Einführung in die Stochastik für Informatiker
Sommersemester 2002
Prof. Mathar

Ursprüngliche Version (SS 2000) geTEXt von

René Wörzberger
rene@woerzberger.de

Bilder

Thorsten Uthke

Review

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

Update Sommersemester 2001

Martin Habbecke
M.Habbecke@gmx.de

Ergänzungen und Update Sommersemester 2002

Prof. Dr. R. Mathar, Daniel Catrein

8. Oktober 2004

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 (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), -netze (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
Kolmogorov	(1973)	Axiomatische Entwicklung

Mit Hilfe der nun folgenden Beispiele werden einige typische Problemstellungen und Ergebnisse der Stochastik anschaulich eingeführt.

Beispiel 1.1 (Münzwurf). Eine Münze werde wiederholt, in der Vorstellung ohne jemals abzubrechen, geworfen. Das Ergebnis Kopf wird durch $+1$, das Ergebnis Zahl durch -1 kodiert. "Kopf" trete mit Wahrscheinlichkeit p, "Zahl" mit Wahrscheinlichkeit $1-p$ auf, wobei $0 \leq p \leq 1$. Die Münze heißt "fair" wenn $p = 1/2$, wenn also Kopf und Zahl mit gleicher Wahrscheinlichkeit auftreten. Das Ergebnis der Münzwürfe wird als Auszahlung eines Spiels zwischen zwei Spielern interpretiert. Bei Auftreten von "Kopf" erhält der erste Spieler einen Euro vom zweiten, bei "Zahl" ist es gerade umgekehrt.
Jedes Ergebnis einer Münzwurfserie kann durch eine unendliche Folge aus den Zahlen -1 und $+1$ dargestellt werden. Die Menge aller möglichen Ergebnisse ist also

$$\Omega = \big\{(x_1, x_2, x_3, \dots) \mid x_i \in \{-1, +1\}\big\}.$$

Interessant ist in diesem Spiel der Kontostand des ersten Spielers nach n Würfen. Bei Anfangskapital 0 besitzt er die Darstellung

$$s_n = \sum_{i=1}^{n} x_i.$$

Es stellt sich die Frage, ob es für den Kontostand eine "Gesetzmäßigkeit des Zufalls", ein "Verteilungsgesetz" gibt? Anschaulich wird diese Frage durch einen Versuchsaufbau von Galton, das sogenannte "Galton-Brett" beantwortet.
Man stelle sich hierzu Reihen von untereinander versetzt eingeschlagenen Näglen auf einem stark geneigten Brett vor. Der Versatz ist jeweils so, daß eine symmetrisch auf den obersten Nagel gesetzte Kugel beim Herunterfallen den nachfolgenden Nagel jeweils wieder genau in der Mitte trifft. Abbildung 1.1 stellt den Aufbau schematisch dar. Fällt die Kugel nach links, verliert der Spieler 1,– EUR, fällt sie nach rechts, gewinnt er 1,– EUR.
Gemäß Abbildung 1.1 werden unter der n-ten Nagelreihe Schächte angebracht, die die Kugel auffangen, wenn sie vom letzten Nagel herunterfällt. Die in dem jeweiligen Schacht notierte Zahl gibt dann den Kontostand nach n Spielen an.

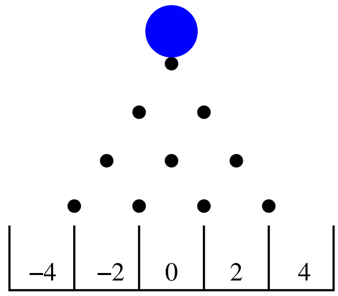


Abbildung 1.1: Galton-Brett mit vier Nagelreihen.

Die Stochastik beschäftigt sich hierbei mit den folgenden interessanten Fragen.

- Mit welcher Wahrscheinlichkeit landet eine Kugel nach n Nagelreihen in Schacht Nummer k? Diese Frage wird später mit Hilfe der Binomialverteilung beantwortet.
- Welches Bild des Füllstands der Schächte ergibt sich, wenn sehr viele Kugeln durch eine große Zahl von Nagelreihen fallen? Einen Eindruck von dem zu erwartenden Ergebnis gibt Abbildung 1.2, das aus einer Computersimulation stammt. Die analytische Antwort hierzu wird der Zentraler Grenzwertsatz geben.
- Wird das Galton Brett um 90^0 gegen den Uhrzeigersinn gedreht, lassen sich die Wege der Kugeln als Pfade mit Sprüngen nach oben und unten interpretieren. Deren asymptotisches Wachstum mit $n \to \infty$ wir durch das *Gesetz vom iterierten Logarithmus* beschrieben. Es besagt, daß

$$\begin{aligned} \limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \ln\ln n}} &= +1 \quad \text{und} \\ \liminf_{n\to\infty} \frac{S_n}{\sqrt{2n \ln\ln n}} &= -1 \end{aligned}$$

mit Wahrscheinlichkeit 1. Die graphische Darstellung einiger zufälliger Pfade nebst den begrenzenden Kurven $\pm\sqrt{2n \ln\ln n}$ findet sich in Abbildung 1.3. Die zugehörige Begriffswelt zum Verständnis des Satzes wird im Laufe der Vorlesung entwickelt, den Beweis selbst werden wir mit den Mitteln der Vorlesung nicht führen können.

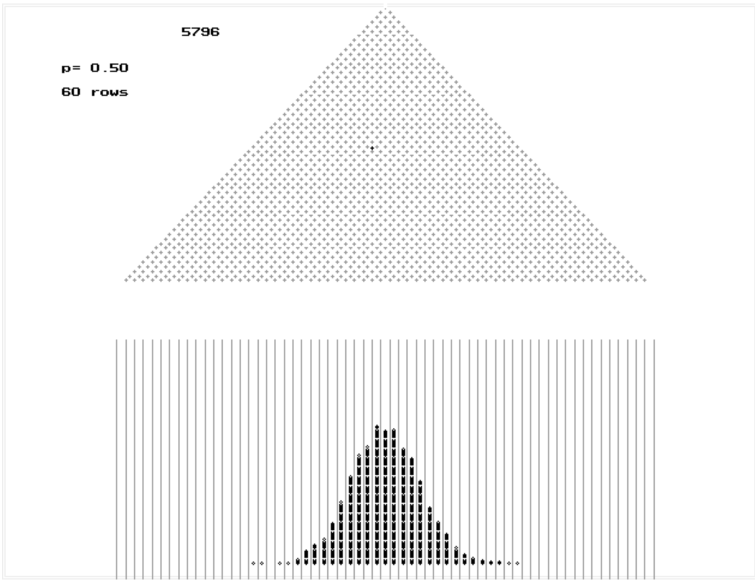
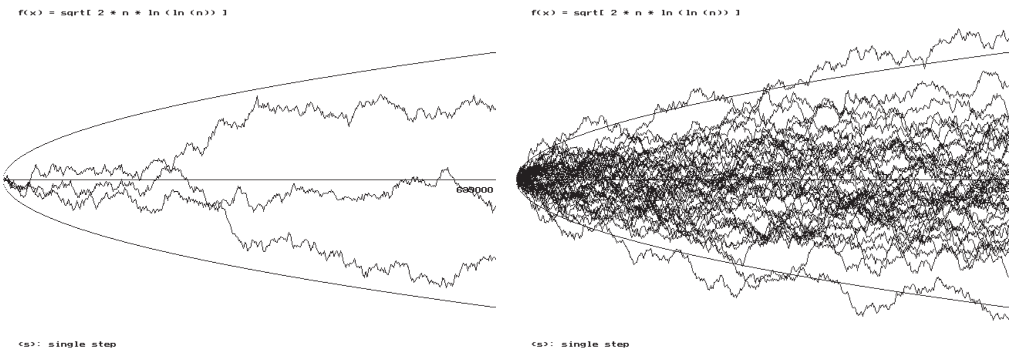


