aligned}$$

Die Wahrscheinlichkeit für einen zufälligen Wert im Intervall $[a,b]$ ist also gleich der Länge des Intervalls $(a,b]$, nämlich $b-a$.

Analog ist die Rechteckverteilung auf einem Intervall $(a,b]$ definiert ($X \sim R(a,b), \ a < b \in \mathbb{R}$), nämlich

$$F_X(x) = \begin{cases} 0 & : \quad x \le a \\ \frac{1}{b-a}(x-a) & : \quad a \le x \le b \\ 1 & : \quad x \ge b \end{cases}$$

Beispielsweise ist für $P(c < X \le d)$ mit $a < c < d < b$

$$\begin{aligned} P(c < X \le d) &= F_X(d) - F_X(c) = \frac{1}{b-a}(d-a) - \frac{1}{b-a}(c-a) \\ &= \frac{1}{b-a}(d-c). \end{aligned}$$

b) X heißt *exponentialverteilt* ($X \sim Exp(\lambda)$ mit Parameter $\lambda > 0$), wenn

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & : \quad x \geq 0 \\ 0 & : \quad x \leq 0 \end{cases} = (1 - e^{-\lambda x})\, \mathbb{I}_{[0,\infty)}(x).$$

In letzterer Darstellung wurde die Indikatorfunktion ($\mathbb{I}$) benutzt. Diese ist wie folgt definiert

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

c) Sei X eine diskrete Zufallsvariable mit geordnetem Träger $T = \{t_1, t_2, \ldots\} \subset \mathbb{R}$ und der Zähldichte $f_X(t_i) = p_i,\ p_i \geq 0,\ \sum_{i=1}^{\infty} p_i = 1$. Dann ist

$$\begin{aligned} F_X(x) &= P(X \leq x) = \sum_{k:\ t_k \leq x} P(X = t_k) = \sum_{t:\ t_k \leq x} p_k \\ &= \sum_{j=1}^{k-1} p_j,\ \text{falls } t_{k-1} \leq x < t_k,\ k = 1, 2, \ldots\ (t_0 = -\infty). \end{aligned}$$

Sei beispielsweise X geometrisch verteilt ($X \sim Geo(p)$), also

$$f_X(k) = (1-p)^k p, \quad k \in \mathbb{N}_0\ (0 < p \leq 1)$$

Es gilt dann

$$\sum_{j=0}^{k} (1-p)^j p = p \frac{1 - (1-p)^{k+1}}{1 - (1-p)} = 1 - (1-p)^{k+1}.$$

Also ist

$$F_X(x) = \begin{cases} 0 & : \quad x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & : \quad x \geq 0 \end{cases}$$

3.2.1 Berechnung von Wahrscheinlichkeiten durch Verteilungsfunktionen

Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x) = P(X \leq x),\ a < b \in \mathbb{R}$.

- Es gilt

$$\begin{aligned} P(a < X \leq b) &= P^X\left((-\infty, b] \setminus (-\infty, a]\right) = P^X\left((-\infty, b]\right) - P^X\left((-\infty, a]\right) \\ &= F_X(b) - F_X(a). \end{aligned}$$

- Sei $a \in \mathbb{R},\ a_n < a_{n+1} < a,\ a_n \overset{n\to\infty}{\longrightarrow} a$. Dann ist $(a_n, a]$ absteigend mit $\lim_{n\to\infty}(a_n, a] = \{a\}$ (Mengenfolge nicht Zahlenfolge) und

$$\begin{aligned} P(X=a) = &= P^X(\{a\}) = P^X\left(\lim_{n\to\infty}(a_n,a]\right) \overset{2.11e}{=} \lim_{n\to\infty} P^X((a_n,a]) \\ &= \lim_{n\to\infty}(F_X(a) - F_X(a_n)) = F_X(a) - \lim_{n\to\infty} F_X(a_n) \\ &= F_x(a) - \underbrace{F_X(a-)}_{\text{linksseitiger Grenzwert von } F_X(a)} . \end{aligned}$$

  Insbesondere gilt $P(X=a) = 0\ \forall a \in \mathbb{R}$, falls F_X stetig ist.

- Es gilt

$$\begin{aligned} P(a \le X \le b) &= P(X=a) + P(a < X \le b) \\ &= F_X(a) - F_X(a-) + F_X(b) - F_X(a) \\ &= F_X(b) - F_X(a-). \end{aligned}$$

Beispiel: Sei X die Bedienzeit von Anforderungen an einem Server und sei $X \sim Exp(\lambda),\ \lambda > 0$. Bestimme nun die Zeit x_α, unterhalb derer die Bedienzeit mit vorgegebender Wahrscheinlichkeit α liegt. Bestimme also x_α mit $P(X \le x_\alpha) = \alpha$. Sei beispielsweise $P(X \le x_\alpha) = \alpha = 0,99$. Das bedeutet, daß in 99% der Fälle die Bedienzeit kürzer als x_α ist.

Definition 3.12 Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x)$. Dann heißt

$$x_\alpha = \min\{x\ :\ F_X(x) \ge \alpha\}, \quad 0 \le \alpha \le 1$$

α-Quantil (*α-Percentil*, *$(1-\alpha)$-Fraktil*) von F_X.

$$F_X^-(t) = \min\{x\ :\ F_X(x) \ge t\}, \quad t \in (0,1)$$

heißt *Pseudoinverse* von F_X.

- Graphisch bedeutet diese Definition bei nicht stetigen Verteilungsfunktionen:

- Ist F invertierbar, so gilt $F^- = F^{-1}$ (Inverse)

$x_{\frac{1}{2}}$, also $\alpha = \frac{1}{2}$ heißt *Median* von F_X. Es gilt $P\left(X \le x_{\frac{1}{2}}\right) \ge \frac{1}{2}$ und $P\left(X < x_{\frac{1}{2}}\right) \le \frac{1}{2}$. Der Median „halbiert“ die Verteilung, in diesem Sinn ist er der „mittlere“ Wert. „Der Median ist ein Schätzer für den mittleren Wert, der robust gegen Ausreißer ist.“.

3.3 Dichten

Dieser Abschnitt behandelt Methoden zur Konstruktion von Verteilungsfunktionen und Verteilungen.

Definition 3.13 Sei die Funktion $f : \mathbb{R} \to \mathbb{R}^+$ (uneigentlich Riemann-) integrierbar mit $\int_{-\infty}^{\infty} f(x)dx = 1$. Durch

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

wird eine Verteilungsfunktion definiert. Gilt für eine Zufallsvariable X, daß $F_X(x) = F(x)$, so heißt f *(Verteilungs-) Dichte* von X (bzw. P^X) (f: *probability density function (pdf)*). X (bzw. P^X) heißt dann *absolut-stetig*.

Beispiel 3.14

a) Die Rechteckverteilung auf $[a, b]$, $a < b \in \mathbb{R}$ $(X \sim R(a, b))$

$$f(x) = \begin{cases} \frac{1}{b-a} & : \quad a \leq x \leq b \\ 0 & : \quad sonst \end{cases} = \frac{1}{b-a}\, \mathbb{I}_{[a,b]}(x),\ x \in \mathbb{R}$$

hat die Verteilungsfunktion

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \frac{1}{b-a}(x-a) & : \quad a \leq x \leq b \\ 0 & : \quad x \leq a \\ 1 & : \quad x \geq b \end{cases} .$$

Die Dichte $\hat{f}(x) = \frac{1}{b-a}\, \mathbb{I}_{(a,b)}(x)$ führt zu derselben Verteilungsfunktion F. Dichten sind nicht eindeutig, — sie sind „fast sicher eindeutig“[3].

b) Die Exponentialverteilung $(X \sim Exp(\lambda),\ \lambda > 0)$

$$f(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0,\infty)}(x),\ x \in \mathbb{R}$$

hat die Verteilungsfunktion

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, \quad x \geq 0$$

c) Die *Normalverteilung* $X \sim N(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

hat keine geschlossene Darstellung ihrer Verteilungsfunktion

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Die Berechnung erfolgt daher numerisch, approximativ oder mit Tabellen.

[3] siehe später

3.3.1 Berechung von Wahrscheinlichkeiten mit Dichten

Sei X eine Zufallsvariable mit Verteilungsfunktion F_X, Dichte f_X und $a < b \in \mathbb{R}$

- Es gilt

$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(t)dt - \int_{-\infty}^{a} f_X(t)dt \\ &= \int_{a}^{b} f_X(t)dt. \end{aligned}$$

- Es gilt $P(X = a) = F_X(a) - F_X(a^-) = 0$, da $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ stetig ist.

- Es gilt

$$P(a \leq X \leq b) = \int_{a}^{b} f_X(t)dt.$$

- Allgemein gilt bei absolut stetigen Zufallsvariablen, daß

$$P = (X \in \langle a, b \rangle) = \int_{a}^{b} f_X(t)dt,$$

wobei $\langle\rangle$ beliebig für „abgeschlossen“ oder „offen“ stehen.

- Man schreibt allgemein für Mengen $B \in \mathfrak{B}^1$

$$P(X \in B) = \int_{B} f_X(t)dt,$$

auch wenn B kein Intervall ist.

3.4 Erzeugende Funktionen und Laplace-Transformierte

In diesem Abschnitt wird eine weitere Methode zur eindeutigen Beschreibung von Verteilungen behandelt.

Definition 3.15

a) Sei X eine diskrete Zufallsvariable mit dem Träger $T = \{t_1, t_2, \ldots\}$ und Zähldichte $f_X(t_k) = p_k,\ k \in \mathbb{N}_0$. Dann heißt

$$G_X(z) := \sum_{k=0}^{\infty} p_k z^k,\ |z| < 1,$$

erzeugende Funktion von X bzw. P^X (engl.: *probability generating function*).

b) Sei X eine absolut-stetige Zufallsvariable mit Dichte f_X, wobei $f_X(x) = 0$, falls $x < 0$ (d.h. $P(X < 0) = 0$). Dann heißt

$$L_X(s) := \int_0^\infty e^{-sx} f_X(x)dx, \; s \geq 0$$

Laplace-Transformierte von X bzw. P^X.

Analog zu Satz 3.10 für Verteilungsfunktionen gilt auch hier Eindeutigkeit.

Satz 3.16

a) X und Y seien diskrete Zufallsvariablen (Verteilungen) mit demselben Träger T. X und Y besitzen dieselben Verteilungen genau dann, wenn

$$G_X(z) = G_Y(z), \quad \forall |z| \leq 1.$$

b) X und Y seien absolut-stetig mit $f_X(x) = f_Y(x) = 0$, $\forall x < 0$ X und Y haben dieselbe Verteilungen genau dann, wenn

$$L_X(s) = L_Y(s) \quad s \geq 0.$$

Beweis:

a) Eindeutigkeitssatz für Potenzreihen

b) Feller II, p. 403, Chapter XIII,1 oder Satz von Stone-Weierstraß

Satz 3.17 (Inversionsformel)

a) $G_X(z)$ sei die erzeugende Funktion einer diskreten Zufallsvariablen X mit Träger $T = \{t_1, t_2, \ldots\}$. Dann gilt

$$P(X = t_k) = \frac{1}{k!} G_X^{(k)}(0).$$

b) $L_X(s)$ sei eine Laplace-Transformierte einer absolut stetigen Zufallsvariablen X. Dann gilt

$$f_X(x) = \lim_{y \to \infty} \frac{1}{2\pi i} \int_{c-iy}^{c+iy} e^{sx} L_X(s)ds \quad \forall c \in \mathbb{R}$$

(Integration über den Weg $c - iy \to c + iy$)

Beispiel 3.18 Beispiele für erzeugende Funktionen und Laplace-Transformierte wichtiger Verteilungen:

a) Geometrische Verteilung ($X \sim Geo(p),\ 0 < p \leq 1$)

$$\begin{aligned} G_X(z) &= \sum_{k=0}^{\infty}(1-p)^k p z^k = p\sum_{k=0}^{\infty}((1-p)z)^k \\ &= \frac{p}{1-z+pz}, \quad |z| \leq 1 \end{aligned}$$

b) Poissonverteilung ($X \sim Poi(x),\ \lambda > 0$)

$$\begin{aligned} G_X(z) &= \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}z^k = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda z)^k}{k!} \\ &= e^{-\lambda}e^{\lambda z} = e^{-\lambda(1-z)}, \quad z \in \mathbb{R} \end{aligned}$$

c) Rechteckverteilung ($X \sim R(0,1)$)

$$L_X(s) = \int_0^1 e^{-sx}dx = \frac{1-e^{-s}}{s}, \quad s \geq 0$$

d) Exponentialverteilung ($X \sim Exp(\lambda)$)

$$\begin{aligned} L_X(s) &= \int_0^{\infty} e^{-sx}\lambda e^{-\lambda x}dx \\ &= \frac{\lambda}{\lambda+s}\underbrace{\int_0^{\infty}(\lambda+s)e^{-(s+\lambda)x}dx}_{=1,\ da\ Int.\ \ddot{u}ber\ Dichte} = \frac{\lambda}{\lambda+s} \end{aligned}$$

Kapitel 4

Produkträume und Zufallsexperimente

Dieses Kapitel behandelt Zufallsexperimente, die mehrere Ausgänge haben (gleichzeitig oder dasselbe Experiment mehrfach wiederholt). Man betrachtet also eine Abbildung der Form

$$(\Omega, \mathfrak{A}, P) \stackrel{X=(X_1,\dots,X_n)}{\longrightarrow} (\mathbb{R}^n, ?, ?)$$

4.1 Produkträume

Gegeben seien n Meßräume $(\Omega_i, \mathfrak{A}_i)$, $i = 1, \dots, n$ (häufig von der Form $(\mathbb{R}^1, \mathfrak{B}^1)$). Definiere nun einen neuen Meßraum Ω mit

$$\Omega = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n) \ : \ \omega_i \in \Omega_i\}.$$

Es muß nun eine $\sigma - Algebra$ über Ω gewählt werden. Der naheliegenste Ansatz wäre, diese durch

$$\varepsilon = \{A_1 \times \ldots \times A_n \ : \ A_i \in \mathfrak{A}_i, \ i = 1, \dots, n\}$$

zu definieren. ε ist aber i.A. keine σ-Algebra über Ω, denn wähle $\Omega_1 = \Omega_2 = \mathbb{R}^1$ und $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathfrak{B}^1$. Dann ist

$$\begin{aligned} c_1 &= [-1, 0] \times [-1, 0] \in \varepsilon \\ c_2 &= [0, 1] \times [0, 1] \in \varepsilon, \end{aligned}$$

aber

$$c_1 \cup c_2 \notin \varepsilon,$$

denn die Vereinigung ist nicht als kartesisches Produkt zweier Menge darstellbar.

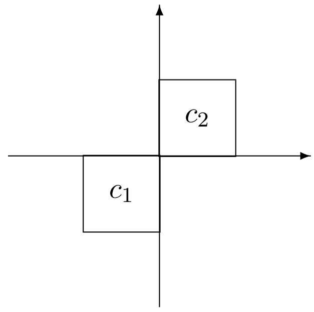

Abbildung 4.1: Mengen im $\mathbb{R}^2$

Definition 4.1 (Produkt-σ-Algebra) $(\Omega_i, \mathfrak{A}_i),\ i = 1, \ldots, n$ seien Meßräume und $\varepsilon = \{A_1 \times \ldots \times A_n \ : \ A_i \in \mathfrak{A}_i\}$. Dann heißt

$$\bigotimes_{i=1}^{n} \mathfrak{A}_i := \mathfrak{A}(\varepsilon)$$

Produkt-σ-Algebra von $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ über $\Omega = \Omega_1 \times \ldots \times \Omega_n$ und

$$\left(\Omega_1 \times \ldots \times \Omega_n, \bigotimes_{i=1}^{n} \mathfrak{A}_i \right)$$

Produkt-Meßraum des Meßraumes $(\Omega_i, \mathfrak{A}_i),\ i = 1, \ldots, n$.