Abbildung 1.2: Kugelverteilung bei 60 Nagelreihen und 5796 Kugeln.



(a) 3 verschiedene Pfade

(b) 70 verschiedene Pfade

Abbildung 1.3: Einige Pfade des Kontostands jeweils bis zum 639.000sten Wurf.

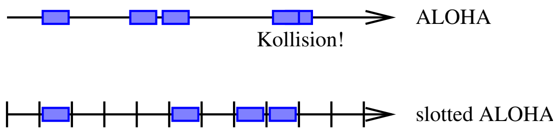


Abbildung 1.4: Aloha und slotted ALOHA.

Beispiel 1.2 (ALOHA). ALOHA ist ein sogenanntes *contention Protokoll*. Mehrere Stationen teilen sich ein gemeinsames Medium zur Übertragung von Datenpaketen fester Länge, etwa Kabel, Glasfaser oder einen Funkkanal. Wenn zwei Pakete bei der Übertragung überlappen, zerstören sie sich gegenseitig durch Interferenz.
Bei reinem ALOHA wählt jede Station unabhängig von den anderen zufällig Zeitpunkte, zu denen sie ein Paket überträgt. Bei slotted ALOHA wird der Kanal in Zeitschlitze eingeteilt, in denen genau ein Paket übertragen werden kann, s. Abbildung 1.4.
Zur Analyse dieses Protokolls müssen die folgenden Fragen beantwortet werden.

- Welches stochastische Modell beschreibt das über der Zeit zufällige Auftreten von Paketen und welche stochastischen Eigenschaften hat die Überlagerung solcher Ströme? Ein geeignetes Modell hierfür ist der Poisson-Prozess.

- Wie groß ist die Wahrscheinlichkeit für die Zerstörung eines Pakets?

- Welcher Durchsatz ist zu erzielen? Mit Durchsatz ist dabei die erwartete Anzahl unzerstörter Pakete pro Zeiteinheit gemeint.

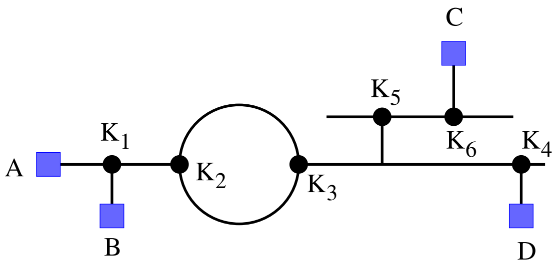


Abbildung 1.5: Verzögerungszeiten von Knoten in einem Netzwerk.

Beispiel 1.3 (Laufzeiten in Netzwerken). Man betrachte ein Netzwerk mit Terminals A bis D und Knoten $K_1, \dots, K_6$, wie in Abbildung 1.5 dargestellt. Im praktischen Betrieb sollen die Verzögerungzeiten in den Knoten $K_1, \dots, K_6$ bestimmt werden.
Hierzu wird von bestimmten Terminals aus zu einigen anderen per "ping" ein Datenpaket geschickt und seine Laufzeit bis zur Rückkehr gemessen. v_i bezeichne die Verzögerung in Knoten i und $L_{X\leftrightarrow Y}$ die Laufzeit von Terminal X nach Y und zurück. Folgender Zusammenhang besteht zwischen den Lauf- und Verzögerungszeiten. Zum Beispiel gilt zwischen A und D bzw. zwischen B und C

$$\begin{aligned} L_{A\leftrightarrow D} &= 2(v_1 + v_2 + v_3 + v_4) + \varepsilon_1, \\ L_{B\leftrightarrow C} &= 2(v_1 + v_2 + v_3 + v_5 + v_6) + \varepsilon_2. \end{aligned}$$

Die Laufzeit ist eine zufällige Größe, da sie von einem zufälligen Meßfehler und der gerade zufällig vorherrschenden Last im Netz abhängt, was durch die Zufallsvariablen ε_i ausgedrückt wird.
Es stellt sich die Frage nach einem vernünftigen Modell für den Meßfehler und die Lastverteilung. In vielen Fällen wird die Normalverteilung geeignet sein. Hat man ein geeignetes Modell gefunden, bleibt das Problem, aus den zufälligen Meßergebnissen die deterministischen Verzögerungen v_i zu bestimmen. Dies ist eine typische statistische Fragestellung, die mit Hilfe von statistischen Schätzfunktionen oder mit Hilfe von Konfidenzintervallen beantwortet werden kann.

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. $P(\Omega) = 1,\ P(\emptyset) = 0$ $\qquad (*)$
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. $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{P}(\Omega),\ A_i \cap A_j = \emptyset \quad \forall i,j \ \ i \neq j$ $\qquad (*)$

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

$$P(A) = \frac{|A|}{|\Omega|},\ A \in \mathfrak{P}(\Omega) \tag{2.1}$$

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: $\omega \equiv$ der Position des gesuchten Elements y und $\omega = 0$, 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.

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

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 zusammengesetzten 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 $A_{k,n}$ aller Möglichkeiten, 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

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

mit $\ln x \leq x - 1, x \geq 0 \quad \Leftrightarrow \quad ln(1-x) \leq -x, x \leq 1$ folgt

$$P(A_{k,n}) \leq \exp\left(-\sum_{i=0}^{k-1} \frac{i}{n}\right) = \exp\left(-\frac{(k-1)k}{2n}\right).$$

Beispielsweise ergibt sich mit den Werten $n = 365$, $k = 23$ eine Wahrscheinlichkeit von $P(A_{k,n}) \leq 0,499998$. 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). Um z.B. die Frage zu klären, wann beim wiederholten Werfen einer Münze zum ersten mal Kopf auftritt, muss eine unendliche Folge von Würfen betrachtet werden. Jeder beliebig weit in der Zukunft liegende Wurf könnte der erste sein. Das Ergebnis einer Münzwurfserie wird durch eine Folgen von Nullen und Einsen dargestellt, wobei die Nullen Zahl und die Einsen Kopf repräsentieren, also

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

mit Ω überabzählbar (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 $A = \{\omega_1, \omega_2, \ldots\} \subset \Omega$ 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 zur Konstruktion einer Wahrscheinlichkeitsverteilung 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

$$\text{(i)} \quad \Omega \in \mathfrak{A} \tag{2.2}$$

$$\text{(ii)} \quad A \in \mathfrak{A} \Rightarrow A^{\complement} \in \mathfrak{A} \tag{2.3}$$

$$\text{(iii)} \quad A_n \in \mathfrak{A},\, n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A} \tag{2.4}$$

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

$$\begin{aligned} A_n \in \mathfrak{A} \overset{\text{(ii)}}{\Rightarrow} A_n^{\complement} \in \mathfrak{A} \overset{\text{(iii)}}{\Rightarrow} \bigcup_{n=1}^{\infty} A_n^{\complement} \in \mathfrak{A} \overset{\text{(ii)}}{\Rightarrow} \left(\bigcup_{n=1}^{\infty} A_n^{\complement} \right)^{\complement} \in \mathfrak{A} \\ \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A} \;\; \forall n \in \mathbb{N} \end{aligned} \tag{2.5}$$

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 $\varepsilon = \{(a, b] \mid a < b \in \mathbb{R}\}$ 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*.

Beweis.

$$\text{(i) } \forall\, i \in I: \ \Omega \in \mathfrak{A}_i \ \Rightarrow \ \Omega \in \bigcap_{i\in I} \mathfrak{A}_i = \mathfrak{A}$$

$$\text{(ii) } A \in \mathfrak{A} \ \Rightarrow \ A \in \bigcap_{i\in I} \mathfrak{A}_i \ \Rightarrow \ \forall i: \ A \in \mathfrak{A}_i \ \Rightarrow \forall i: A^{\complement} \in \mathfrak{A}_i$$

$$\Rightarrow \ A^{\complement} \in \bigcap_{i\in I} \mathfrak{A}_i = \mathfrak{A}$$

$$\text{(iii) } \forall\, k \in \mathbb{N}: \ A_k \in \mathfrak{A} \ \Rightarrow \ \forall\, i \in I,\ k \in \mathbb{N}: \ A_k \in \mathfrak{A}_i$$

$$\Rightarrow \ \forall\, i \in I: \ \bigcup_{k\in\mathbb{N}} A_k \in \mathfrak{A}_i \ \Rightarrow \ \bigcup_{k\in\mathbb{N}} A_k \in \bigcap_{i\in I} \mathfrak{A}_i = \mathfrak{A}$$

□

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:

$$\text{(i)} \quad P(\Omega) = 1 \tag{2.6}$$

$$\text{(ii)} \quad P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{A}, \text{ paarweise disjunkt} \tag{2.7}$$

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

Der Laplacesche Wahrscheinlichkeitsbegriff ist ein Spezialfall von Definition 2.10 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) $P(A^{\complement}) = 1 - P(A)$ (2.8)

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (2.9)

c) $A \subseteq B \Rightarrow P(A) \leq P(B)$ (2.10)

d) $A \subseteq B \Rightarrow P(B \backslash A) = P(B) - P(A)$ (2.11)

e) Stetigkeit von unten:

$$\{A_n\} \text{ aufsteigend } \Rightarrow P\left(\lim_{n\to\infty}(A_n)\right) = \lim_{n\to\infty} P(A_n) \tag{2.12}$$

Stetigkeit von oben:

$$\{A_n\} \text{ absteigend } \Rightarrow P\left(\lim_{n\to\infty}(A_n)\right) = \lim_{n\to\infty} P(A_n) \tag{2.13}$$

Beweis. Aussagen a) bis d) zur Übung.
Aussage e): Teil 1: 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

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

Teil 2: $\{A_n\}$ absteigend $\Rightarrow \left\{A_n^{\complement}\right\}$ aufsteigend. Es gilt:

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

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)*

$$\begin{aligned} P\Big(\bigcup_{k=1}^{n} A_k\Big) &= \sum_{k=1}^{n}(-1)^{k+1} \sum_{1\le i_1<\ldots<i_k\le n} P\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) \\ &= \sum_{k=1}^{n} P(A_k) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) \\ &\quad + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n} A_k\Big) \end{aligned} \tag{2.14}$$

b) *Bonferroni-Ungleichung (Bonferroni inequality)*

$$\sum_{k=1}^{n} P(A_k) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) \;\le\; P\Big(\bigcup_{k=1}^{n} A_k\Big) \;\le\; \sum_{k=1}^{n} P(A_k). \tag{2.15}$$

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

Beweis. (2.14) wird mit vollständiger Induktion bewiesen. Für $n = 1$ ist die Formel trivial gültig. Für $n = 2$ erhält man

$$\begin{aligned} P(A_1\cup A_2) &= P\Big(\big((A_1\cup A_2)\setminus(A_1\cap A_2)\big)\cup(A_1\cap A_2)\Big) \\ &= P\big((A_1\cup A_2)\setminus(A_1\cap A_2)\big) + P(A_1\cap A_2) \\ &= P\big(A_1\setminus(A_1\cap A_2)\big) + P\big(A_2\setminus(A_1\cap A_2)\big) + P(A_1\cap A_2) \\ &= P(A_1) + P(A_2) - 2\cdot P(A_1\cap A_2) + P(A_1\cap A_2) \\ &= P(A_1) + P(A_2) - P(A_1\cap A_2).) \end{aligned} \tag{2.16}$$

Unter der Induktionsvoraussetzung, dass die Siebformel von Poincare-Sylvester für $n \in \mathbb{N}$ gilt, folgt mit (2.16) somit:

$$\begin{aligned} P\Big(\bigcup_{k=1}^{n+1} A_k\Big) &= P\Big(\bigcup_{k=1}^{n} A_k\Big) + P(A_{n+1}) - P\Big(\bigcup_{k=1}^{n}(A_k\cap A_{n+1})\Big) \\ &= \sum_{k=1}^{n+1} P(A_k) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n} A_k\Big) \\ &\quad - \Bigg(\sum_{k=1}^{n} P(A_k\cap A_{n+1}) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}\cap A_{n+1} \\ &\quad + - \cdots + (-1)^{n+1} P\Big(\bigcap_{k=1}^{n+1} A_k\Big)\Bigg) \\ &= \sum_{k=1}^{n+1} P(A_k) - \sum_{1\le i_1<i_2\le n+1} P(A_{i_1}\cap A_{i_2}) + - \cdots + (-1)^{n+2} P\Big(\bigcap_{k=1}^{n+1} A_k\Big), \end{aligned}$$

d.h. die Aussage für $n + 1$. Hierbei wurde die Siebformel zusätzlich auf den Term $P\Big(\bigcup_{k=1}^{n}(A_k\cap A_{n+1})\Big)$ angewandt.