Beispiel 4.2 Sei $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}^1, \mathfrak{B}^1)$ und $\varepsilon = \{B_1 \times \ldots \times B_n \ : \ B_i \in \mathfrak{B}^1\}$, dann heißt $\mathfrak{A}(\varepsilon) = \mathfrak{B}^n$ *n-dimensionale Borelsche-σ-Algebra*. Zum Beispiel wäre ein Element aus ε für $n = 2$ graphisch:

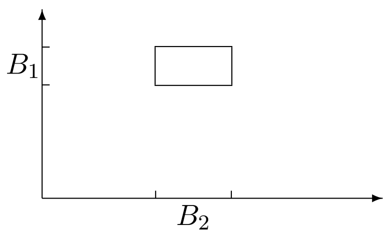

Abbildung 4.2: $B_1 \times B_2 \in \varepsilon$

Ebenfalls enthalten sind

- alle abzählbaren Vereinigungen von Rechtecken
- Linien (durch abzählbare Schnitte und Vereinigungen von Rechtecken)

- „Dreiecksflächen“ ($\{x_i \geq 0 \ : \ \sum_{i=0}^{n} x_i \leq t\}$)
- Kreisflächen ($\{x_1^2 + x_2^2 \leq c\}$)
- Approximation des Viertelkreises durch Rechtecke

4.2 Zufallsvektoren und Folgen von Zufallsvariablen

Definition 4.3 Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X = (X_1, \ldots, X_n),\ \Omega \to \mathbb{R}^n$ heißt *Zufallsvektor*, wenn $X^{-1}(B) \in \mathfrak{A} \quad \forall B \in \mathfrak{B}^n$ (Meßbarkeit). Bezeichnung:

$$X: \ (\Omega, \mathfrak{A}) \to (\mathbb{R}^n, \mathfrak{B}^n)$$

$X = (X_1, \ldots, X_n)$ ist genau dann ein Zufallsvektor, wenn alle $X_1, \ldots, X_n$ Zufallsvariablen sind.

Lemma 4.4 Sei X ein Zufallsvektor. Durch

$$P^X(B) = P\left(X^{-1}(B)\right) = P\left(\{\omega \ : \ X(\omega) \in B\}\right) =: P(X \in B), \ B \in \mathfrak{B}^n$$

wird eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}^n, \mathfrak{B}^n)$ definiert.

$$P^X = P^{(X_1, \ldots, X_n)}$$

heißt *gemeinsame Verteilung* von $(X_1, \ldots, X_n)$. Im folgenden werden gemeinsame Verteilungen mit Verteilungsfunktionen und Dichten beschrieben. Es ist klar, daß, falls $X_1, \ldots, X_n$ jeweils diskrete Verteilungen sind, auch $X = (X_1, \ldots, X_n)$ eine diskrete Verteilung mit Zähldichte wie in Kapitel 3 beschrieben ist.

Satz 4.5 Sei X ein Zufallsvektor. Dann ist P^X eindeutig bestimmt durch die (n-dimensionale) Verteilungfunktion

$$\begin{aligned}
F_X(x_1, \ldots, x_n) &= P^X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) \\
&= P\left(\{\omega \ : \ X_1(\omega) \in (-\infty, x_1) \wedge \ldots \wedge X_n(\omega) \in (-\infty, x_n)\}\right) \\
&= P\left(\{\omega \ : \ X_1(\omega) \leq x_1\} \cap \ldots \cap \{\omega \ : \ X_n(\omega) \leq x_n\}\right) \\
&= P(X_1 \leq x_1, \ldots, X_n \leq x_n)
\end{aligned}$$

Bezeichnung: $X \sim F_X$.
Die Beschreibung von Verteilungen mit n-dimensionalen Dichten ist oft einfacher.

Definition 4.6 Sei $F_X(x_1, \ldots, x_n)$ die Verteilungsfunktion des Zufallsvektors X. Eine (uneigentlich Riemann) integrierbare Funktion

$$f_X: \ \mathbb{R}^n \to \mathbb{R}^+$$

heißt *Dichte* von X, wenn

$$F_X(x_1,\ldots,x_n) = \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1} f_x(t_1,\ldots,t_n)\,dt_1\ldots dt_n \quad \forall x_1,\ldots,x_n \in \mathbb{R}.$$

X (bzw. F_X oder P^X) heißt dann *absolut-stetig* mit Dichte f_X.

Bezeichnung: $X \sim f_X$.

Sprechweisen: „X (hat|besitzt) Dichte f_X" oder „X ist verteilt nach Verteilungfunktion F_X".

Berechnung von Wahrscheinlichkeiten mit Dichten:
Sei $X = (X_1,\ldots,X_n)$ absolut-stetig mit Verteilung P^X, Verteilungsfunktion F_X und Dichte f_X. Dann ist

$$\begin{aligned} P^X(\langle a_1,b_1\rangle \times \ldots \times \langle a_n,b_n\rangle) &= P(a_1 \le X_1 \le b_1,\ldots,a_n \le X_n \le b_n) \\ &= \int_{a_n}^{b_n}\cdots\int_{a_1}^{b_1} f_X(t_1,\ldots,t_n)dt_1\ldots dt_n \\ &\quad \forall a_i \le b_i \in \mathbb{R}, \end{aligned}$$

wobei „$\langle\rangle$" beliebig für „()", „[]", „[)" oder „(]" stehen.

Allgemein für $B \in \mathfrak{B}^n$ gilt

$$P^X(B) = \underset{B}{\int\cdots\int} f_X(t_1,\ldots,t_n)dt_1\ldots dt_n.$$

Mit dem Hauptsatz der Differential- und Integralrechung folgt, daß

$$f_X(x_1,\ldots,x_n) = \frac{\partial F_X(x_1,\ldots,x_n)}{\partial x_1,\ldots,\partial x_n}$$

in allen Stetigkeitspunkten $(x_1,\ldots,x_n)$ von f.

Beispiel 4.7 Im folgenden wird die Indizierung mit X zur Vereinfachung weggelassen.

a) Die Gleichverteilung auf $T \in \mathfrak{B}^n$ ist gegeben durch

$$f(x_1,\ldots,x_n) = \frac{1}{c}\mathbb{I}_T(x_1,\ldots,x_n) = \begin{cases} \frac{1}{c} & : \ (x_1,\ldots,x_n) \in T \\ 0 & : \ sonst \end{cases}$$

$$\text{mit } c := \int\limits_T \cdots \int dx_1 \ldots dx_n < \infty$$

- auf dem Einheitswürfel des $\mathbb{R}^n$ gilt dann beispielsweise

$$T = \{(x_1,\ldots,x_n) \ : \ 0 \le x_i \le 1, \ i = 1,\ldots,n\} = [0,1]^n$$
$$f(x_1,\ldots,x_n) = \mathbb{I}_T(x_1,\ldots,x_n)$$
$$F(x_1,\ldots,x_n) = \begin{cases} 0 & : \ \exists i : \ x_i < 0 \\ \min\{x_1,1\}\cdot\ldots\cdot\min\{x_n,1\} & : \ sonst \end{cases}$$

- auf der Einheitskugel des $\mathbb{R}^n$

$$T = \left\{(x_1,\ldots,x_n) \ : \ \sum_{i=1}^n x_i^2 \le 1\right\}, \quad c = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)}$$

$$\Gamma(n) = (n-1)! \ n \in \mathbb{N}, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(x+1) = x\Gamma(x)$$

$$f(x_1,\ldots,x_n) = \frac{\Gamma\left(\frac{n}{2}+1\right)}{\pi^{\frac{n}{2}}}\mathbb{I}_T(x_1,\ldots,x_n), \quad (x_1,\ldots,x_n) \in \mathbb{R}^n$$

  Die Verteilungsfunktion ist schwierig zu berechnen. Man benötigt dazu Dirichlet-Integrale.

b) Für die n-dimensionale Normalverteilung $N(\mu,\Sigma)$ mit $\mu \in \mathbb{R}^n$ und positiv definite Matrix $\Sigma \in \mathbb{R}^{n\times n}$ gilt

$$f(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}} \ \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$
$$x = (x_1,\ldots,x_n)^T \in \mathbb{R}^n$$

c) $f_1,\ldots,f_k : \ \mathbb{R}^n \to \mathbb{R}^+$ seien n-dimensionale Dichten (auch $n=1$) und $\alpha_1,\ldots,\alpha_k \ge 0$ mit $\sum_{i=1}^k \alpha_i = 1$. Dann ist

$$f(x) = \sum_{i=1}^k \alpha_i f_i(x), \ x = (x_1,\ldots,x_n) \in \mathbb{R}^n$$

eine n-dimensionale Dichte. $f(x)$ heißt *Mischung* der Dichten $f_1,\ldots,f_n$.

Definition 4.8 (stochastische Unabhängigkeit) $X_1,\ldots,X_n$ seien Zufallsvariablen (somit ist $X = (X_1,\ldots,X_n)$ ein Zufallsvektor). $X_1,\ldots,X_n$ heißen *stochastisch unabhängig*, wenn

$$F_X(x_1,\ldots,x_n) = F_{X_1}(x_1)\cdot\ldots\cdot F_{X_n}(x_n) \quad \forall (x_1,\ldots,x_n)\in\mathbb{R}^n.$$

Also sind $X_1,\ldots,X_n$ stochastisch unabhängig, wenn

$$\begin{aligned} & P(X_1\le x_1,\ldots,X_n\le x_n) = P(X_1\le x_1)\cdot\ldots\cdot P(X_n\le x_n)\\ \Leftrightarrow\quad & P\left(\{\omega\ :\ X_1(\omega)\le x_1\}\cap\ldots\cap\{\omega\ :\ X_n(\omega)\le x_n\}\right)\\ =\quad & P\left(\{\omega\ :\ X_1(\omega)\le x_1\}\right)\cdot\ldots\cdot P\left(\{\omega\ :\ X_n(\omega)\le x_n\}\right)\ \forall x_i \end{aligned}$$

Das heißt $X_1,\ldots,X_n$ sind stochastisch unabhängig genau dann, wenn die Ereignisse

$$\{\omega\ :\ X_i(\omega)\le x_i\},\ i=1,\ldots,n,\ \forall x_1,\ldots,x_n$$

stochastisch unabhängig im Sinne der Definition 2.16 sind. Bestimmte $i_1,\ldots,i_k$ lassen sich auswählen, durch Setzen der anderen $x_j=\infty$.

Lemma 4.9 Sei $(X_1,\ldots,X_n)$ ein absolut-stetiger Zufallsvektor mit Dichte $f_{(X_1,\ldots,X_n)}$. Dann hat jedes der X_i eine Dichte

$$f_{X_i}(x) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n.$$

Wenn gilt

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

sind $X_1,\ldots,X_n$ stochastisch unabhängig. Sind umgekehrt die $X_1,\ldots,X_n$ stochastisch unabhängig mit Dichten f_{X_i}, so ist

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

eine Dichte des Zufallsvektors $(X_1,\ldots,X_n)$.
Beweis:

$$f_{X_i} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n.$$

ist eine Dichte von X_i, da

$$F_{X_i}(x) = P(X_i\le x) = \int_{-\infty}^{x} f_{X_i}(t_i)dt_i.$$

Es gilt

$$\begin{aligned} F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) &= \int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1}\prod_{i=1}^n f_{X_i}(t_i)dt_1\ldots dt_n\\ = \quad \prod_{i=1}^n\int_{-\infty}^{x_i} f_{X_i}(t_i)dt_i &= \prod_{i=1}^n F_{X_i}(x_i) \quad \forall x_1,\ldots,x_n\in\mathbb{R}. \end{aligned}$$

Also sind die $X_1, \ldots, X_n$ stochastisch unabhängig (siehe Definition 4.8).
Sind umgekehrt die $X_1, \ldots, X_n$ stochastisch unabhängig, so folgt

$$\begin{aligned} F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) &= \prod_{i=1}^n F_{X_i}(x_i) \\ = \prod_{i=1}^n \int_{-\infty}^{x_i} f_{X_i}(t_i)dt_i &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \prod_{i=1}^n f_{X_i}(t_i)dt_i. \end{aligned}$$

Somit ist

$$\prod_{i=1}^n f_{X_i}(t_i)$$

eine Dichte von $(X_1, \ldots, X_n)$.

Im Fall von Zähldichten vereinfacht sich Lemma 4.9 zu

Lemma 4.10 Seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen mit Trägern $T_1, \ldots, T_n$. Dann sind die $X_1, \ldots, X_n$ genau dann stochastisch unabhängig, wenn

$$P(X_1 = t_1, \ldots, X_n = t_n) = \prod_{i=1}^n P(X_i = t_i) \quad \forall t_i \in T_i.$$

Definition 4.11 Eine Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ heißt *stochastisch unabhängig*, wenn die $X_1, \ldots, X_n$ stochastisch unabhängig sind für alle $n \in \mathbb{N}$. Besitzen stochastisch unabhängige Zufallsvariablen alle dieselbe Verteilung, so heißen sie *stochastisch unabhängig, identisch verteilt*, kurz: *stid* (i.i.d. oder iid: „independent identically distributed“.).

Beispiel 4.12

a) Seien X_1, X_2 stochastisch unabhängig, identisch binomialverteilt (X_1, X_2 *stid* $\sim Bin(n,p)$. Dann gilt

$$\begin{aligned} P(X_1 = X_2) &= P^{(X_1,X_2)}(\{(i,j) \ : \ 0 \le i,j \le n, \ i = j\}) \\ = \sum_{i=0}^n P^{(X_1,X_2)}(\{(i,i)\}) &= \sum_{i=0}^n P(X_1 = i, X_2 = i) \\ = \sum_{i=0}^n P(X_1 = i)P(X_2 = i) &= \sum_{i=0}^n \left[\binom{n}{i} p^i (1-p)^{n-i}\right]^2. \end{aligned}$$

b) Seien X_1, X_2 stochastisch unabhängig, absolut-stetig mit Dichten f_1, f_2. Dann gilt

$$\begin{aligned} P(X_1 = X_2) &= \int_{x_1=x_2}\int f(x_1,x_2)dx_1dx_2 \\ &= \int_{-\infty}^{\infty} \underbrace{\int_{x_2}^{x_2} f_1(x_1)dx_1}_{=0} f_2(x_2)dx_2 = 0. \end{aligned}$$

Beispiel 4.13 Seien $X_1, \dots, X_n$ *stid* $\sim F$ (Verteilungsfunktion). Dann gilt

a) Sei $Y : \Omega^n \to \mathbb{R} : (\omega_1, \dots, \omega_n) \mapsto \max\{X_1(\omega_1), \dots, X_n(\omega_n)\}$. Dies könnte z.B. interpretiert werden, als das Maximum der „Laufzeiten" von $X_1, \dots, X_n$. Dann gilt

$$P(Y \le x) = P(X_1 \le x, \dots, X_n \le x) = \prod_{i=1}^{n} P(X_i \le x) = F^n(x) \quad x \in \mathbb{R}.$$