Die rechte Ungleichung der Bonferroni-Ungleichung folgt mit vollständiger Induktion aus (2.16), die linke ergibt sich wie folgt:

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= P\Big(A_1 \setminus \big(A_1 \cap (A_2 \cup \ldots \cup A_n)\big) \cup \\ &\quad A_2 \setminus \big(A_2 \cap (A_3 \cup \ldots \cup A_n)\big) \cup \ldots \\ &\qquad \ldots \cup A_{n-1} \setminus (A_{n-1} \cap A_n) \cup A_n\Big) \\ &= P(A_1) - P\big((A_1 \cap A_2) \cup \ldots \cup (A_1 \cap A_n)\big) + P(A_2) \\ &\quad - P\big((A_2 \cap A_3) \cup \ldots \cup (A_2 \cap A_n)\big) + \ldots \\ &\qquad \ldots + P(A_{n-1}) - P\big((A_{n-1} \cap A_n) + P(A_n) \\ &\geq \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j). \qquad \square \end{aligned}$$

Aus Lemma 2.12 b) folgt die sogenannte *Subadditivität* von Wahrscheinlichkeitsmaßen, nämlich

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} P(A_n)$$

für beliebige Ereignisfolgen A_n, $n \in \mathbb{N}$, wie folgt.

$$\begin{aligned} P\Big(\bigcup_{n=1}^{\infty} A_n\Big) &= P\Big(\lim_{k\to\infty} \bigcup_{n=1}^{k} A_n\Big) \quad &&(\textstyle\bigcup_{n=1}^{k} A_n \text{ ist eine aufsteigende Mengenfolge}) \\ &= \lim_{k\to\infty} P\Big(\bigcup_{n=1}^{k} A_n\Big) &&(\text{wegen Lemma 2.11 e)}) \\ &\leq \lim_{k\to\infty} \sum_{n=1}^{k} P(A_n) &&(\text{wegen Lemma 2.12 b)}) \\ &= \sum_{n=1}^{\infty} P(A_n) \end{aligned}$$

Beispiel 2.13 (Sortieren (Recontre-Problem)). Betrachte ein Feld der Länge n von verschiedenen, untereinander vergleichbaren Elementen. Alle Anordnungen der Elemente seien gleichwahrscheinlich. Diese Situation kann man durch folgenden Wahrscheinlichkeitsraum modellieren.

$$\begin{aligned} &\Omega = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega \text{ ist Permutation von } \{1, \ldots, n\}\}, \\ &\mathfrak{A} = \mathfrak{P}(\Omega), \\ &P(\{\omega\}) = \frac{1}{n!} \quad \forall\, \omega \in \Omega. \end{aligned}$$

a) Bestimme die Wahrscheinlichkeit, daß mindestens ein Element an der richtigen Stelle steht, also bereits vorsortiert ist. Definiere dazu Ereignisse $A_j \in \mathfrak{A}$, die jeweils alle

Permutationen enthalten, bei denen Element j an der j-ten Stelle steht, formal

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

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

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

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

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

weil die gegebenen l Elemente an den jeweils festen richtigen Positionen stehen und die verbleibenden $(n-l)$ Elemente beliebig auf die restlichen $(n-l)$ Positionen verteilt werden können. Die Wahrscheinlichkeit dafür, daß sich l Elemente $i_1, \ldots, i_l$ auf den richtigen Positionen befinden, ist daher

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

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

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

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

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

Die Wahrscheinlichkeit dafür, daß mindestens ein Element vorsortiert ist, konvergiert somit gegen den oben angegenen Wert. Das bedeutet, daß die Wahrscheinlichkeit, mindestens ein Element in einem Feld der Länge n an der richtigen Position vorzufinden, für große n fast unabhängig von der Länge des Felds ist, ein überraschendes Ergebnis.

b) Eine obere Schranke für die Wahrscheinlichkeit, daß mindestens k Elemente vorsortiert sind, läßt sich wie folgt gewinnen.

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

Die Wahrscheinlichkeit fällt also sehr schnell mit wachsendem k.

c) Abschließend wird die Wahrscheinlichkeit bestimmt, daß *genau* k Elemente vorsortiert sind. Nach Teil a) beträgt die Wahrscheinlichkeit, in einem Feld der Länge $n-k$ kein Element vorsortiert zu finden, gerade

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

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

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

Es gibt $\binom{n}{k}$ Möglichkeiten, ein Feld der Länge n in eines der Länge k mit einem Restfeld der Länge $n-k$ aufzuteilen. Somit ergibt sich für die Wahrscheinlichkeit, daß genau k Elemente vorsortiert sind,

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

Im folgenden bahandeln wir die Frage, wie Wahrscheinlichkeiten festgelegt werden können, wenn schon bekannt ist, daß das Ergebnis in einer bestimmten Teilmenge liegt?

Beispiel. Wir betrachten 5000 Chips, die von zwei verschiedenen Firmen stammen. Firma A hat 1000 Chips und Firma B hat 4000 Chips geliefert. Unter den 5000 Chips sind 300 Chips defekt. 100 defekte Chips stammen von Firma A, 200 defekte stammen von Firma B. Ein geeignetes Modell zur Beschreibung des zufälligen, gleichverteilten Ziehens eines Chips aus dieser Menge ist das folgende.

$$\begin{aligned}
\Omega &: \text{Menge aller Chips, } |\Omega| = 5000 \\
A &: \text{Menge der Chips von Firma } A,\ |A| = 1000 \\
B &: \text{Menge der Chips von Firma } B,\ |B| = 4000 \\
D &: \text{Menge der defekten Chips, } |D| = 300 \\
A \cap D &: \text{Menge defekter Chips von Firma } A,\ |A \cap D| = 100 \\
B \cap D &: \text{Menge defekter Chips von Firma } B,\ |B \cap D| = 200
\end{aligned}$$

Ein Chip wird nun rein zufällig gezogen. Dieses Zufallsexperiment wird durch das Laplace-Modell beschrieben. Wie groß ist die Wahrscheinlichkeit, dass der Chip defekt ist, wenn er von Firma A stammt?

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

Umgekehrt lässt sich auch nach der Wahrscheinlichkeit fragen, das der Chip von Firma A stammt, wenn der defekt ist.

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

Diese Vorüberlegungen legen die folgende allgemeine Definition für den Begriff der bedingten Wahrscheinlichkeit nahe.

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

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.17}$$

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

$$\forall A \in \mathfrak{A}: \quad P(A) = \sum_{n=1}^{\infty} P(A|B_n) \cdot P(B_n) \qquad (\text{Konvention: } * \cdot 0 = 0) \tag{2.18}$$

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

$$P(B_n|A) = \frac{P(A|B_n) \cdot P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)} \tag{2.19}$$

Beweis. a):

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

b):

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

□

Wichtiger Spezialfall: Gelte

$$P(A|B) = P\left(A \middle| B^{\complement}\right), \text{ falls } 0 < P(B) < 1,$$

d.h., die Wahrscheinlichkeit für das Eintreten von A hängt nicht vom Eintreten von B ab, dann gilt

$$\begin{aligned} P(A) &= P(A \cap B) + P\left(A \cap B^{\complement}\right) = P(A|B)\, P(B) + P\left(A \middle| B^{\complement}\right) P\left(B^{\complement}\right) \\ &= P(A|B) \cdot \left(P(B) + P\left(B^{\complement}\right)\right) = P(A|B) = \frac{P(A \cap B)}{P(B)} \\ \iff & P(A)\, P(B) = P(A \cap B) \quad \text{(auch falls } P(B) = 0). \end{aligned}$$

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. $A_1, \ldots,\ A_n$ heißen (gemeinsam) *stochastisch unabhängig* (s.u.), wenn

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

Eine Folge $\{A_n\}_{n \in \mathbb{N}}$ von Ereignissen heißt stochastisch unabhängig, wenn $\forall n \in \mathbb{N}$ die Ereignisse $A_1, \cdots, A_n$ stochastisch unabhängig sind.

Beachte. Beachte: Aus paarweiser stochastischer Unabhängigkeit folgt nicht die (gemeinsame) stochastische Unabhängigkeit.

Lemma 2.17. $(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathfrak{A}$ Ereignisse. Folgende Aussagen sind äquivalent.

a) $A_1, \ldots, A_n$ sind stochastisch unabhängig.

b) Für alle $B_i \in \{A_i, A_i^c\}$ sind $B_1, \ldots, B_n$ stochastisch unabhängig.

c) Für alle $B_i \in \{A_i, A_i^c\}$ gilt $P\big(\bigcap_{i=1}^n B_i\big) = \prod_{i=1}^n P(B_i)$.

Beweis. Die Aussagen "b)$\Rightarrow$a)" und "b)$\Rightarrow$c)" sind trivial.
"a)$\Rightarrow$b)" folgt per Induktion aus der folgenden Aussage.
Für alle $\ell \in \{1, \ldots, n\}$ gilt

$$A_1, \ldots, A_\ell, \ldots, A_n \text{ stoch. unabhängig} \Rightarrow A_1, \ldots, A_\ell^c, \ldots, A_n \text{ stoch. unabhängig} \tag{2.20}$$

Um (2.20) zu zeigen, sei $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ eine beliebige Teilmenge. Falls $\ell \notin I$, gilt offensichtlich

$$P\big(\bigcap_{i \in I} A_i\big) = \prod_{i \in I} P(A_i).$$

Falls $\ell \in I$, gilt

$$\begin{aligned}\prod_{i \in I, i \neq \ell} P(A_i) &= P\big(\bigcap_{i \in I, i \neq \ell} A_i\big) = P\big(\bigcap_{i \in I, i \neq \ell} A_i \cap (A_\ell \cup A_\ell^c)\big) \\ &= P\big(\bigcap_{i \in I} A_i\big) + P\big(\bigcap_{i \in I, i \neq \ell} A_i \cap A_\ell^c\big) = \prod_{i \in I} P(A_i) + P\big(\bigcap_{i \in I, i \neq \ell} A_i \cap A_\ell^c\big).\end{aligned}$$

Durch Auflösen dieser Gleichung erhält man

$$P\big(\bigcap_{i \in I, i \neq \ell} A_i \cap A_\ell^c\big) = \big(1 - P(A_\ell)\big) \prod_{i \in I, i \neq \ell} P(A_i) = P(A_\ell^c) \prod_{i \in I, i \neq \ell} P(A_i).$$

Insgesamt folgt (2.20).
"c)$\Rightarrow$b)" sieht man wie folgt ein. Beide Seiten der folgenden Gleichungen werden addiert

$$\begin{aligned}P(B_1 \cap B_2 \cap \cdots \cap B_n) &= P(B_1)P(B_2)\cdots P(B_n), \\ P(B_1^c \cap B_2 \cap \cdots \cap B_n) &= P(B_1^c)P(B_2)\cdots P(B_n),\end{aligned}$$

mit dem Ergebnis

$$P(B_2 \cap \cdots \cap B_n) = P(B_2)\cdots P(B_n).$$

Dies kann man für jeden beliebigen Index und iteriert durchführen, so daß b) folgt. $\square$

Beispiel 2.18. Man betrachte ein Netzwerk aus 5 Komponenten (Abb. 2.1). Jede der Komponenten $K_1, \ldots, K_5$ ist 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. Das gesamte System ist intakt, wenn mindestens 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

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

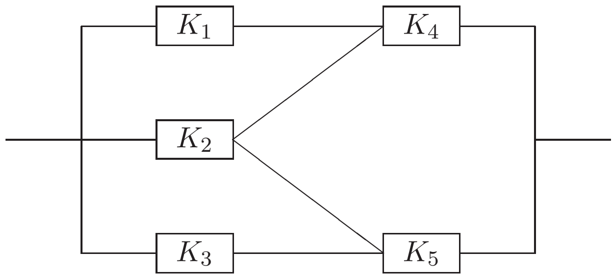


Abbildung 2.1: Netzwerk

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

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

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 dem Satz von der totalen Wahrscheinlichkeit (Satz 2.15 a)) folgt:

$$\begin{aligned} P(S) &= P(S|A_2)\,P(A_2) + P\left(S \,\middle|\, A_2^{\complement}\right)\, P\left(A_2^{\complement}\right) \text{ mit} \\ P(S|A_2) &= P(A_4 \cup A_5) = 1 - P\left(\underbrace{A_4^{\complement} \cap A_5^{\complement}}_{\text{s.u.}}\right) = 1 - P\left(A_4^{\complement}\right)\, P\left(A_5^{\complement}\right) \\ &= 1 - (1-p_4)(1-p_5) \\ P\left(S \,\middle|\, A_2^{\complement}\right) &= P\left((A_1 \cap A_4) \cup (A_3 \cap A_5)\right) = 1 - P\Big(\underbrace{(A_1 \cap A_4)}_{\text{s.u.}}{}^{\complement} \cap \underbrace{(A_3 \cap A_5)}_{\text{s.u.}}{}^{\complement}\Big) \\ &= 1 - (1-p_1p_4)(1-p_3p_5) \end{aligned}$$

und somit

$$\begin{aligned} P(S) &= (1-(1-p_4)(1-p_5))p_2 + (1-(1-p_1p_4)(1-p_3p_5)) \cdot (1-p_2) \\ &= 1 - p_2(1-p_4)(1-p_5) - (1-p_2)(1-p_1p_4)(1-p_3p_5) \\ &= 0,90062 \end{aligned}$$

Das System ist also mit einer Wahrscheinlichkeit von ca. 90% intakt. Mit der Bayes-Formel (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 $\Omega = \{\omega = (x_1, x_2, \ldots) \mid x_i \in \{0,1\}\}$. Das Ereignis, daß im n-ten Wurf Kopf fällt, ist $A_n = \{\omega = (x_1, x_2, \ldots) \mid x_n = 1\}$. Das Ereignis A sei: Es fällt unendlich oft Kopf, es treten also unendlich viele A_n ein.