Also hat $Y = \max\{X_1, \dots, X_n\}$ die Verteilungsfunktion $F^n(x)$

b) Sei $Y : \Omega^n \to \mathbb{R} : (\omega_1, \dots, \omega_n) \mapsto \min\{X_1(\omega_1), \dots, X_n(\omega_n)\}$. Dann gilt

$$P(Y > x) = P(X_1 > x, \dots, X_n > x) = (1 - F(x))^n.$$

Also hat $Y = \min\{X_1, \dots, X_n\}$ die Verteilungsfunktion $1 - (1 - F(x))^n$. Beispielsweise gilt für X_i *stid* $\sim Exp(\lambda)$, $(\lambda > 0)$

$$P(Y \le x) = 1 - e^{-n\lambda x}$$

d.h. $Y \sim Exp(n\lambda)$

Beispiel 4.14 Seien die Zufallsvariablen X_1, X_2 stochastisch unabhängig, identisch exponentialverteilt (X_1, X_2 *stid* $Exp(\lambda)$, $\lambda > 0$). Dann gilt

$$\begin{aligned} & P(X_1 + X_2 \le z) \\ = & \int_{0 \le x_1 + x_2 \le z}\int \lambda e^{-\lambda x_1} \cdot \mathbb{I}_{(0,\infty)}(x_1) \cdot \lambda e^{-\lambda x_2} \cdot \mathbb{I}_{(0,\infty)}(x_2)dx_1dx_2 \\ = & \int_0^z \int_0^{z-x_2} \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}dx_1dx_2 = \int_0^z \lambda e^{-\lambda x_2}\left[1 - e^{-\lambda(z-x_2)}\right]dx_2 \\ = & \int_o^z \left(\lambda e^{-\lambda x_2} - \lambda e^{-\lambda z}\right)dx_2 = 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} \\ = & F_{X_1+X_2}(z), \quad z \ge 0. \end{aligned}$$

Durch Differenzieren erhält man die Dichte

$$f_{X_1+X_2}(z) = \lambda e^{-\lambda z} - \lambda\left(e^{-\lambda z} - \lambda z e^{-\lambda z}\right) = \lambda^2 z e^{-\lambda z}, \quad z \ge 0.$$

Die Verteilung der Summe von stochastisch unabhängigen, identisch $Exp(\lambda)$-verteilten Zufallsvariablen ist absolut-stetig mit Dichte

$$f(x) = \lambda^2 x e^{-\lambda x} \mathbb{I}_{[0,\infty)}(x).$$

Eine allgemeine Klasse von Verteilungen, die diese Dichte als Spezialfall enthält, sind die Γ-*Verteilungen* mit Dichten

$$f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x), \quad \alpha, \lambda > 0,$$

wobei

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

das Γ-*Integral* ist (insbesondere gilt: $\Gamma(n) = (n-1)!\ \forall n \in \mathbb{N}$). Eine Zufallsvariable mit Dichte $f_{\alpha,\lambda}(x)$ heißt $\Gamma(\alpha, \lambda)$-*verteilt*.
Spezialfälle:

(i) $\alpha = 1$: $Exp(\lambda)$-Verteilungen

(ii) $\alpha = 2$: siehe oben. Verteilung der Summe von zwei stochastisch unabhängigen, identisch exponentialverteilten Zufallsvariablen.

(iii) $\alpha = n \in \mathbb{N}$: *Erlang-Verteilung* mit Parametern n, λ und

$$f_{n,\lambda}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x)$$

Beispiel 4.15 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigigen, identisch verteilten Zufallsvariablen und $B \in \mathfrak{B}^1$. Dann heißt

$$S := \min\{n \in \mathbb{N} \ : \ X_n \in B\}$$

die *erste Eintrittszeit* in B. Die Verteilung von S ist

$$\begin{aligned} P(S = k) &= P(X_1 \notin B, \ldots, X_{k-1} \notin B, X_k \in B) \\ &= P(X_1 \notin B) \cdot \ldots \cdot P(X_{k-1} \notin B) \cdot P(X_k \in B) \\ &= (1-p)^{k-1} p, \ k \in \mathbb{N}, \quad \text{wobei } p := P(X_1 \in B) \end{aligned}$$

$S - 1$ ist also geometrisch verteilt mit dem Träger $\mathbb{N}$. $S - 1 \sim Geo(p)$. Konkret betrachte man z.B. eine unabhängige Folge von Münzwürfen, so daß X_n *stid* $\sim Bin(1, p)$ und $B = \{1\}$. Dann ist

$$P(X_n \in \{1\}) = P(X_n = 1) = p.$$

$\{S = k\}$ sei das Ereignis, daß beim k-ten Wurf erstmalig die 1 (entspricht beispielsweise „Kopf“) auftritt. Setze $X = S - 1$, die Wartezeit bis zum ersten Treffer. Dann ist $X \sim Geo(p)$.

Satz 4.16 $X_1, \ldots, X_n$ seien stochastisch unabhängige Zufallsvariablen und $I, J \subseteq \{1, \ldots, n\}$ mit $I \cap J = \emptyset$. Dann sind $(X_i)_{i \in I}$ und $(X_j)_{j \in J}$ stochastisch unabhängige Zufallsvektoren.
Sind f, g meßbare Abbildungen (mit entsprechenden Bild-Meßräumen), so sind $f((X_i)_{i \in I})$ und $g((X_j)_{j \in J})$ stochastisch unabhängig.

Beispielweise gilt: $X_1, \ldots, X_4$ stochastisch unabhängig $\Longrightarrow X_1 + X_2, X_3 + X_4$ stochastisch unabhängig.

Kapitel 5

Transformationen von Zufallsvariablen und Zufallsverteilungen

Im folgenden werden allgemeine Hilfsmittel zur Berechnung der Verteilung von Funktionen von Zufallsvariablen beschrieben.

Satz 5.1 (Transformationssatz für Dichten) Sei $X = (X_1, \ldots, X_n)$ ein absolut stetiger Zufallsvektor [1] auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit Dichte f_X. Es gelte

$$f_X(x_1, \ldots, x_n) = 0 \quad \forall (x_1, \ldots, x_n) \in M^{\complement}$$

für eine offene Menge $M \subseteq \mathbb{R}^n$. Desweiteren sei

$$T: \ (\mathbb{R}^n, \mathfrak{B}^n) \to (\mathbb{R}^n, \mathfrak{B}^n)$$

eine meßbare Abbildung (d.h. $T^{-1}(B) \in \mathfrak{B}^n \ \forall B \in \mathfrak{B}^n$) mit

(i) $\widetilde{T} = T|_M$ ist injektiv (Restriktion von T auf M).

(ii) $\widetilde{T}$ ist stetig differenzierbar auf M.

(iii) Die *Funktionaldeterminante*

$$\det\left(\left(\frac{\partial \widetilde{T}_i}{\partial x_j}\right)_{1 \le i,j \le n}\right) \neq 0 \text{ auf } M.$$

[1] Beispielweise repräsentiere jedes X_i einen Router in einem Netzwerk, der jeweils gewisse Verzögerungszeiten besitzt.

Dann ist $Y = \widetilde{T}(x)$ absolut-stetig mit der Dichte

$$\begin{aligned} f_Y(y_1, \ldots, y_n) &= \frac{f_X\left(\widetilde{T}^{-1}(y_1, \ldots, y_n)\right) \cdot \mathbb{I}_{\widetilde{T}(M)}(y_1, \ldots, y_n)}{\left|\det\left(\left(\frac{\partial \widetilde{T}_i}{\partial x_j}\right)_{1\le i,j\le n}\Big|_{\widetilde{T}^{-1}(y_1,\ldots,y_n)}\right)\right|} \\ &= \left|\det\left(\left(\frac{\partial \widetilde{T}_i^{-1}}{\partial y_j}\right)_{1\le i,j\le n}\right)\right| f_X\left(\widetilde{T}^{-1}(y_1 \ldots, y_n)\right) \mathbb{I}_{\widetilde{T}(M)}(y_1, \ldots, y_n). \end{aligned}$$

Notation: Wir schreiben jetzt zur Vereinfachung T statt $\widetilde{T}$.

Beispiel 5.2 Der Zufallsvektor $X = (X_1, X_2)$ sei gleichverteilt auf dem Einheitsquadrat $Q = (0,1)^2$ mit der Dichte $f_X(x_1, x_2) = \mathbb{I}_Q(X_1, X_2)$, d.h. X_1, X_2 sind stochastisch unabhängig und beide X_i sind rechteckverteilt ($X_i \sim R(0,1)$). Desweiteren definiere

$$T(x_1, x_2) := (\underbrace{\sqrt{x_1}\cos(2\pi x_2)}_{T_1}, \underbrace{\sqrt{x_1}\sin(2\pi x_2)}_{T_2}) \quad x_1, x_2 \in Q.$$

Dann gilt

$$\begin{aligned} \det\left(\frac{\partial T_i}{\partial x_j}\right)_{1\le i,j\le 2} &= \det\begin{pmatrix} \frac{1}{2\sqrt{x_1}}\cos(2\pi x_2) & -2\pi\sqrt{x_1}\sin(2\pi x_2) \\ \frac{1}{2\sqrt{x_1}}\sin(2\pi x_2) & 2\pi\sqrt{x_1}\cos(2\pi x_2) \end{pmatrix} \\ &= \pi\cos^2(2\pi x_2) + \pi\sin^2(2\pi x_2) \\ &= \pi\left(\cos^2(2\pi x_2) + \sin^2(2\pi x_2)\right) = \pi. \end{aligned}$$

Nach Satz 5.1 besitzt $Y = T(X)$ eine Dichte

$$f_Y(y_1, y_2) = \frac{1}{\pi}\mathbb{I}_{T(Q)}(y_1, y_2)$$

mit $\widehat{K} = T(Q)$ (Einheitskreis ohne positive x-Achse). Also ist $Y = T(X)$ gleichverteilt auf dem Einheitskreis (vgl. Bsp. 4.7).

Beispiel 5.3 (Rayleigh-Verteilung) Seien X, Y zwei stochastisch unabhängige, identisch verteilte Zufallsvariablen, die normalverteilt $N(0, \sigma^2)$ sind und eine gemeinsame Dichte

$$f_{(X,Y)}(x, y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}} \quad (x, y) \in \mathbb{R}^2$$

besitzen (Produkt der Randdichten). Die Abbildung

$$T:\ (x, y) \mapsto (r, \varphi)\ :\ \mathbb{R}^2 \setminus \{x = 0\} \to (0, \infty) \times (0, 2\pi)$$

mit

$$\begin{aligned} r &= \sqrt{x^2+y^2} \quad \text{(Länge) und} \\ \varphi &= \begin{cases} \arctan \frac{y}{x} & : \ y > 0 \\ \pi + \arctan \frac{y}{x} & : \ y < 0 \end{cases} \quad \text{(Winkel) (mit } \arctan(\pm\infty) = \tfrac{\pi}{2}) \end{aligned}$$

transformiert nun auf Polarkoordinaten. Die Umkehrabbildung ist dann entsprechend $T^{-1}: \ (r,\varphi) \mapsto (r\cos\varphi, r\sin\varphi)$. Mit dem Transformationssatz kann die Dichte berechnet werden durch

$$\det\left(\frac{\partial T_i^{-1}}{\partial z_j}\right)_{1\le i,j\le 2} = \det\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = r\left(\cos^2\varphi + \sin^2\varphi\right) = r.$$

Somit besitzt $(R,\Phi) = T(X,Y)$ nach dem Transformationssatz eine Dichte

$$\begin{aligned} f_{(R,\Phi)}(r,\varphi) &= \frac{1}{2\pi\sigma^2} e^{-\frac{r^2\cos^2\varphi + r^2\sin^2\varphi}{2\sigma^2}} \cdot r \cdot \mathbb{I}_{(0,\infty)}(r)\mathbb{I}_{(0,2\pi)}(\varphi) \\ &= \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbb{I}_{(0,\infty)}(r) \cdot \frac{1}{2\pi}\mathbb{I}_{(0,2\pi)}(\varphi) \end{aligned}$$

Also sind R, Φ stochastisch unabhängig mit

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbb{I}_{(0,\infty)}(r),$$

Rayleigh-Verteilung genannt (kurz: $R \sim \mathrm{Ray}(\sigma^2)$), und

$$f_\Phi(\varphi) = \frac{1}{2\pi}\mathbb{I}_{(0,2\pi)}(\varphi),$$

wobei Φ rechteckverteilt ist ($\Phi \sim \mathrm{R}(0,2\pi)$).

Lemma 5.4 Sei $X = (X_1, X_2)$ ein Zufallsvektor, absolut-stetig mit Dichte $f_X(x_1,x_2)$. Dann ist $Y = X_1 + X_2$ absolut-stetig mit Dichte

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(t, y-t)dt, \ \ y \in \mathbb{R}.$$

Beweis: Mit dem Transformationssatz folgt, unter der Wahl von $T(x_1,x_2) = (x_1, x_1+x_2)$, $T^{-1}(y_1,y_2) = (y_1, y_2 - y_1)$

$$\det\left(\frac{\partial T_i}{\partial x_j}\right)_{1\le i,j\le 2} = \det\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1.$$

Also gilt $f_Y(y_1,y_2) = f_X(y_1, y_2 - y_1)$. Es folgt $f_Y(y) = \int_{-\infty}^{\infty} f_X(y_1, y-y_1)dy_1$ (zweite Randverteilung, Integration über erste Komponente).

Speziell folgt aus Lemma 4.9, daß, wenn X_1, X_2 stochastisch unabhängig sind (also $f_{(X_1,X_2)}(x_1,x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$), dann gilt

$$f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(y-t)dt \quad y \in \mathbb{R}.$$

Bei stochastisch unabhängigen Zufallsvariablen X_1, X_2 heißt die Verteilung von $X_1 + X_2$ *Faltung* (*convolution*) der Verteilung von X_1 und X_2.
Bezeichnung: $P^{X_1+X_2} = P^{X_1} * P^{X_2}$.

Beispiel 5.5

a) Seien X_1, X_2 zwei stochastisch unabhängige, Γ-verteilte Zufallsvariablen mit Parametern $\alpha, \beta, \lambda > 0$ ($X_1 \sim \Gamma(\alpha,\lambda)$, $X_2 \sim \Gamma(\beta,\lambda)$). Dann gilt

$$X_1 + X_2 \sim \Gamma(\alpha+\beta,\lambda)$$

d.h., $\Gamma(\alpha,\lambda) * \Gamma(\beta,\lambda) = \Gamma(\alpha+\beta,\lambda)$.

b) X_1, X_2 seien stochastisch unabhängige, normalverteilte Zufallsvariablen mit Parametern $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ $\left(X_1 \sim \mathrm{N}\left(\mu_1, \sigma_1^2\right),\right.$ $\left.X_2 \sim \mathrm{N}\left(\mu_2, \sigma_2^2\right)\right)$. Dann gilt

$$X_1 + X_2 \sim \mathrm{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right).$$

d.h.,

$$\mathrm{N}\left(\mu_1, \sigma_1^2\right) * \mathrm{N}\left(\mu_2, \sigma_2^2\right) = \mathrm{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Beweis:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1}\, e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2}\, e^{-\frac{(y-t-\mu_2)^2}{2\sigma_2^2}}\, dt \\
&= \ldots = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\, e^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}
\end{aligned}$$

c) Sei die Zufallsvariable X normalverteilt auf dem Intervall $(0,1)$ ($X \sim \mathrm{N}(0,1)$). Dann gilt

$$X^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$

Beweis: Sei $x > 0$. Dann gilt

$$\begin{aligned}
P\left(X^2 \le x\right) &= P\left(-\sqrt{x} \le X \le \sqrt{x}\right) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}} dt \\
&\quad (\text{Substitution } u = t^2,\ dt = \tfrac{1}{2\sqrt{u}} du) \\
&= 2\int_0^x \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u}{2}} \frac{1}{2\sqrt{u}} du
\end{aligned}$$

Die Dichte von X^2 ist also

$$f_{X^2}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} = \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} x^{-\frac{1}{2}} e^{-\frac{x}{2}} & : \ x \geq 0 \\ 0 & : \ \text{sonst} \end{cases} .$$

(Dichte einer $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$-Verteilung) Mit a) folgt: Seien die Zufallsvariablen $X_1, \ldots, X_n$ stid $\sim \mathrm{N}(0,1)$. Dann gilt $X_1^2 + \ldots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ heißt χ^2-*Verteilung* mit n Freiheitsgraden (Bezeichnung: χ_n^2), mit der Dichte

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \, x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{I}_{(0,\infty)}(x).$$

Mit Beispiel 5.3 folgt: Sei $Y \sim \chi_2^2 = \Gamma\left(1, \frac{1}{2}\right)$. Dann ist $X = \sqrt{Y} \sim \mathrm{Ray}(1)$.

d) (siehe Beipiel 4.14) Seien $X_1, \ldots, X_n$ stid $\sim \mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$ mit Parameter $\lambda > 0$. Mit a) folgt

$$S_n = X_1 + \ldots + X_n \sim \Gamma(n, \lambda)$$

(Erlang-Verteilung mit Parameter n, λ). Als Interpretation stelle man sich ein X_i als Bedienzeit für eine Anforderung an einen Server vor. Dann wäre S_n die Gesamtbedienzeit für n Anforderungen.