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

Analoges gilt für Ereignis B: Fast alle (bis auf endlich viele) Würfe zeigen Kopf, also fast alle (bis auf endlich viele) A_n treten ein

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

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 $A_n \in \mathfrak{A},\ n \in \mathbb{N}$. Dann gelten

a)

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P(\limsup_{n\to\infty} A_n) = 0. \tag{2.21}$$

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

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P\left(\limsup_{n\to\infty} A_n\right) = 1 \tag{2.22}$$

Beweis. a): Wegen der Konvergenz der Reihe gilt

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

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): 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 $\ln x \leq x - 1$ für alle $x > 0$

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

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

Beispiel 2.22. Betrachte einen unendlichen Würfelwurf, wobei die Ergebnisse der einzelnen Würfe unabhängig voneinander sind. Sei $\Omega = \{\omega = (\omega_1, \omega_2, \ldots) \mid \omega_i \in \{1,2,3,4,5,6\}\}$. Gesucht ist nun die Wahrscheinlichkeit, daß unendlich oft die Sequenz $(1,2,3,4,5,6)$ fällt. Setze $A_n = \{\omega \mid \omega_n = 1, \ldots, \omega_{n+5} = 6\}$, also Ereignis dafür, daß ab dem n-ten Wurf die Sequenz fällt. Die Folge $\{A_{6n}\}_{n\in\mathbb{N}}$ ist stochastisch unabhängig, da Überlappungen ausgeschlossen sind. Mit dem Borel-Cantelli-Lemma folgt

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

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

Kapitel 3

Zufallsvariable 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. im Fall vom 5000 Stichproben mit dem Ausgang „gut“ oder „schlecht“, wobei nur die Anzahl der „schlechten“ Proben von Interesse ist. Bei dieser Fragestellung ist nicht mehr die ursprüngliche Ergebnismenge $\Omega = \{(x_1, \ldots, x_{5000}) \mid x_i \in \{g, s\}\}$ von Interesse, sondern das Ergebnis stammt aus der Menge $T = \{0, \ldots, 5000\}$.
Die Ergebnisse von Zufallsexperimenten sind oft Zahlen oder Vektoren. Dabei sind arithmetische Operationen oft sehr hilfreich:

- 5000-facher Münzwurf mit dem Ergebnis 0 oder 1,
  $x_1, \ldots, x_{5000} \in \{0, 1\}$.
  Von Interesse: Anzahl der Einsen $= \sum_{i=1}^{5000} x_i$

- Verzögerungszeiten an einem Switch:
  $x_1, \ldots, x_n \in \mathbb{R}^+$.
  Von Interesse: mittlere Verzögerung: $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Die Modellierung solcher Probleme erfolgt allgemein mit Zufallsvariablen.

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

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

Diese Bedingung[1] heißt *Meßbarkeit von X*.
Schreibweise: $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ (X ist eine Abbildung und meßbar).

Bemerkung. Der Begriff „Zufallsvariable“ hat sich eingebürgert, obwohl es sich eigentlich um „Funktionen“ mit Messbarkeit handelt. Betont wird die Modellierung des Zufalls, wichtig ist die *Verteilung* von Zufallsvariablen.

[1] Existenz des Urbildes

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

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

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

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

□

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

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

Modelliert 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 ein n -facher Münzwurf, in dem Kopf der Eins und Zahl der Null entspreche. Die Würfe seien unabhängig voneinander, die Wahrscheinlichkeit für Kopf sei p und die für Zahl $(1-p)$ in jedem Wurf.
Mathematisches Modell:

- $\Omega = \{\omega = (x_1, \ldots, x_n) \mid x_i \in \{0, 1\}\}$
- $\mathfrak{A} = \mathfrak{P}(\Omega)$
- $P\left(\{(x_1, \ldots, x_n)\}\right) = \underbrace{p \cdot \ldots \cdot p}_{\text{Anzahl Einsen}} \cdot \underbrace{(1-p) \cdot \ldots \cdot (1-p)}_{\text{Anzahl Nullen in}(x_1, \cdots, x_n)}$
$$= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i}$$

Die Zufallsvariable X beschreibe die Anzahl der Einsen,

$$X(\omega) = X((x_1, \dots, x_n)) = \sum_{i=1}^{n} x_i.$$

X hat den Wertebereich $T = \{0, 1, 2, \dots, n\}$.
Die Verteilung von X ist somit

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

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

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

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

Im Folgenden werden allgemeine Methoden zur Beschreibung von Wahrscheinlichkeitsverteilungen angegeben, direkt formuliert für Zufallsvariablen. Mit $X = id$ lassen sich diese folgenden Überlegungen aber 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 $T = \{t_1, t_2, \dots\}$ 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 $T = \{t_1, t_2, \dots\},\ A \in \mathfrak{B}^1$. Dann gilt

$$P^X(A \cap T) \underset{\text{La. 2.11 b)}}{=} P^X(A) + \underbrace{P^X(T)}_{=1} - \underbrace{P^X(A \cup T)}_{=1} = P^X(A).$$

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

Also gilt

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

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.

$X \sim \text{Bin}(n,p)$ (siehe Beispiel 3.3) ist ein Beispiel für eine diskrete Zufallsvariable mit Zähldichte

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

Beispiel 3.6 (geometrische Verteilung). Betrachtet werde der unendliche, unabhängige Münzwurf wie in Beispiel 3.3. Die Zufallsvariable X beschreibe die „Wartezeit“ bis zum erstmaligen Auftreten einer Eins, also die Anzahl der Würfe, bis zum ersten Mal Kopf fällt, ohne diesen Wurf mitzuzählen. Der Träger der zugehörigen Verteilung ist offensichtlich $T = \mathbb{N}_0$. Die Wahrscheinlichkeit, dass die Wartezeit genau k Würfe beträgt, also im $(k+1)$-ten Wurf erstmalig eine 1 fällt, ist

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

Diese Verteilung mit Zähldichte

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

heißt *geometrische Verteilung*. Bezeichnung:

$$X \sim \text{Geo}(p), \quad 0 < p \leq 1.$$

Es handelt sich hierbei um eine Verteilung, da $f_X(k) \geq 0\ \forall\ k \in \mathbb{N}_0$ und

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

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

$$e^{-\lambda}\frac{\lambda^k}{k!} \geq 0 \quad \forall\ k \in T \quad \text{und} \quad e^{-\lambda}\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{=e^{\lambda}} = 1.$$

Eine diskrete Zufallsvariable X mit Träger $\mathbb{N}_0$ und 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:

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

Zur Interpretation stelle man sich das Intervall $[0, 1]$ in n Stücke der Länge $\frac{1}{n}$ unterteilt vor. Dann ist die Wahrscheinlichkeit für das Auftreten eines Ereignisses, das insgesammt mit Wahrscheinlichkeit p auftritt, in jedem Teilstück durch $p_n = \frac{\lambda}{n}$ gegeben. Bei stochastischer Unabhängigkeit ist die Gesamtzahl des Auftretens der Ereignisse $\mathrm{Bin}(n, p_n)$-verteilt. Die zugehörige Zähldichte konvergiert mit $n \to \infty$ gegen die Zähldichte einer $\mathrm{Poi}(\lambda)$-Verteilung.

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).

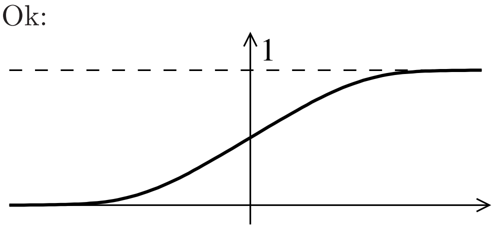
(ii) $\lim\limits_{x\to+\infty} F(x) = 1$ und $\lim\limits_{x\to-\infty} F(x) = 0$

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

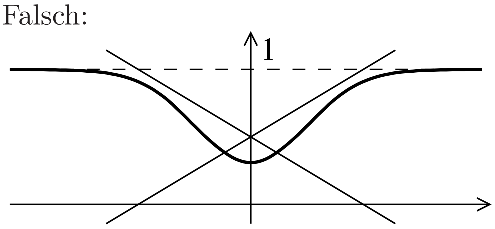
heißt *Verteilungsfunktion* (VF) (engl.: (cumulative) *distribution function* (cdf)).

Beispiele:

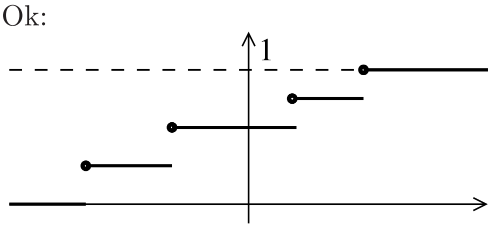
Ok:



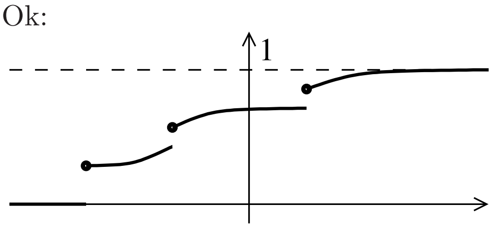
Falsch:



Ok:



Ok:



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

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) = P^X((-\infty, x]), \quad x \in \mathbb{R}$$

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

Beweis. Die drei Bedingungen aus Definition 3.8 sind nachzuweisen.

(i) Seien $x \leq y \in \mathbb{R}$. Dann gilt $(-\infty, x] \subseteq (-\infty, y]$, und mit Lemma 2.11 c) folgt $P^X\big((-\infty, x]\big) \leq P^X\big((-\infty, y]\big)$, d.h. $F_X(x) \leq F_X(y)$.

(ii) Sei $\{x_n\}$ eine monoton steigende Folge mit $\lim_{n\to\infty} x_n = \infty$. Die Mengenfolge $(-\infty, x_n]$ ist dann aufsteigend, ihr Limes ist wohldefiniert, und es folgt mit Lemma 2.11 e), daß

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P^X\big((-\infty, x_n]\big) = P^X\big(\lim_{n\to\infty}(-\infty, x_n]\big) = P^X(\mathbb{R}) = 1.$$

Sei ferner $\{x_n\}$ eine monoton fallende Folge mit $\lim_{n\to\infty} x_n = -\infty$. Analog gilt

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P^X\big((-\infty, x_n]\big) = P^X\big(\lim_{n\to\infty}(-\infty, x_n]\big) = P^X(\emptyset) = 0.$$

(iii) Sei schließlich $x_0 \in \mathbb{R}$ und $\{x_n\}$ eine monoton fallende Folge mit $\lim_{n\to\infty} x_n = x_0$. Die rechtsseitige Stetigkeit folgt wieder mit Lemma 2.11 e), da

$$\begin{aligned}\lim_{n\to\infty} F_X(x_n) &= \lim_{n\to\infty} P^X\big((-\infty, x_n]\big) \\ &= P^X\big(\lim_{n\to\infty}(-\infty, x_n]\big) = P^X\big((-\infty, x_0]\big) = F_X(x_0).\end{aligned}$$

□

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

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

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

Ferner existiert zu jeder Verteilungsfunktion F genau eine Verteilung P auf $(\mathbb{R}^1, \mathfrak{B}^1)$ mit $F(x) = P\big((-\infty, x]\big)$ für alle $x \in \mathbb{R}$.

Beweis.

"$\Rightarrow$" einfach

"$\Leftarrow$" benutzt den Fortsetzungs- und Eindeutigkeitssatz der Maßtheorie

□

Beispiel 3.11 (Verteilungsfunktionen).

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

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

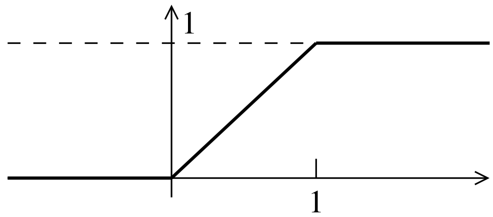


Abbildung 3.1: Verteilungsfunktion der Rechteckverteilung auf $[0,1]$.

Bezeichnung: $X \sim \mathrm{R}(0,1)$.

Seien $0 \le a < b \le 1$. Dann gilt

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

Die Wahrscheinlichkeit, daß unter der $R(0,1)$-Verteilung ein zufälliger Wert im Intervall $[a,b]$ liegt, ist also gleich der Länge des Intervalls $(a,b]$, nämlich $b-a$.

Analog ist die Rechteckverteilung auf einem Intervall $[a,b]$ definiert. Eine Zufallsvariable heißt rechteckverteilt auf dem Intervall $[a,b]$, bezeichnet mit $X \sim \mathrm{R}(a,b)$, $a < b \in \mathbb{R}$, wenn

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

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

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{falls} \quad x \ge 0 \\ 0, & \text{falls} \quad x \le 0 \end{cases} = (1 - e^{-\lambda x})\, \mathbb{I}_{[0,\infty)}(x).$$

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

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

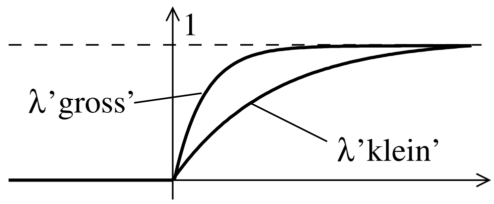


Abbildung 3.2: Verteilungsfunktion der Exponentialverteilung

c) Sei X eine diskrete Zufallsvariable mit geordnetem Träger $T = \{t_1, t_2, \ldots\} \subset \mathbb{R}$, $t_1 < t_2 < t_3 < \ldots$, und der Zähldichte $f_X(t_i) = p_i,\ p_i \geq 0,\ \sum_{i=1}^{\infty} p_i = 1$. Dann ist

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

Hierbei wird $t_0 := -\infty$ gesetzt.