Umgekeht könnte man auch fragen, mit welcher Wahrscheinlichkeit im Intervall $[0, t]$ genau n Anforderungen bedient werden.

Definition 5.6 Sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge von stochastisch unabhängigen Zufallsvariablen (z.B. Bedienzeiten) die exponentialverteilt sind ($X_i \sim \mathrm{Exp}(\lambda), i \in \mathbb{N}, \lambda > 0$).

$$N(t) = \max \left\{ n \in \mathbb{N} \ : \ \sum_{i=1}^{n} X_i \leq t \right\} = \left| \left\{ n \in \mathbb{N} \ : \ \sum_{i=1}^{n} X_i \leq t \right\} \right|$$

heißt dann *Poisson-Prozess* mit Parameter $\lambda > 0$ (Bezeichnung: $\mathrm{PP}(\lambda)$). Die X_N heißen auch *Zwischenankunftszeiten* oder *Verweilzeiten* (*sojourn times*, *interarrival times*, *dwell times*). Die $S_n = \sum_{i=1}^{n} X_i$ heißen *Ankunftszeiten* (*arrival times*).

Lemma 5.7 Für alle $t \geq 0$ besitzt $N(t)$ eine Poissonverteilung mit Parameter λt, d.h. $N(t) \sim \mathrm{Poi}(\lambda t)$

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k \in \mathbb{N}_0$$

Beweis: Für k=0 gilt

$$P(N(t)=0) = P(X_1 > t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{\lambda t} = e^{-\lambda t}\frac{(\lambda t)^0}{0!}$$

Für $k \geq 1$ gilt

$$\begin{aligned}
P(N(t)=k) &= P\left(\underbrace{\sum_{i=1}^{k} x_i}_{S_k} \leq t, \underbrace{\sum_{i=1}^{k+1} x_i}_{S_{k+1}} > t\right) \\
&= P\left(\{S_k \leq t\} \setminus \{S_{k+1} \leq t\}\right) \\
&= P\left(\{S_k \leq t\}\right) - P\left(\{S_{k+1} \leq t\}\right) \\
&\quad (S_K \sim \mathrm{Erl}(k,\lambda), S_{k+1} \sim \mathrm{Erl}(k+1,\lambda)) \\
&= \frac{\lambda^k}{(k-1)!}\int_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda^{k+1}}{k!}\int_0^t x^k e^{-\lambda x}dx. \\
&= \frac{\lambda^k}{(k-1)!}\left(\int_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda}{k}\int_0^t x^k e^{-\lambda x}dx\right) \\
&= \frac{\lambda^k}{(k-1)!}\left(\int_0^t x^{k-1}e^{-\lambda x}dx + \int_0^t \frac{x^k}{k}\left(-\lambda e^{-\lambda x}\right)dx\right) \\
&= \frac{\lambda^k}{(k-1)!}\frac{x^k}{k}e^{-\lambda x}\Big|_0^t = \frac{\lambda^k}{k!}t^k e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}
\end{aligned}$$

Es gilt sogar für $N_{(s,t]} = N(t) - N(s),\ 0 \leq s < t$.
Zuwachs in $(s,t]$:

- $N_{(s,t]} \sim \mathrm{Poi}(\lambda(t-s))$
- $N_{(s_i,t_i]},\ i \in \mathbb{N}$ sind stochastisch unabhängig und Poisson-verteilt mit Parameter $\lambda(t_i - s_i)$, falls die Intervalle $(s_i, t_i]$ paarweise disjunkt sind. Die Zuwächse eines Poisson-Prozesses sind stochastisch unabhängige, Poisson-verteilte Zufallsvariablen.

Im folgenden werden nun Summen von diskreten Zufallsvariablen behandelt.

Lemma 5.8 Seien X_1, X_2 stochastisch unabhängige diskrete Zufallsvariablen mit Träger $\mathbb{N}_0$ und den Zähldichten f_{X_1}, f_{X_2}. Dann besitzt die Zufallsvariable $X_1 + X_2$ die Zähldichte

$$f_{X_1+X_2}(k) = \sum_{i=0}^{k} f_{X_1}(i) \cdot f_{X_2}(k-i), \quad k \in \mathbb{N}_0$$

Beweis:

$$
\begin{aligned}
P(X_1 + X_2 = k) &= P\left(\bigcup_{i=0}^{k}(\{x_1 = i\} \cap \{x_2 = k-i\})\right) \\
&= \sum_{i=0}^{k} P(X_1 = i, X_2 = k-i) \\
&= \sum_{i=0}^{k} \underbrace{P(X_1 = i)}_{f_{X_1}(i)} \cdot \underbrace{P(X_2 = k-i)}_{f_{X_2}(k-i)}
\end{aligned}
$$

Beispiel 5.9 Seien $X_1, \ldots, X_n$ stid und geometrisch verteilt mit Parameter $0 < p < 1$. Dann gilt

$$
P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n+k-1}{n-1}(1-p)^k p^n, \quad k \in \mathbb{N}_0
$$

Diese Verteilung heißt *negative Binomialverteilung* (Bezeichnung: $\overline{\text{Bin}}(n,p)$). Es gilt also

$$
\underbrace{\text{Geo}(p) * \ldots * \text{Geo}(p)}_{n \ mal} = \overline{\text{Bin}}(n,p)
$$

$X_1 + \ldots + X_n$ entspricht dabei z.B. der Wartezeit bis zum n-ten Treffer (ohne die Treffer mitzuzählen).

Lemma 5.10 Es gilt

a) $\text{Bin}(n_1,p) * \text{Bin}(n_2,p) = \text{Bin}(n_1+n_2,p)$ mit $n_1, n_2 \in \mathbb{N},\ 0 \leq p \leq 1$. Insbesondere gilt

$$
\underbrace{\text{Bin}(1,p) * \ldots * \text{Bin}(1,p)}_{m \ mal} = \text{Bin}(n,p).
$$

b) $\overline{\text{Bin}}(n_1,p) * \overline{\text{Bin}}(n_2,p) = \overline{\text{Bin}}(n_1+n_2,p)$ mit $n_1, n_2 \in \mathbb{N},\ 0 < p < 1$.

c) $\text{Poi}(\lambda_1) * \text{Poi}(\lambda_2) = \text{Poi}(\lambda_1 + \lambda_2)$.

Diese drei Verteilungsklassen sind *faltungsstabil*.

Lemma 5.11 Seien X_1, X_2 stochastisch unabhängig, absolut-stetig mit Dichten f_{X_1}, f_{X_2}, wobei $f_{X_i}(x) = 0,\ i \in \{1,2\}$, falls $x \leq 0$. Dann gilt

a) $Y = X_1 \cdot X_2$ ist absolut-stetig mit Dichte

$$
f_Y(y) = \int_0^\infty \frac{1}{t} f_{X_1}\left(\frac{y}{t}\right) f_{X_2}(t)dt \cdot \mathbb{I}_{(0,\infty)}(y).
$$

b) $Z = \frac{X_1}{X_2}$ ist absolut-stetig mit Dichte

$$f_Z(y) = \int_0^\infty t f_{X_1}(yt) f_{X_2}(t) dt \cdot \mathbb{I}_{(0,\infty)}(y).$$

Der Beweis benutzt den Transformationssatz für Dichten mit $T(x, y) = (x, x \cdot y)$ bzw. $T(x, y) = \left(x, \frac{x}{y}\right)$.

Kapitel 6

Erwartungswerte und Momente von Zufallsvariablen

Zur Motivation der in diesem Kapitel eingeführten Begriffe zunächst zwei Beispiele:

a) Betrachte ein einfaches Würfelspiel mit einem fairen Würfel. Ein mathematisches Modell für die stochastische Analyse wäre eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ und der Verteilung $P(X = i) = 1/6$, $i = 1, \ldots, 6$. Möglich wären nun folgende Problemstellugen:

- Angenommen, es werde bei jedem Würfelwurf die Anzahl der geworfenen Augen in EURO ausgezahlt. Dann ist der festgelegte Spieleinsatz die „mittlere" oder zu „erwartende" Auszahlung. Dieser ist ganz intuitiv

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

  Die „mittlere" Auszahlung beträgt dann also 3,5 EURO

- Angenommen, es werde bei jedem Würfelwurf das Quadrat der geworfenen Augen in EURO ausgezahlt. Dann ist die „zu erwartende" Auszahlung

  $$\begin{aligned} g(X) &= X^2 \\ E(g(X)) &= \frac{1}{6} \cdot 1^2 + \ldots + \frac{1}{6} 6^2 = 15\frac{1}{6} \end{aligned}$$

b) Unter dem Namen „*Petersburger Paradoxon*“ ist folgende Überlegung von *Nikolaus Bernoulli* (1695-1726) bekannt geworden. Für ein vereinfachtes Roulettespiel, bei dem die Farben „Rot“ und „Schwarz“ jeweils mit der Wahrscheinlichkeit 1/2 auftreten, gibt es eine Gewinnstrategie. Dabei wird solange gespielt, bis das erste Mal gewonnen wird. Solange man verliert, verdoppelt man den Einsatz in jedem Spiel. Angenommen man beginnt das erste Spiel mit einem Einsatz von 1 DM und gewinnt das erste Mal beim n-ten Spiel. Dann beläuft sich der Gesamteinsatz bis dahin auf $1+2+4+\ldots+2^n = 2^{n+1}-1$ DM. Man bekommt allerdings 2^{n+1} DM ausgezahlt, womit sich der Gewinn auf 1 DM beläuft.
Eine mathematisch formalere Modellierung könnte so aussehen: Sei $\{X_n\}_{n\in\mathbb{N}_0}$ eine Folge stochastisch unabhängig, identisch verteilter Zufallsvariablen, die mit Parameter $(1, \frac{1}{2})$ binomialverteilt sind ($X \sim \text{Bin}(1, 1/2)$). Dabei bedeute $\{X_n = 1\}$, daß im n-ten Spiel „Rot“ auftritt (o.B.d.A. werde immer auf „Rot“ gesetzt). Dann ist $S = \min\{n \in \mathbb{N}_0 : X_n = 1\}$ der Zeitpunkt, an dem erstmalig „Rot“ auftritt mit $S \sim \text{Geo}\left(\frac{1}{2}\right)$ und $P(S=k) = (1-p)^k p$, $k \in \mathbb{N}_0$. Die Auszahlung A ist 1 DM $A = 1$, falls $S < \infty$, also $P(A=1) = P(S<\infty) = \sum_{k=0}^{\infty} P(S=k) = 1$. Somit beträgt die erwartete Auszahlung $E(A) = 1 \cdot P(A=1) = 1$.
Diese sichere Gewinnstrategie erfordert dummerweise unendlich viel Kapital und unendlich viel Zeit (Zumindest letzteres steht aber dem Informatikstudenten an der RWTH nicht zur Verfügung.). Insofern ist die Frage nach der zu erwartenden Auszahlung bei begrenztem Kapital interessant. Angenommen das maximale Kapital beträgt $2^L - 1$ DM. Das heißt, daß höchstens $L-1$ Spiele spielbar sind. Also ist die Auszahlung $A = 1$ genau dann, wenn $S \leq L-1$ und $A = -(2^L - 1)$ genau dann, wenn $S \geq L$. Somit gilt für die Verteilung

$$\begin{aligned} P(A=1) &= P(S \leq L-1) = \sum_{k=0}^{L-1} \frac{1}{2^{k+1}} = 1 - \frac{1}{2^L} \\ P\left(A = -\left(2^L - 1\right)\right) &= P(S \geq L) = 1 - P(A=1) = \frac{1}{2^L} \end{aligned}$$

Für die erwartete Auszahlung ergibt sich also

$$E(A) = 1 \cdot \left(1 - \frac{1}{2^L}\right) - \left(2^L - 1\right) \cdot \frac{1}{2^L} = 0$$

In beiden Beispielen ist $E(X) = \sum_i i \cdot P(X=i)$ bzw. $E(g(X)) = \sum_i g(i) P(X = i)$, wobei die i die Trägerpunkte sind. Beachte bei der Erweiterung von E auf unendlich viele Trägerpunkte oder absolut-stetige Zufallsvariablen, daß die Summe bzw. das Integral wohldefiniert sind.

Definition 6.1 (Erwartungswert von Zufallsvariablen) Sei g eine reellwertige Funktion.

a) Sei X eine diskrete Zufallsvariable mit Träger $T = \{x_1, x_2, \ldots\} \subset \mathbb{R}$ und Zähldichte f. Falls

$$\sum_{i=1}^{\infty} |g(x_i)| f(x_i) < \infty,$$

so heißt

$$E(g(X)) = \sum_{i=1}^{\infty} g(x_i) f(x_i) = \sum_{i=1}^{\infty} g(x_i) P(X = x_i)$$

der *Erwartungswert* von $g(X)$.

b) Sei X absolut-stetig mit Dichte f. Falls

$$\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty,$$

so heißt

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

der *Erwartungswert* von $g(X)$

Insbesondere gilt für die Identität $g(X) = X$

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) \text{ bei diskreten Zufallsvariablen}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \text{ bei absolut-stetigen Zufallsvariablen}$$

Für beliebige Zufallsvariablen kann der Erwartungswert mit Hilfe der Verteilungsfunktion wie folgt berechnet werden.

Lemma 6.2 Sei X eine Zufallsvariable mit Verteilungsfunktion F. Falls $\int_{-\infty}^{0} F(x)dx < \infty$ und $\int_0^\infty (1 - F(x))dx < \infty$, so gilt

$$E(X) = -\int_{-\infty}^{0} F(x)dx + \int_0^{\infty} (1 - F(x))dx$$

Beweis: Nur für den Fall differenzierbarer Verteilungsfunktionen F ($F'(x) = f(x)$ ist dann eine Dichte). Es gilt

$$\int_{-\infty}^{0} F(x) \cdot 1 dx = F(x) \cdot x|_{-\infty}^{0} - \int x f(x) dx = -\int_{-\infty}^{0} x f(x) dx$$

als auch

$$\int_0^{\infty} (1 - F(x)) \cdot 1 dx = (1 - F(x)) \cdot x|_0^{\infty} - \int_0^{\infty} -f(x) \cdot x dx = \int_0^{\infty} f(x) \cdot x dx.$$

Zusammen

$$E(X) = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{\infty} xf(x)dx = -\int_{-\infty}^{0} F(x)dx + \int_{0}^{\infty} (1 - F(x))dx.$$

Beispiel 6.3 Der Erwartungswert für eine...

a) ...geometrisch verteilte Zufallsvariable X ($X \sim Geo(p)$) ($P(X = k) = f(k) = (1-p)^k p,\ k \in \mathbb{N}_0,\ 0 < p \leq 1$)

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k(1-p)^k p = \sum_{k=1}^{\infty} k(1-p)^k p = \sum_{k=0}^{\infty} (k+1)(1-p)^{k+1} p \\ &= (1-p) \underbrace{\sum_{k=0}^{\infty} k(1-p)^k p}_{E(X)} + (1-p) \underbrace{\sum_{k=0}^{\infty} (1-p)^k \cdot p}_{=1} \\ &\Rightarrow\ E(X) = \frac{1-p}{p}, \quad 0 < p \leq 1 \end{aligned}$$

b) ... exponentialverteilte Zufallsvariable X ($X \sim Exp(\lambda)$) ($f(x) = \lambda e^{-\lambda x},\ x \geq 0,\ \lambda > 0$)

$$\begin{aligned} E(X) &= \int_{0}^{\infty} x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_{0}^{\infty} + \int_{0}^{\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x}\Big|_{0}^{\infty} = \frac{1}{\lambda} \end{aligned}$$

Also ist $E(X) = \frac{1}{\lambda}$, falls X mit Parameter λ exponentialverteilt ist.

c) ...normalverteilte Zufallsvariable X ($X \sim \mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma > 0$) mit Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}$$

ist

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x+\mu) e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \underbrace{\int_{-\infty}^{\infty} xe^{-\frac{x^2}{2\sigma^2}} dx}_{=0} + \mu \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx}_{=1} = \mu \end{aligned}$$

Also ist $E(X) = \mu$, falls X normalverteilt mit Parameter μ, σ^2 ist ($X \sim N(\mu, \sigma^2)$).

Lemma 6.4 Sei X eine diskrete Zufallsvariable mit Träger $\mathbb{N}_0$. Dann gilt

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

Beweis: mit Lemma 6.2

$$E(X) = \int_0^{\infty} (1 - F(x))dx = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=1}^{\infty} P(X \geq k)$$

Satz 6.5 (Eigenschaften des Erwartungswertes) (auftretende Erwartungswerte sollen existieren)

a) $E(aX + bY) = aE(X) + bE(Y)\ \forall a, b \in \mathbb{R}$ (*Linearität* des Erwartungswertes).
Beweis: für absolut-stetige Zufallsvariablen

$$\begin{aligned} E(aX + bY) &= \iint (ax + by) f_{(X,Y)} dxdy \\ &= a \iint x f_{(X,Y)}(x,y) dxdy + b \iint y f_{(X,Y)}(x,y) dxdy \\ &= a \int x f_X(x) dx + b \int y f_Y(y) dy \\ &= aE(X) + bE(Y) \end{aligned}$$

b) $X \leq Y \Rightarrow E(X) \leq E(Y)$ (*Monotonie*)

c) Für $X = \mathbb{I}_A,\ A \in \mathfrak{A}$ gilt

$$E(X) = E(\mathbb{I}_A) = P(A)$$

Beweis:

$$E(\mathbb{I}_A) = 1 \cdot P(A) + 0 \cdot P\left(A^{\complement}\right) = P(A)$$

d) $P(|X| > c) \leq \frac{E(|X|)}{c} \quad \forall c > 0$ (*Markoff-Ungleichung*)
Beweis:

$$\forall c > 0 \quad c \cdot \mathbb{I}_{\{|X|>c\}} = \left\{ \begin{array}{ccc} c & : & |X| > c \\ 0 & : & |X| \leq c \end{array} \right\} \leq |X|$$

$$E\left(c \cdot \mathbb{I}_{\{|X|>c\}}\right) = c \cdot P(|X| > c) \leq E(|X|)$$

$$\text{Also } P(|X| > c) \leq \frac{E(|X|)}{c} \quad \forall c > 0$$

e) X, Y seien stochastisch unabhängig. Dann gilt

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Beweis: für diskrete, stochastisch unabhängige Zufallsvariablen X, Y

$$\begin{aligned} E(X \cdot Y) &= \sum_{i,j} x_i y_j f_{(X,Y)}(x_i, y_j) \\ &\overset{s.u.}{=} \sum_{i,j} x_i y_j f_X(x_i) f_Y(y_j) \\ &= \left(\sum_i x_i f_X(x_i)\right)\left(\sum_j y_j f_Y(y_j)\right) = E(X) \cdot E(Y) \end{aligned}$$

Erwartungswerte von Funktionen von Zufallsvektoren werden wie folgt berechnet.

Satz 6.6 Sei $(X_1, \ldots, X_n)$ ein Zufallsvektor und $g: \mathbb{R}^n \to \mathbb{R}^1$ eine meßbare Funktion

a) Sei $(X_1, \ldots, X_n)$ diskret mit Träger $T = \{t_1, t_2, \ldots\} \subset \mathbb{R}^n$ und Zähldichte f. Falls

$$\sum_{i=1}^{\infty} |g(t_i)| f(t_i) < \infty,$$

so gilt

$$E(g(X_1, \ldots, X_n)) = \sum_{i=1}^{\infty} g(t_i) f(t_i)$$

b) Sei $(X_1, \ldots, X_n)$ absolut-stetig mit Dichte f. Falls

$$\int \cdots \int |g(x_1 \ldots, x_n)| f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n < \infty,$$

so gilt

$$E(g(X_1, \ldots, X_n)) = \int \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

Definition 6.7 X, Y seien Zufallsvariablen. X und Y heißen *unkorreliert* (*uncorrelated*), wenn $E(X \cdot Y) = E(X) \cdot E(Y)$. Mit Satz 6.5.e folgt, daß, falls X, Y stochastisch unabhängig sind, X, Y auch unkorreliert sind (Umkehrschluß ist i.A. falsch, außer z.B. bei der Normalverteilung).

Transformationen mit besonderer Bedeutung sind

$$\begin{aligned} g(X) &= X^k, \quad k \in \mathbb{N} \\ g(X,Y) &= (X - EX)(Y - EY) \end{aligned}$$

Definition 6.8 X, Y seien Zufallsvariablen. Alle im folgenden auftretenden Erwartungswerte sollen existieren.

a) $E\left(X^k\right)$, $k \in \mathbb{N}_0$ heißt k-tes *Moment* (k^{th} *moment*) von X (speziell k=1: Erwartsungswert von X).

b) $E\left((X - EX)^k\right)$ heißt k-tes *zentrales Moment* (*central moment*) von X. Speziell für den Fall $k = 2$ heißt $E\left((X - EX)^2\right) = \operatorname{Var}(X)$ die *Varianz* (*variance*) von X. $\sqrt{\operatorname{Var}(X)}$ heißt *Standardabweichung* (*standard deviation*) von X.

c) $\operatorname{Cov}(X,Y) = E((X - EX)(Y - EY))$ heißt *Kovarianz* (*covariance*) von X und Y.

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}$$

heißt *Korrelation* (*correlation*) von X und Y.

Lemma 6.9 $X, Y, X_1, \ldots, X_n$ seien Zufallsvariablen, auftretende Momente sollen existieren.

a)

$$\begin{aligned} \operatorname{Cov}(X,Y) &= E(X \cdot Y) - E(X)E(Y) \\ \operatorname{Cov}(X,X) &= \operatorname{Var}(X) = E\left(X^2\right) - (EX)^2 \end{aligned}$$

b)

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X) \quad \forall a, b \in \mathbb{R}$$

c)

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) + 2 \cdot \sum_{i<j} \operatorname{Cov}(X_i, X_j)$$

Falls $X_1, \ldots, X_n$ paarweise unkorreliert sind, gilt

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i).$$

d) Es gilt die *Cauchy-Schwarz-Ungleichung*

$$|\operatorname{Cov}(X,Y)| \leq \sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)}.$$

Insbesondere folgt $|\operatorname{Corr}(X,Y)| \leq 1$.

Satz 6.10 $G_{X_1}(z), G_{X_2}(z)$ seien erzeugende Funktionen von diskreten, stochastisch unabhängigen Zufallsvariablen X_1, X_2, bzw. $L_{X_1}(s), L_{X_2}(s)$ die Laplace-Transformierten von absolut-stetigen stochastisch unabhängigen Zufallsvariablen $X_1, X_2 \geq 0$. Dann gilt

$$\begin{aligned} G_{X_1+X_2}(z) &= G_{X_1}(z) \cdot G_{X_2}(z),\ |z| \leq 1 \quad \text{bzw.} \\ L_{X_1+X_2} &= L_{X_1}(s) \cdot L_{X_2}(s),\ s \geq 0 \end{aligned}$$

Beweis: Es gilt

$$\begin{aligned} G_X(z) &= \sum_{k=0}^{\infty} p_k z^k = E\left(z^X\right) \\ L_X(s) &= \int_0^{\infty} e^{-sx} f(x) dx = E\left(e^{-sx}\right) \end{aligned}$$

Also für $X_1 + X_2$

$$\begin{aligned} G_{X_1+X_2}(z) &= E\left(z^{X_1+X_2}\right) \stackrel{s.u.}{=} E\left(z^{X_1} \cdot z^{X_2}\right) = E\left(z^{X_1}\right) \cdot E\left(z^{X_2}\right) \\ &= G_{X_1}(z) \cdot G_{X_2}(z), \quad |z| \leq 1 \\ L_{X_1+X_2}(s) &= E\left(e^{-s(X_1+X_2)}\right) \stackrel{s.u.}{=} E\left(e^{-sX_1} \cdot e^{-sX_2}\right) \\ &= E\left(e^{-sX_1}\right) \cdot E\left(e^{-sX_2}\right) = L_{X_1}(s) \cdot L_{X_2}(s), \quad s \geq 0 \end{aligned}$$

Beispiel 6.11

a) Sei X binomialverteilt mit Parametern n, p. Dann ist $G_X(z) = (1 - p + pz)^n,\ |z| \leq 1$. Wenn $X_1 \sim \mathrm{Bin}(n_1, p),\ X_2 \sim \mathrm{Bin}(n_2, p)$ stochastisch unabhängig sind, dann ist

$$G_{X_1+X_2}(z) = (1 - p + pz)^{n_1} \cdot (1 - p + pz)^{n_2} = (1 - p + pz)^{n_1+n_2}$$

die erzeugende Funktion einer Binomialverteilung mit Parametern $n_1 + n_2$ und p ($\mathrm{Bin}(n_1 + n_2, p)$).

b) Sei

$$X_1 \sim \text{Exp}(\lambda),\ L_{X_1}(s) = \frac{\lambda}{s+\lambda},\ s \geq 0$$
$$X_2 \sim \text{Exp}(\mu),\ L_{X_2}(s) = \frac{\mu}{s+\mu},\ s \geq 0$$

Die Dichte

$$f(x) = \left(\frac{\mu}{\mu-\lambda}\lambda e^{-\lambda x} + \frac{\lambda}{\lambda-\mu}\mu e^{-\mu x}\right)\mathbb{I}_{(0,\infty)}(x)$$

hat als Laplace-Transformierte

$$L_{X_1+X_2}(s) = \frac{\lambda}{s+\lambda}\cdot\frac{\mu}{s+\mu} = \frac{\mu}{\mu-\lambda}\frac{\lambda}{s+\lambda} + \frac{\lambda}{\lambda-\mu}\frac{\mu}{s+\mu}.$$

Also hat $X_1 + X_2$ die obige Dichte.

Satz 6.12 (Transformierte und Momente)

a) Sei $G(z) = \sum_{k=0}^{\infty} P(X=k)z^k$ die erzeugende Funktion einer diskreten Zufallsvariable X mit Träger $\mathbb{N}_0$. Dann gilt

$$E(X) = G'(1), \quad E(X(X-1)\dots(X-k+1)) = G^{(k)}(1)$$

(k-te *faktorielle Moment*). Demzufolge ist $E\left(X^2\right) = G'(1) + G''(1)$. Beachte: Es gilt $G^{(i)}(1) = \lim_{z\uparrow 1} G^{(i)}(z)$, falls $G(z)$ nicht existiert für $z > 1$.
„Beweis:“

$$G'(z) = \sum_{k=0}^{\infty} p_k k\cdot z^{k-1}, \quad |z| < 1 \quad \text{(gliedweises differenzieren)}$$

Also ist $G'(1) = \sum_{k=0}^{\infty} p_k \cdot k = E(X)$.

b) Sei $L(s) = \int_0^\infty e^{-sx}f(x)dx$ eine Laplace-Transformierte einer absolutstetigen Zufallsvariable $X \geq 0$ mit Dichte f. Dann gilt

$$E(X) = -L'(0), \quad E\left(X^k\right) = (-1)^k L^{(k)}(0).$$

Beispiel 6.13 (Berechnung von Momenten)

a) Sei X binomialverteilt mit Parameter $n, p,\ 0 \leq p \leq 1$. Berechung des Erwartungswertes (3 Methoden):

(i) $E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \ldots = np.$

(ii) Sei $X \sim \sum_{i=1}^{n} X_i$ mit X_i stid und alle $X_i \sim \text{Bin}(1, p)$. Dann ist für alle X_i der Erwartungswert $E\left(X_i\right) = 1 \cdot p + 0 \cdot (1-p) = p$ und für X

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E\left(X_i\right) = \sum_{i=1}^{n} p = np.$$

(iii) Über die erzeugende Funktion

$$\begin{aligned} G(z) = (pz + 1 - p)^n \Rightarrow G'(z) &= n(pz+1-p)^{n-1} \cdot p \\ G'(1) &= np = E(X) \end{aligned}$$

Die Varianz wird wie in (ii) berechnet:

$$\begin{aligned} & \text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right) \stackrel{*}{=} \sum_{i=1}^{n} p(1-p) \\ & = np(1-p) \\ (*) \quad & \text{Var}\left(X_i\right) = E\left(X_i^2\right) - \left(E\left(X_i\right)\right)^2 = E\left(X_i^2\right) - p^2 \stackrel{**}{=} p - p^2 \\ & = p(1-p) \\ (**) \quad & E\left(X_i^2\right) = 1^2 \cdot p + 0^2(1-p) = p \end{aligned}$$

b) Sei X exponentialverteilt mit Parameter $\lambda > 0$. Dann gilt

$$\begin{aligned} L_X(s) = \frac{\lambda}{s+\lambda}, \quad L_X^{(k)}(s) &= (-1)^k k! \frac{\lambda}{(s+\lambda)^{k+1}} \\ L_X^{(k)}(0) &= (-1)^k k! \lambda^{-k} \end{aligned}$$

Also gilt

$$\begin{aligned} E\left(X^k\right) = k!\lambda^{-k},\ E(X) = \frac{1}{\lambda},\ E\left(X^2\right) &= \frac{2}{\lambda^2} \\ \text{Var}(X) = E\left(X^2\right) - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} &= \frac{1}{\lambda^2} \end{aligned}$$

c) Sei X normalverteilt mit Parametern $\mu = 0$, $\sigma^2 = 1$. Dann ist $E(X) = 0$ und

$$\begin{aligned}
E\left(X^2\right) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-x^2/2} dx \\
&\quad \left(\text{Substitution: } \frac{x^2}{2} = y,\ dx = \frac{dy}{\sqrt{2y}}\right) \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} 2y e^{-y} \frac{1}{\sqrt{2y}} dy \\
&= \underbrace{\frac{2}{\sqrt{\pi}} \int_0^{\infty} \sqrt{y} e^{-y} dy}_{\text{Integral über Dichte einer } \Gamma\left(\frac{3}{2}, 1\right)\text{-Verteilung}} = 1
\end{aligned}$$

Es folgt: $\mathrm{Var}(X) = E\left(X^2\right) - (EX)^2 = 1 - 0 = 1$ (Varianz einer Standardnormalverteilung $= 1$). $Y = \sigma X + \mu$ besitzt die Dichte

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

d.h. $Y \sim \mathrm{N}\left(\mu, \sigma^2\right)$.
Also gilt $E(Y) = E(\sigma X + \mu) = \sigma E(X) + \mu = \mu$ und $\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \sigma^2 \mathrm{Var}(X) = \sigma^2$. Das bedeutet, daß der erste Parameter einer Normalverteilung den Erwartungswert und der zweite die Varianz darstellt.

Kapitel 7

Bedingte Verteilungen und Erwartungswerte

7.1 Diskreter Fall

Seien X, Y diskrete Zufallsvariablen mit Trägern T_X, T_Y und gemeinsamer Zähldichte

$$f_{(X,Y)}(x, y) = P(X = x, Y = y), \quad (x, y) \in T_X \times T_Y.$$

Es heißt

$$f_{X|Y}(x|y) = \begin{cases} \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & : \quad f_Y(y) > 0 \\ f_X(x)(\text{oder bel. andere Zähldichte}) & : \quad f_Y(y) = 0 \end{cases}$$

bedingte Zähldichte von X unter $Y = y$ (analog zu Definition 2.14). Beachte: $f_{X|Y}(x|y)$ ist eine Zähldichte für alle festen $y \in T_Y$. Die zugehörige Verteilung heißt *bedingte Verteilung* von X unter $Y = y$, also

$$P(X \in A|Y = y) = P^{X|Y=y}(A) = \sum_{X \in A} f_{X|Y}(x|y), \quad A \in \mathfrak{A}, y \in T_Y.$$

Der Satz von der totalen Wahrscheinlichkeit (Satz 2.15a) liefert

$$f_X(x) = P(X = x) = \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y), \quad x \in T_X \quad (*)$$

Beispiel 7.1 Seien $f_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$, also $P^{X|N=n} = \mathrm{Bin}(n, p)$ und N poissonverteilt mit Parameter λ, $N \sim Poi(\lambda)$. Als Interpretation stelle man sich bespielsweise ein zweistufiges Experiment vor, in dem zunächst die Anzahl der Münzwürfe ermittelt wird (poissonverteilt) und anschließend entsprechend oft eine Münze geworfen wird, wobei letztendlich die Anzahl der auftretenden „Köpfe“ interessiert. Für die Wahrscheinlichkeit, daß genau k „Köpfe“ auftreten gilt

$$\begin{aligned}
P(X=k) &= \sum_{n=0}^{\infty} P(X=k|N=n)P(N=n) \\
&= \sum_{n=k}^{\infty} P(X=k|N=n)P(N=n) \\
&= \sum_{n=k}^{\infty} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= e^{-\lambda} \frac{\lambda^k p^k}{k!} \underbrace{\sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!}}_{e^{\lambda(1-p)}} \\
&= e^{-\lambda} \frac{(\lambda)^k p^k}{k!} e^{\lambda(1-p)} \\
&= e^{-\lambda p} \frac{(\lambda p)^k}{k!} \sim \mathrm{Poi}(\lambda p).
\end{aligned}$$

X ist also poissonverteilt mit Parameter λp, $X \sim Poi(\lambda p)$.

Betrachte nun den Erwartungswert bzgl. der bedingten Verteilung.

$$E(g(X)|Y=y) = \sum_{x \in T_X} g(x) f_{X|Y}(x|y), \quad y \in T_Y$$

heißt *bedingter Erwartungswert* von $g(X)$ unter $Y = y$ (sofern existent). Es berechnet sich der Erwartungswert mit den bedingten Erwartungswerten wie folgt:

$$\begin{aligned}
E(g(X)) &= \sum_{x \in T_X} g(x) f_X(x) = \sum_{x \in T_X} g(x) \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y) \\
&= \sum_{y \in T_Y} \left(\sum_{x \in T_X} g(x) f_{X|Y}(x|y) \right) f_Y(y) \\
&= \sum_{y \in T_Y} E(g(X)|Y=y) f_Y(y)
\end{aligned}$$

7.2 Absolut-stetiger Fall

X, Y seien absolut-stetige Zufallsvariablen mit gemeinsamer Dichte $f_{(X,Y)}(x,y)$. Definiere die *bedingte Dichte* von X unter $Y = y$ durch

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & : \quad f_Y(y) > 0 \\ f_X(x) \text{ (oder beliebige andere Dichte)} & : \quad f_Y(y) = 0 \end{cases}$$

Beachte: $f_{X|Y}(x|y)$ ist eine Dichte für alle $y \in \mathbb{R}$. Die zugehörige Verteilung heißt *bedingte Verteilung* von X unter $Y = y$, also

$$P(X \in A|Y = y) = P^{X|Y=y}(A) = \int_A f_{X|Y}(x|y)dx, \quad A \in \mathfrak{A},\ y \in \mathbb{R}.$$

Speziell heißt

$$F_{X|Y=y}(x) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(z|y)dz$$

bedingte Verteilungsfunktion von X unter $Y = y$. Ähnlich wie im diskreten Fall (totale Wahrscheinlichkeit) gilt

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy.$$

Ebenso analog definiere

$$E(g(X)|Y = y) = \int g(x)f_{X|Y}(x|y)dx, \quad y \in \mathbb{R}$$

den *bedingten Erwartungswert* von $g(X)$ unter $Y = y$. Demnach gilt dann auch hier

$$\begin{aligned} E(g(X)) &= \int g(x)f_X(x)dx = \int g(x) \int f_{X|Y}(x|y)f_Y(y)dydx \\ &= \int \left(\int g(x)f_{X|Y}(x|y)dx \right) f_Y(y)dy \\ &= \int E(g(X)|Y = y)f_Y(y)dy. \end{aligned}$$

7.3 Gemischter Fall

Beispiel 7.2 (Wartezeit in einem Wartesystem) Sei N die Anzahl der Kunden in einer Warteschlange und $N \sim \text{Geo}(p)$. $X_n,\ n \in \mathbb{N}$ sei die Bedienzeit von Kunde n mit X_n stid $\sim \text{Exp}(\lambda)$. Gesucht ist die Gesamtwartezeit für neu ankommende Kunden $W = \sum_{i=1}^{N+1} X_i$. Bekannt ist $P^{W|N=n} = \text{Erl}(n+1, \lambda)$.

Es gilt

$$\begin{aligned}
P(W \le t) &= \sum_{n=0}^{\infty} P(W \le t | N = n) P(N = n) \\
&= \sum_{n=0}^{\infty} \int_0^t \frac{\lambda^{n+1}}{n!} y^n e^{-\lambda y} dy (1-p)^n p \\
&= \int_0^t \lambda p e^{-\lambda y} \underbrace{\sum_{n=0}^{\infty} \frac{(\lambda y(1-p))^n}{n!}}_{= e^{\lambda y(1-p)}} dy \\
&= \int_0^t \lambda p e^{-\lambda y} e^{\lambda y(1-p)} dy = \int_0^t \lambda p e^{-\lambda p y} dy \\
&= 1 - e^{-\lambda p t}, \quad t \ge 0
\end{aligned}$$

Also ist W exponentialverteilt mit Parameter λp und Erwartungswert $E(W) = \frac{1}{\lambda p}$.

Beispiel 7.3 (Ankünfte eines Poisson-Prozesses in zufälligen Intervallen) Sei $N(t)$ die Anzahl der Ankünfte in einem Zeitintervall, ein Poisson-Prozess mit Parameter $\lambda > 0$ ($N(t) \sim \text{Poi}(\lambda t)$) und Y eine exponentialverteilte Zufallsvariable mit Parameter $\mu > 0$ ($Y \sim \text{Exp}(\mu)$). N und Y seien stochastisch unabhängig. Gesucht ist die Verteilung der von $N(Y) = \max\{n \in \mathbb{N}_0 : \sum_{i=1}^n X_i \le Y\}$, wobei die X_i stid $\sim \text{Exp}(\lambda)$. Gesucht ist also die Verteilung der Anzahl der Ankünfte in einem zufälligen Zeitintervall $[0, Y]$. Bekannt ist, daß $P^{N(Y)|Y=t} = \text{Poi}(\lambda t)$ für alle $t > 0$. Es gilt

$$\begin{aligned}
P(N(Y) = k) &= \int_0^{\infty} P(N(Y) = k | Y = t) \mu e^{-\mu t} dt \\
&= \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \mu e^{-\mu t} dt \\
&= \frac{\lambda^k \mu}{k!} \underbrace{\int_0^{\infty} t^k e^{-(\lambda+\mu)t} dt}_{\left(\frac{(\lambda+\mu)^{k+1}}{k!}\right)^{-1}, \text{ da } \Gamma\text{-Dichte}} \\
&= \frac{\mu}{\lambda+\mu} \left(\frac{\lambda}{\lambda+\mu}\right)^k, \ k \in \mathbb{N}_0 \\
&= \left(1 - \frac{\mu}{\lambda+\mu}\right)^k \frac{\mu}{\lambda+\mu}
\end{aligned}$$

Also ist $N(t)$ geometrisch verteilt mit Parameter $\frac{\mu}{\lambda+\mu}$.

Kapitel 8

Grenzwertsätze

Zur Motivation zwei Problemstellungen:

a) Man betrachte einen unendlichen Münzwurf, wobei die Wahrscheinlichkeit p dafür, daß „Kopf" fällt unbekannt ist. Die Frage ist also, ob die Münze fair ist, d.h. wie groß p ist. Intuitiv wäre natürlich eine Lösung, die Münze sehr oft zu werfen und die Ergebnisse zu untersuchen. Sei dabei $h_n = \frac{\text{Anzahl der Würfe mit Kopf}}{n = \text{Anzahl aller Würfe}}$. Die Frage ist nun, ob $h_n \overset{n\to\infty}{\longrightarrow} p$.

b) In einem Wartesystem sei die Anzahl der ankommenden Kunden im Zeitintervall i (jeweils der Länge 1) $X_i \sim \text{Poi}(\lambda)$. Demnach ist die Anzahl der bis zur Zeit n angekommenen Kunden $Y_n = \sum_{i=1}^{n} X_i \sim \text{Poi}(n\lambda)$. Da für große n die numerische Berechnung recht lästig wird interessiert man sich dafür, welche Verteilung sich asymptotisch für große n ergibt. Es existieren Konstanten a_n, b_n mit

$$\frac{Y_n - a_n}{b_n} \overset{asymptotisch}{\sim} N(0,1) \qquad \text{kurz: } \frac{Y_n - a_n}{b_n} \overset{as}{\sim} N(0,1)$$

Dies gilt für beliebige Verteilung der X_i, insbesondere auch für $X_i \sim \text{Bin}(k,p)$.

Zunächst ein paar geeignete Konvergenzbegriffe:

Definition 8.1 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X ebenfalls eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$. Die Folge heißt ...

a) ... *P-fast sicher konvergent* gegen X (almost surely, almost everywhere), wenn

$$P\left(\left\{\omega \;:\; \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = P\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Bezeichnung:

$$X_n \overset{n\to\infty}{\longrightarrow} X \text{ P-f.s., oder } \lim_{n\to\infty} X_n = X \text{ P-f.s.}$$

b) ... *P-stochastisch konvergent* gegen X (stochastically convergent), wenn

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

Bezeichnung:

$$X_n \overset{n\to\infty}{\longrightarrow} X \text{ P-stoch.}, \; \lim_{n\to\infty} X_n = X \text{ P-stoch.}$$

c) ... *schwach* oder *verteilungskonvergent* (convergence in distribution) gegen X, wenn

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

in allen Stetigkeitspunkten von F, wobei $X_n \sim F_n$ und $X \sim F$. Bezeichnung:

$$X_n \overset{as}{\sim} X, \; X_n \overset{\mathcal{D}}{\longrightarrow} X$$

Lemma 8.2 Es gilt

$$P(\limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}) = 0 \Longleftrightarrow X_n \overset{n\to\infty}{\longrightarrow} X \; P - f.s.$$

Also gilt

$$X_n \to X \; P - f.s. \Longleftrightarrow P(|X_n - X| > \varepsilon \text{ für unendlich viele } n) = 0 \; \forall \varepsilon > 0$$

Beweis: von „$\Longleftarrow$“

$$\forall \varepsilon = \frac{1}{k} : \; P\left(|X_n - X| > \frac{1}{k} \text{ für } \infty \text{ viele } n\right) = 0$$

$$\text{Es folgt: } P\underbrace{\left(\bigcup_k \left\{\omega \;:\; |X_n(\omega) - X(\omega)| > \frac{1}{k} \text{ für } \infty \text{ viele } n\right\}\right)}_{=\left\{\omega \,:\, \exists k \forall n_0 \exists n \geq n_0 \,:\, |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}}$$

$$\text{Also: } P\left(\left\{\omega \;:\; \forall k \exists n_0 \forall n \geq n_0 \;:\; |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}\right) = 1$$

d.h. $X_n \to X$ P-f.s.

Satz 8.3 (Zusammenhänge zwischen Konvergenzarten)

a) $X_n \to X \; P - f.s. \overset{(i)}{\Longrightarrow} X_n \to X \; P - stoch \overset{(ii)}{\Longrightarrow} X_n \xrightarrow{\mathcal{D}} X$. Die Umkehrungen sind im Allgemeinen falsch.

b) „Schnelle“ stochastische Konvergenz impliziert fast sichere Konvergenz.

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \; \forall \varepsilon > 0 \Longrightarrow X_n \to X \; P - f.s.$$

Beweis:

a) erste Folgerung: $\forall \varepsilon > 0$ gilt:

$$\begin{aligned}
0 &\leq \lim_{n\to\infty} P(|X_n - X| > \varepsilon) \\
&\leq \lim_{n\to\infty} P\left(\bigcup_{m\geq n} \{|X_m - X| > \varepsilon\}\right) \\
&\overset{Stetigkeit}{=} P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\right) \\
&= P\left(\limsup_{n\to\infty} \{|X_n - X| > \varepsilon\}\right) \\
&= P(|X_n - X| > \varepsilon \text{ für } \infty \text{ viele } n) \\
&= 0 \qquad \text{da } X_n \to X \; P - f.s. \text{ (wg. La 8.2)}
\end{aligned}$$

zweite Folgerung: siehe „Mathar und Pfeiffer“, Lemma 2.3.2, Seite 137

b) Sei $\varepsilon > 0$. Mit dem Borel-Cantelli-Lemma (Lemma 8.2) folgt

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \overset{BCL}{\Longrightarrow} P\left(\limsup_{n\to\infty} \{|X_n - X| > \varepsilon\}\right) = 0,$$

woraus die Behauptung folgt.

Lemma 8.4 (Tschebyscheff-Ungleichung) (Chebyshev inequality) Sei X eine Zufallsvariable mit $\mathrm{Var}(X) < \infty$. Dann gilt für alle $\varepsilon > 0$

$$P(|X - E(X)| > \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Beweis: Mit der Markoff-Ungleichung folgt

$$P(|X - EX| > \varepsilon) = P\left((X - EX)^2 > \varepsilon^2\right) \leq \frac{E\left(|X - EX|^2\right)}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Satz 8.5 (Starkes Gesetz großer Zahlen, SGGZ) (Strong Law of Large Numbers, LLN) Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von paarweise unkorrelierten Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit $\text{Var}(X_n) \leq M < \infty\ \forall n \in \mathbb{N}$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{P-fast sicher.}$$

Beweis: Ohne Beschränkung der Allgemeinheit sei $EX_i = 0$ für alle i, wobei alle X_i paarweise unkorreliert seien. Setze $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ und zeige, daß für alle $\varepsilon > 0$ gilt $\sum_{n=1}^{\infty} P\left(\left|\overline{X_n}\right| > \varepsilon\right) < \infty$. Dann ist $\overline{X_n} \to 0$ P-fast sicher konvergent (Satz 8.3.b).
Betrachte zunächst die Teilfolge $\overline{X_{n^2}}$. Mit der Tschebyscheff-Ungleichung folgt

$$P\left(\left|\overline{X_{n^2}}\right| > \varepsilon\right) \leq \frac{\text{Var}\left(\overline{X_{n^2}}\right)}{\varepsilon^2} \leq \frac{\frac{1}{n^4}n^2 M}{\varepsilon^2} = \frac{M}{n^2\varepsilon^2} \quad \forall n \in \mathbb{N},\ \varepsilon > 0$$

Also gilt

$$\sum_{n=1}^{\infty} P\left(\left|\overline{X_{n^2}}\right| > \varepsilon\right) \leq \sum_{n=1}^{\infty}\frac{M}{n^2\varepsilon^2} = \frac{M}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty$$
$$\Rightarrow \quad \overline{X_{n^2}} \to 0 \text{ P-fast sicher konvergent}$$

Sei $n = n(k)$ definiert durch $n^2 \leq k < (n+1)^2,\ k \in \mathbb{N}$. Dann ist

$$\text{Var}\left(k\overline{X_k} - n^2\overline{X_{n^2}}\right) = \text{Var}\left(\sum_{i=n^2+1}^{k} X_i\right) \leq (k-n)^2 M.$$

Es folgt mit der Tschebyscheff-Ungleichung

$$P\left(\left|k\overline{X_k} - n^2\overline{X_{n^2}}\right| \geq \varepsilon n^2\right) \leq \frac{(k-n)^2 M}{\varepsilon^2 n^4},$$

also mit Satz 8.3.b

$$\frac{k}{n(k)^2}\overline{X_k} - \overline{X_{n(k)^2}} \overset{k\to\infty}{\longrightarrow} 0 \quad \text{P-f.s.}$$

Insgesamt:

$$\exists F_1 \in \mathfrak{A},\ P(F_1) = 1\ \forall\omega \in F_1 :\ \overline{X_{n^2}} \overset{n\to\infty}{\longrightarrow} 0$$
$$\exists F_2 \in \mathfrak{A},\ P(F_2) = 1\ \forall\omega \in F_2 :\ \tfrac{k}{n(k)^2}\overline{X_k} - \overline{X_{n(k)^2}} \overset{k\to\infty}{\longrightarrow} 0$$
$$\tfrac{k}{n(k)^2} \overset{k\to\infty}{\longrightarrow} 1$$

Es folgt: $\overline{X_n} \to 0 \; \forall \omega \in F_1 \cap F_2$, wobei

$$P(F_1 \cap F_2) = \underbrace{P(F_1)}_{=1} + \underbrace{P(F_2)}_{=1} - \underbrace{P(F_1 \cup F_2)}_{=1} = 1,$$

d.h. $\overline{X_n} \to 0$ P-fast sicher konvergent.

Bemerkung 8.6

a) Aus Satz 8.3.a (fast sichere Konvergenz impliziert stochastische Konvergenz) folgt

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \stackrel{n\to\infty}{\longrightarrow} 0 \text{ P-fast sicher konvergent} \\ \Longrightarrow \quad & \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \stackrel{n\to\infty}{\longrightarrow} 0 \text{ P-stochastisch konvergent} \end{aligned}$$

Die gefolgerte Aussage heißt „*Schwaches Gesetz großer Zahlen*“ *WLLN*.

b) Das starke Gesetz großer Zahlen gilt auch, wenn $\{X_n\}_{n\in\mathbb{N}}$ stid und $E(X_1)$ existiert (ohne dabei die Existenz der Varianz zu fordern). Mit der Existenz von $E(X_1)$ folgt wegen der stochastischen Unabhängigkeit und Gleichverteilung der X_i die Existenz von $E(X_i)$. Dann gilt mit $\mu = E(X_i)$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{n\to\infty}{\longrightarrow} \mu \text{ P-fast sicher konvergent}$$

Beweis: aufwendig (s. Shiryayev)

Beispiel 8.7 (Anwendung des SGGZ auf Wahrscheinlichkeiten) Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen, identisch verteilten Zufallsvariablen, $A \in \mathfrak{B}^1, P(X_n \in A) = p \; \forall n \in \mathbb{N}$. Es ist $\sum_{i=1}^{n} \mathbb{I}_A(X_i)$ die Anzahl des Auftretens $\{X_i \in A\}$ bis zum n-ten Versuch. Die $Y_i = \mathbb{I}_A(X_i)$ sind ebenfalls stochastisch unabhängig, identisch verteilt mit $E(Y_i) = P(X_i \in A) = p \; \forall i \in \mathbb{N}$. Mit dem starken Gesetz großer Zahlen folgt

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(X_i)}_{\text{s. Fußnote } 1} \stackrel{n\to\infty}{\longrightarrow} p \text{ P-fast sicher konvergent}$$

Beispiel 8.8 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen, identisch verteilten Zufallsvariablen mit $\mu = E(X_i), \; \sigma^2 = \mathrm{Var}(X_i)$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X_n} \stackrel{n\to\infty}{\longrightarrow} \mu \text{ P-fast sicher konvergent}$$

d.h. $\overline{X_n}$ ist ein „stark konsistenter Schätzer“ für μ. Desweiteren gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \overset{n\to\infty}{\longrightarrow} E\left(X_i^2\right) \text{ P-fast sicher konvergent}$$

wegen des starken Gesetzes großer Zahlen, mit X_i ersetzt durch X_i^2. Außerdem gilt

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X_n}\right)^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X_n}\right)^2 \\ \longrightarrow \quad & E\left(X_i^2\right)-\mu^2 = \operatorname{Var}(X_1) = \sigma^2 \end{aligned}$$

ebenfalls wegen des starken Gesetzes großer Zahlen, d.h. $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ ist ein „stark konsistenter Schätzer“ für σ^2.

Satz 8.9 (Zentraler Grenzwertsatz, ZGWS (CLT)) Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge stochastisch unabhängiger, identisch verteilter Zufallsvariablen mit $\mu = E(X_n)$ und $\sigma^2 = \operatorname{Var}(X_n) > 0$ existent. Dann gilt

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \overset{as}{\sim} \mathrm{N}(0,1) \text{ für } n\to\infty,$$

d.h.

$$P\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)\le z\right) \longrightarrow \phi(z) = \frac{1}{2\pi}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx \ \forall z\in\mathbb{R}.$$

Beweis: z.B. „Casella und Berger“, Seite 216 ff.
Für $S_n = \sum_{i=1}^{n} X_i$ gilt $E(S_n) = n\mu,\ \operatorname{Var}(S_n) = n\sigma^2$. Dann ist

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \overset{as}{\sim} \mathrm{N}(0,1) \text{ für } n\to\infty.$$

Die standardisierte Summe konvergiert nach Verteilung gegen die N(0,1)-Verteilung.

Beispiel 8.10 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge stochastisch unabhängiger, identisch binomialverteilter Zufallsvariablen mit Parametern $1, p$ ($\mathrm{Bin}(1,p)$). Es gilt $E(X_n) = p$ und $\operatorname{Var}(X_n) = p(1-p)$. Mit dem zentralen Grenzwertsatz folgt

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \overset{as}{\sim} \mathrm{N}(0,1) \text{ für } n\to\infty.$$

Also ist

$$P\left(\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le \frac{z-np}{\sqrt{np(1-p)}}\right) \approx \phi\left(\frac{z-np}{\sqrt{np(1-p)}}\right),$$

wobei

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx.$$

Diese Approximation ist brauchbar für $n \geq 30$.

Kapitel 9

Schätzfunktionen und Konfidenzintervalle

Zur Motivation der folgenden Begriffe zunächst drei Beispiele und Problemstellungen

a) Man werfe eine Münze in unabhängigen Versuchen. Die Wahrscheinlichkeit, daß „Kopf" auftritt sei p und nicht bekannt. Daher interessiert man sich für gute Schätzungen von p. Ein mathematisches Modell wäre z.B., daß man Zufallsvariablen $X_1, \ldots, X_n$ betrachtet, die stochastisch unabhängig, identisch binomialverteilt mit Parameter $1, p$ sind ($\mathrm{Bin}(1,p)$). Dann ist $\hat{p}$ mit

$$\hat{p} = \hat{p}(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow p \quad \text{P-fast sicher konvergent, (bekannt SGGZ)}$$

ein vernünftiger Schätzer für p. Dabei ist $\hat{p}(X_1, \ldots, X_n)$ wieder eine Zufallsvariable. Setzt man *Realisationen* $x_1, \ldots, x_n$ ein, so erhält man den Schätzwert $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

b) $X_1, \ldots, X_n$ seien stochastisch unabhängig, identisch verteilte Zufallsvariablen, die exponentialverteilt sind mit Parameter λ (Die Familie der Exponentialverteilungen ist eine parametrische Familie). Dann ist

$$\hat{\vartheta} = \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \frac{1}{\lambda} = E(X_1) \quad \text{P-fast sicher, (SGGZ)}$$

ein Schätzer für $g(\lambda) = 1/\lambda = E(X_1)$.

c) In a) und b) wird der Erwartungswert durch das arithmetische Mittel geschätzt. Es gibt aber auch schwierigere Situationen, wie z.B. die Frage nach der Anzahl M der regelmäßigen Besucher einer Web-Seite. Um M zu schätzen, geht man intuitiv wie folgt vor. Zunächst merke man sich die Adressen von n Besuchern ("markieren"). Zu einem späteren Zeitpunkt merke man sich m Adressen und bestimme den Anteil x bereits markierter Besucher. Unter diesen Voraussetzungen sollte $n/m \approx x/m$. Somit ergibt sich als vernünftiger Schätzer für M:

$$M = \left\lfloor \frac{nm}{x} \right\rfloor .$$

Offen ist nun jedoch noch, ob dieses intuitive Vorgehen exakt begründet werden kann.

9.1 Methoden zur Bestimmung von Schätzern

Definition 9.1 Seien $X_1, \ldots, X_n$ Zufallsvariablen mit gemeinsamer Verteilung $P_\vartheta^{(X_1,\ldots,X_n)}$, $\vartheta \in \Theta$ (Parameterraum) und sei $g(\vartheta) : \Theta \to \mathcal{Y}$ eine Funktion. Dann heißt jede Abbildung $h(X_1, \ldots, X_n)$ mit Wertebereich $\mathcal{Y}$ *statistische Schätzfunktion* oder *(Punkt-) Schätzer* (point estimator) von $g(\vartheta)$.

Beispiel 9.2 Seien $X_1, \ldots, X_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen, die normalverteilt sind mit Parametern μ, σ^2. Der Parameterraum ist also $\Theta = \mathbb{R} \times \mathbb{R}^+$ und $\vartheta = (\mu, \sigma^2) \in \Theta$. Zunächst sei die Funktion

$$g_1(\vartheta) = g_1\left((\mu, \sigma^2)\right) = \mu$$

zu schätzen (man interessiert sich also für den ersten Parameter, daher ist g_1 gerade die Projektion auf diesen). Wie wir bereits wissen, ist

$$h_1(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

ein guter Schätzer dafür. Interessiert man sich für den zweiten Parameter, so suche man einen Schätzer für die Funktion

$$g_2(\vartheta) = g_2\left((\mu, \sigma^2)\right) = \sigma^2.$$

Auch hierfür kennen wir bereits einen guten Schätzer:

$$h_2(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 .$$

Seien zum anderen $X_1, \ldots, X_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen, die exponentialverteilt sind mit Parameter λ. Der Parameterraum ist also $\Theta = \mathbb{R}^+$ und $\vartheta = \lambda$. Definiere die zu schätzende Funktion g_1 nun so:

$$g_1(\vartheta) = g_1(\lambda) = \frac{1}{\lambda} = E(X_1),$$

denn hierfür kennen wir einen guten Schätzer

$$h_1(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(vergleiche Beispiel 6.3.b). Die Frage ist nun, ob es vernünftig ist, für die Identität $g_2(\lambda) = \lambda$, direkt

$$h_2(X_1, \ldots, X_n) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^{-1}$$

als Schätzer zu nehmen.

Im folgenden werden Methoden behandelt, um geeignete Schätzer zu finden.

Definition 9.3 (Maximum-Likelihood-Schätzer) Sei $f(X_1, \ldots, X_n|\vartheta)$, $\vartheta \in \Theta$ eine Zähldichte oder Dichte. Eine Funktion $L(\vartheta|x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\vartheta)$ (als Funktion von ϑ bei gegebenen $x_1, \ldots, x_n$) heißt *Likelihood-Funktion*. $\hat{\vartheta}(x_1, \ldots, x_n)$ heißt *Maximum-Likelihood-Schätzer* (*MLS*) (engl. *maximum-likelihood-estimator* (*MLE*)), falls

$$L\left(\hat{\vartheta}(x_1, \ldots, x_n) \middle| x_1, \ldots, x_n \right) = \sup_{\vartheta \in \Theta} (\vartheta | x_1, \ldots, x_n) \quad \forall x_1, \ldots, x_n.$$

$g\left(\hat{\vartheta}\right)$ heißt Maximum-Likelihood-Schätzer für $g(\vartheta)$.

Das dahinterstehende Konzept ist, daß $\hat{\vartheta}$ bei gegebenen Beobachtungen $x_1, \ldots, x_n$ der „passendste“ Parameter ist, der diese Beobachtungen am wahrscheinlichsten macht. $\hat{\vartheta}$ zu bestimmen ist eine Maximierungsaufgabe, die oft durch Differenzieren (Nullstellen der ersten Ableitung) gelöst werden kann. Oft ist es günstiger, statt $L(\vartheta|x)$ die *logarithmische Likelihood-Funktion* $\log L(\vartheta|x)$ zu betrachten. Das Maximum von $L(\vartheta|x)$ und $\log L(\vartheta|x)$ liegt an der gleichen Stelle.

Beispiel 9.4

a) Seien $X_1, \ldots, X_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen, die normalverteilt mit Parametern μ, σ^2 sind. Gesucht ist ein Schätzer für $\vartheta = (\lambda, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$, d.h. die Familie, aus der die Verteilung stammt, ist bekannt, der Parameter der Verteilung aber nicht. Nach Definition 9.3 gilt

$$
\begin{aligned}
& f(x_1, \ldots, x_n|\vartheta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
= \quad & \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} = L(\vartheta|x_1, \ldots, x_n) \\
\Rightarrow \quad & \log L(\vartheta|x_1, \ldots, x_n) = -\frac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu)^2.
\end{aligned}
$$

Setze nun $\tau = \sigma^2$ und partielle Ableitungen $= 0$. Anschließend wird nach μ und τ aufgelöst.

- $$\frac{\partial \log L}{\partial \mu} = \frac{2}{2\sigma^2} \sum_{i=1}^{n}(x_i - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$$

- $$
\begin{aligned}
\frac{\partial \log L}{\partial \tau} &= -\frac{n}{2}\frac{2\pi}{2\pi\tau} + \frac{1}{2\tau^2} \sum_{i=1}^{n}(x_i - \mu)^2 = 0 \\
\Rightarrow \quad \hat{\tau} &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2
\end{aligned}
$$

Insgesamt ergibt sich

$$\hat{\vartheta} = \left(\hat{\mu}, \hat{\sigma^2}\right) = \left(\overline{x}, \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2\right)$$

als Maximum-Likelihood-Schätzer für $\left(\mu, \sigma^2\right)$, wobei noch zu prüfen ist, ob das Maximum bei $\left(\hat{\mu}, \hat{\sigma^2}\right)$ wirklich angenommen wird.

b) Seien $X_1, \dots, X_n$ stochastisch unabhängige, identisch verteilte Zufallsvariablen, die exponentialverteilt mit Parameter $\lambda > 0$ sind. Gesucht ist wiederum ein Schätzer für $\vartheta = \lambda$. Es gilt

$$f(x_1, \dots, x_n|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = L(\lambda|x_1, \dots, x_n), \quad x_i \geq 0$$

$$\Rightarrow \log L(\lambda|x_1, \dots, x_n) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

Nun ist wieder die erste Ableitung gleich 0 zu setzen und nach λ aufzulösen.

$$\frac{\partial \log L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Leftrightarrow \frac{n}{\lambda} = \sum_{i=1}^{n} x_i \Leftrightarrow \lambda = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{x}}$$

Also ist $\hat{\lambda} = 1/\overline{x}$ ein Maximum-Likelihood-Schätzer für $\vartheta = \lambda$, wobei noch zu zeigen ist, daß $\overline{\lambda}$ ein Maximum von $L(\lambda|x_1, \dots, x_n)$ ist.

c) Seien $X_1, \dots, X_n$ stochastisch unabhängig und binomialverteilt mit Parametern $1, p$. Gesucht ist ein Schätzer für den Parameter $p \in [0, 1]$.

$$\begin{aligned} f(x_1, \dots, x_n|p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n-\sum_{i=1}^n x_i} \\ &= L(p|x_1, \dots, x_n), \quad x_i \in \{0, 1\} \end{aligned}$$

Der logarithmische Likelihood-Schätzer ist dann

$$\log L(p|x_1, \dots, x_n) = \left(\sum_{i=1}^{n} x_i\right) \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$$

Wiederum wird das Maximum über $p \in [0, 1]$ durch die Nullstelle der ersten Ableitung bestimmt.

$$\begin{aligned} \frac{\partial \log L}{\partial p} &= \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i\right) = 0 \\ \Leftrightarrow \quad & \frac{1}{p}\overline{x} - \frac{1}{1-p}(1-\overline{x}) = 0 \Leftrightarrow p = \overline{x} \end{aligned}$$

Also ist $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \overline{x}$ ein Maximum-Likelihood-Schätzer.

9.1.1 Bayes-Methode

Eine weitere Methode zur Konstruktion von Schätzern ist die *Bayes-Methode*

- Modelliere Vorkenntnisse über den Parameter ϑ durch eine Wahrscheinlichkeitsverteilung über dem Parameterraum Θ, die sogenannte *a-priori Verteilung*, beschrieben durch die (Zähl-)Dichte $\pi(\vartheta)$.
- $f(x|\vartheta)$, $x = x_1, \ldots, x_n$ sei die Zähldichte oder Dichte der Verteilung von $X_1, \ldots, X_n$ bei Vorliegen von ϑ, aufgefaßt als bedingte Verteilung von $(X_1, \ldots, X_n)$ bei gegebenem ϑ.
- Gegeben seien die Beobachtungen $x = (x_1, \ldots, x_n)$. Die zu $f(\vartheta|x)$ gehörige Verteilung heißt *a-posteriori Verteilung* von ϑ. $f(\vartheta|x)$ reflektiert den Kenntnisstand über ϑ *nach* Beobachten von $x_1, \ldots, x_n$.

Definition 9.5 Die Zufallsvariable $\hat{\vartheta}(x)$ besitze die (Zähl-) Dichte

$$f(\vartheta|x) = \frac{f(x,\vartheta)}{f(x)} = \frac{f(x|\vartheta)\pi(\vartheta)}{\int f(x|\vartheta)\pi(\vartheta)d\vartheta},$$

wobei sich die normierende Konstante im Nenner durch

$$f(x) = \int f(x|\vartheta)\pi(\vartheta)d\vartheta$$

ergibt (vergleiche Kapitel 7). Dann heißt $E(\vartheta(x))$ *Bayes-Schätzer* von ϑ. Bemerkung: Bei Dichten ist $\int \vartheta f(\vartheta|x)d\vartheta$ der Bayes-Schätzer.

Beispiel 9.6 Sei die Zufallsvariable X binomialverteilt mit Parametern n, p, wobei das n fest sei ($X \sim \text{Bin}(n,p)$). Die a-priori Verteilung für p sei $\text{Beta}(\alpha,\beta)$, $\alpha, \beta > 0$ mit Dichte

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0 \le p \le 1$$

Es gilt: Falls X beta-verteilt mit Parametern α, β ist ($X \sim \text{Beta}(\alpha,\beta)$), dann gilt für den Erwartungswert von X: $E(X) = \frac{\alpha}{\alpha+\beta}$. Sei $x \in \{0, 1, \ldots, n\}$. Dann ist

$$\begin{aligned}
f(p|x) &= \frac{f(x|p)\pi(p)}{\int f(x|p)\pi(p)dp} = \frac{\binom{n}{x} p^x (1-p)^{n-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}}{\int f(x|p)\pi(p)dp} \\
&= \frac{\binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1}}{\underbrace{\int f(x|p)\pi(p)dp}_{=1,\ \text{Int. über Dichte}}} \\
&= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)} p^{x+\alpha-1}(1-p)^{n-x+\beta-1}
\end{aligned}$$

die Dichte einer $\text{Beta}(x+\alpha, n-x+\beta)$-Verteilung. Also ist $\hat{p}(x) = \frac{x+\alpha}{n+\alpha+\beta}$ der Bayes-Schätzer zur a-priori Verteilung π. Speziell für $\alpha = \beta = 1$ (dann ist die Beta-Verteilung nämlich gleich der Rechteckverteilung mit Parametern 0,1 (Beta$(\alpha,\beta) = R(0,1)$)) gilt

$$\hat{p}(x) = \frac{x+1}{n+2}$$

Der Maximum-Likelihood-Schätzer für p lautet in dem Fall: $\hat{p}_{ML}(x) = \frac{x}{n}$.

9.2 Gütekriterien für Schätzer

Definition 9.7 $H = h(X_1,\ldots,X_n)$ sei ein Schätzer für $g(\vartheta) \in \mathbb{R}$. Dann heißt $E_\vartheta(H-g(\vartheta))^2 = E_\vartheta(h(X_1,\ldots,X_n)-g(\vartheta))^2$ der *mittlere quadratische Fehler* (engl.: *MSE= mean squared error*). Der MSE mißt die mittlere quadratische Abweichung vom zu schätzenden Wert $g(\vartheta)$ und ist eine Funktion von ϑ. Es gilt:

$$\begin{aligned}
& E_\vartheta(H-g(\vartheta))^2 = E_\vartheta(H - E_\vartheta H + E_\vartheta H - g(\vartheta))^2 \\
= \;& E_\vartheta(H-E_\vartheta H)^2 + 2(E_\vartheta H - g(\vartheta))\underbrace{E_\vartheta(H-E_\vartheta H)}_{=0} + (E_\vartheta H - g(\vartheta))^2 \\
= \;& E_\vartheta(H-E_\vartheta H)^2 + (E_\vartheta H - g(\vartheta))^2 = \mathrm{Var}_\vartheta(H) + (\mathrm{Bias}_\vartheta(H))^2
\end{aligned}$$

Hierbei ist $\mathrm{Var}_\vartheta(H)$ die *Präzision* oder die *Variablilität* des Schätzers und *Bias*$_\vartheta(H)$ die *Schiefe* oder *Genauigkeit* des Schätzers.

Wichtig ist der Fall, daß $\mathrm{Bias}_\vartheta = 0$ ist.

Definition 9.8 Ein Schätzer H heißt *erwartungstreu* (*unbiased*) für $g(\vartheta)$, wenn $\mathrm{Bias}_\vartheta(H) = 0$ für alle $\vartheta \in \Theta$, – falls also $E_\vartheta(H) = g(\vartheta)$ für alle $\vartheta \in \Theta$. Offensichtlich gilt für erwartungstreue Schätzer

$$\mathrm{MSE}(H) = E_\vartheta(H-g(\vartheta))^2 = \mathrm{Var}_\vartheta(H).$$

Beispiel 9.9

a) Prüfe, ob der Schätzer $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ erwartungstreu ist. Dazu folgende Vorüberlegungen:

$$
\begin{aligned}
Y \sim \mathrm{N}(0,1) &\Rightarrow E\left(Y^2\right) = 1 \quad \text{(Bsp. 6.13)} \\
Y \sim \mathrm{N}(0,1) &\Rightarrow Z := \sigma Y + \mu \sim \mathrm{N}(\mu, \sigma^2) \\
E\left(Z^2\right) &= E(\sigma Y + \mu)^2 = \sigma^2 E\left(Y^2\right) + 2\sigma\mu \underbrace{EY}_{=0} + \mu^2 \\
&= \sigma^2 + \mu^2
\end{aligned}
$$

Demnach ist

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \hat{X}^2$$

mit einem Erwartungswert von

$$
\begin{aligned}
E_\vartheta\left(\hat{\sigma}^2\right) &= \frac{1}{n}\sum_{i=1}^n \underbrace{E\left(X_i^2\right)}_{\sigma^2+\mu^2} - \underbrace{E\left(\overline{X}^2\right)}_{\frac{\sigma^2}{n}+\mu^2,\ da\ \overline{X}\sim\mathrm{N}\left(\mu,\frac{\sigma^2}{n}\right)} \\
&= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

Also ist $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ nicht erwartungstreu für σ^2, aber *asymptotisch erwartungstreu*, d.h.

$$\sigma = \lim_{n\to\infty} E\left(\hat{\sigma}^2\right).$$

Zudem ist

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$$

erwartungstreu für σ^2, denn

$$E\left(S^2\right) = E\left(\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2\right) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Beispiel 9.10 (Gütevergleich von zwei Schätzern) Sei X binomialverteilt mit Parametern p, n, wobei n fest sei. Bekannt sind zwei Schätzer für p,
– zum einen der Maximum-Likelihood-Schätzer

$$\hat{p}_{ML} = \frac{x}{n}$$

und zum anderen der zur a-priori Verteilung Beta(α, β)

$$\hat{p}_B = \frac{x+\alpha}{n+\alpha+\beta} \quad \text{(s. Bsp. 9.6).}$$

Es stellt sich nun die Frage, welcher von beiden besser ist. Man vergleiche beide mit dem mittleren quadratischen Fehler (MSE).

$$\begin{aligned}
E_p\left(\hat{p}_{ML}-p\right)^2 &= \operatorname{Var}\left(\hat{p}_{ML}\right) = \operatorname{Var}\left(\frac{1}{n}X\right) \\
&= \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n} \\
E_p\left(\hat{p}_B-p\right)^2 &= \operatorname{Var}\left(\hat{p}_B\right) + \left(\operatorname{Bias}_p\left(\hat{p}_B\right)\right)^2 \\
&= \operatorname{Var}\left(\frac{X+\alpha}{n+\alpha+\beta}\right) + \left(E_p\left(\frac{X+\alpha}{n+\alpha+\beta}\right)-p\right)^2 \\
&= \frac{np(1-p)}{n+\alpha+\beta)^2} + \left(\frac{np+\alpha}{n+\alpha+\beta}-p\right)^2
\end{aligned}$$

Wähle nun α, β so, daß der mittlere quadratische Fehler von $\hat{p}_B$ konstant ist, d.h. kein Wert von p wird bei der Schätzung bevorzugt, – in dem Fall: $\alpha = \beta = 1/2 \cdot \sqrt{n}$. Dann gilt

$$\hat{p}_B = \frac{X+\frac{1}{2}\sqrt{n}}{n+\sqrt{n}}, \quad E_p\left(\hat{p}_B-p\right)^2 = \frac{n}{4(n+\sqrt{n})^2}.$$

Für kleine Stichproben ist also der Bayes-Schätzer besser, für kleine der Maximum-Likelihood-Schätzer.

9.3 Konfidenzintervalle

Im vorherigen Abschnitt wurden Punktschätzer behandelt, wobei immer ein $\hat{\vartheta}$ bestimmt wurde, das möglichst nahe am tatsächlichen Parameter ϑ liegt. Im folgenden sollen Schranken angegeben werden, so daß der Parameter ϑ mti vorgegebender Wahrscheinlichkeit innerhalb dieser Schranken liegt. Die Idee dabei ist: Seien $X_1, \ldots, X_n$ stochastisch unabhängig und alle binomialverteilt mit Parametern $1, p$. Aus dem schwachen Gesetz großer Zahlen folgt

$$\begin{aligned}
& P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \stackrel{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0 \\
\Longleftrightarrow \quad & P\left(\overline{X}-\varepsilon \leq p \leq \overline{X}+\varepsilon\right) \stackrel{n\to\infty}{\longrightarrow} 1
\end{aligned}$$

Für große n wird der Parameter p durch das zufällige Intervall $\left[\overline{X}-\varepsilon, \overline{X}+\varepsilon\right]$ mit hoher Wahrscheinlichkeit überdeckt. Man nutzt die Kenntnisse über die Verteilung von $\overline{X}$ um $\varepsilon, P(\ldots)$ zu quantifizieren.

Definition 9.11 Seien $X_1, \ldots, X_n$ Zufallsvariablen mit gemeinsamer Verteilung $P^{(X_1,\ldots,X_n)}$. Ein Intervall der Form $[L(X_1,\ldots,X_n), U(X_1,\ldots,X_n)]$ heißt *Konfidenzintervall* zum Niveau $1-\alpha$ für $g(\vartheta)$, falls

$$P_\vartheta\left(g(\vartheta) \in [L(X_1,\ldots,X_n), U(X_1,\ldots,X_n)]\right) \geq 1-\alpha \quad \forall \vartheta \in \Theta.$$

Beispiel 9.12 Seien $X_1, \ldots, X_n$ stochastisch unabhängige, exponentialverteilte Zufallsvariablen mit Parameter μ, σ^2 ($X_1, \ldots, X_n$ stid $\sim \mathrm{N}\left(\mu, \sigma^2\right)$), wobei das σ^2 fest sei. Dann ist der Mittelwert $\overline{X}$ normalverteilt mit Parametern $\mu, \frac{\sigma^2}{n}$ und

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$$

Bestimme nun u so, daß

$$P\left(\left|\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq u\right) = 1-\alpha$$

Wähle $u = u_{1-\frac{\alpha}{2}}$ als $(1-\alpha/2)$-Fraktil der N(0,1)-Verteilung. Auflösen nach μ ergibt

$$\begin{aligned} P\left(\left|\overline{X}-\mu\right| \leq \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right) &= P\left(\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right) \\ &= 1-\alpha \end{aligned}$$

Also ist

$$\left[\overline{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2},\ \overline{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right]$$

ein $(1-\alpha)$-Konfidenzintervall für μ. Die Länge der Konfidenzintervalls fällt mit wachsendem n und wächst mit wachsendem Niveau $1-\alpha$. Es gibt auch einseitige Konfidenzintervalle, festgelegt durch

$$P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq u\right) = 1-\alpha.$$

Wähle $u = u_{1-\alpha}$, denn dann ist

$$P\left(\overline{X}-\mu \leq \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha/2}\right) = P\left(\mu \geq overlineX - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\alpha}\right) = 1-\alpha$$

Also ist $\left(\overline{X} - u_{1-\alpha}\frac{\sigma}{\sqrt{n}},\ \infty\right)$ ein einseitiges $(1-\alpha)$-Konfidenzintervall für μ.

α	0,1	0,05	0,025	0,01	0,005	0,0025
$1-\alpha$	0,9	0,95	0,975	0,99	0,995	0,9975
$u_{1-\alpha}$	1,282	1,645	1,960	2,326	2,576	2,807

Beispiel 9.13 (Approximatives Konfidenzintervall für Wahrscheinlichkeiten)
Seien $X_1, \ldots, X_n$ binomialverteilt mit Parametern $1, p$ ($X_1, \ldots, X_n \sim \mathrm{Bin}(1, p)$).
Dann ist

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{as}{\sim} \mathrm{N}(0,1) \quad \text{(Bsp. 8,10)}$$

Es gilt

$$P\left(\frac{\left|\overline{X} - p\right|}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\alpha/2}\right) \approx 1 - \alpha,$$

wobei $u_{1-\alpha/2}$ das Fraktil der N(0,1)-Verteilung ist.

$$\left|\overline{X} - p\right| = u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \Longleftrightarrow \left(\overline{X} - p\right)^2 = u_{1-\alpha/2}\frac{p(1-p)}{n}$$

ist ein quadratische Gleichung in p mit Lösungen $p_L\left(\overline{X}\right) \leq p_U\left(\overline{X}\right)$. $\left[p_L\left(\overline{X}\right),\ p_U\left(\overline{X}\right)\right]$ ist ein approximatives $(1-\alpha)$-Konfidenzintervall für p. (explizite Lösung der quadratischen Gleichung in „Casella und Berger“, p. 445)

Index