Wie in Abbildung 3.3 dargestellt, ergibt sich eine Treppenfunktion mit Sprungstellen bei t_k der Höhe p_k, $k = 1, 2, \ldots$.

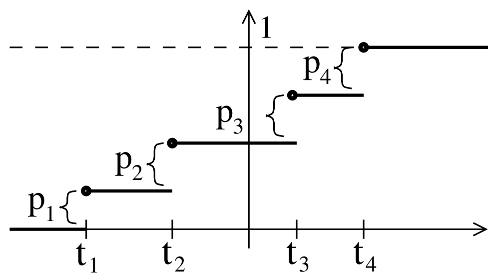


Abbildung 3.3: Treppenfunktion

Sei beispielsweise X geometrisch verteilt ($X \sim \mathrm{Geo}(p)$), mit der Zähldichte

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

Es gilt

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

Also ist

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

die Verteilungsfunktion der geometrischen Verteilung, wobei $\lfloor x \rfloor$ die größte ganze Zahl kleiner oder gleich x bezeichnet.

3.2.1 Berechnung von Wahrscheinlichkeiten durch Verteilungsfunktionen

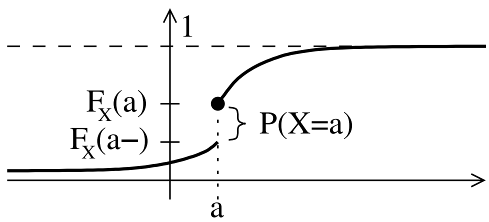
Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x) = P(X \leq x)$ und $a < b \in \mathbb{R}$.

- Es gilt

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

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

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



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

- Es gilt

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

Im folgenden Beispiel wird die Bedeutung des Begriffs *Quantil* motiviert und erklärt. Die Bedienzeit X von Anforderungen an einem Server sei $Exp(\lambda)$-verteilt, $\lambda > 0$. Bestimme nun die Zeit x_α, unterhalb derer die Bedienzeit mit vorgegebender Wahrscheinlichkeit α liegt. Bestimme also x_α mit $P(X \leq x_\alpha) = \alpha$. Im Fall $\alpha = 0,99$ kann x_α als die Zeit interpretiert werden, unterhalb derer die Bedienzeit in 99% der Fälle liegt.

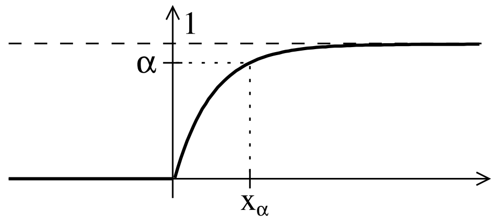


Abbildung 3.4: Verteilungsfunktion der Exp(λ)-Verteilung, $F_X(x) = P(X \leq x) = 1 - e^{-\lambda x}$, $x \geq 0$, mit eingezeichnetem α und zugehörigem x_α.

Definition 3.12. Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x)$, ferner $0 \leq \alpha \leq 1$. Dann heißt

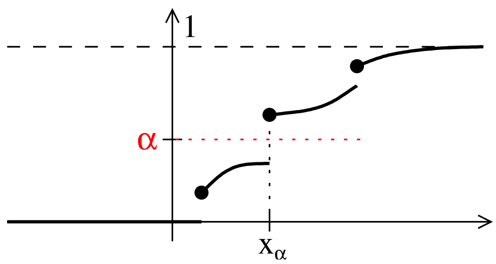
$$x_\alpha = \min\{x \mid F_X(x) \geq \alpha\}$$

das α-*Quantil* $\big($oder α-*Percentil* oder $(1-\alpha)$-*Fraktil*$\big)$ von F_X. Allgemein heißt

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

die *Pseudoinverse* von F_X.

- Diese Definition läßt sich graphisch wie folgt darstellen.



  Wegen der rechtsseitigen Stetigkeit liegt der Wert von $F(x_\alpha)$ bei Unstetigkeitsstellen stets oberhalb von α. Es folgt also, daß $F_X(x_\alpha) \geq \alpha$ für alle $\alpha \in (0,1)$.

- Ist F invertierbar, so gilt $F^- = F^{-1}$, die Inverse von F_X.

Insbesondere heißt für $\alpha = \frac{1}{2}$ der Wert $x_{\frac{1}{2}}$ der *Median* von F_X.
Es gilt $P\big(X \leq x_{\frac{1}{2}}\big) \geq \frac{1}{2}$ und $P\big(X < x_{\frac{1}{2}}\big) \leq \frac{1}{2}$. Der Median „halbiert" also die Verteilung. In diesem Sinn ist er der „mittlere" Wert. Der Median einer Stichprobe ist ein sogenannter *Schätzer* (siehe später) für den mittleren Wert der zugrunde liegenden Verteilung. Weil ungewöhnlich große oder kleine Werte keinen Einfluß auf den Median haben, ist er robust gegen sogenannte *statistische Ausreißer*.

3.3 Dichten

Im folgenden Abschnitt werden sogenannte Verteilungsdichten oder auch Dichten eingeführt. Dichten sind nichtnegative, reellwertige Funktionen aus denen durch Integration Verteilungsfunktionen gewonnen werden können. Dichten bilden also eine weitere Möglichkeit, Verteilungen zu beschreiben.

Definition 3.13. $f: \mathbb{R} \to \mathbb{R}^+$ sei eine (uneigentlich Riemann-) integrierbare Funktion 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* (engl.: *probability density function* (*pdf*)) von X (bzw. P^X). X (bzw. P^X) heißt dann *absolut-stetig*.

Mit dem Hauptsatz der Differential- und Integralrechnung folgt:

$$f(x) \;=\; F'(x)$$

für alle Stetigkeitspunkte x von f.

Geometrisch kann der Zusammenhang zwischen Dichten und Verteilungsfunktionen wie folgt veranschaulicht werden.

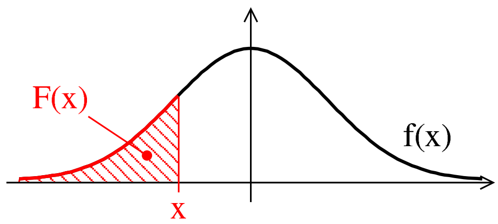
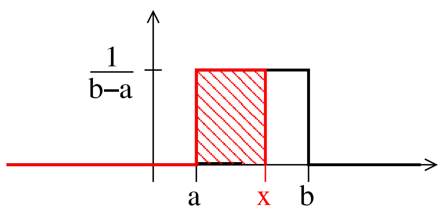


Abbildung 3.5: Der Wert der Verteilungsfunktion $F_X(x)$ an der Stelle x ist durch die Fläche unter der Dichte bis zum Punkt x gegeben.

Beispiel 3.14.

a) Die *Rechteckverteilung* auf $[a, b]$, $a < b \in \mathbb{R}$ ($X \sim \mathrm{R}(a, b)$) besitzt eine Dichte der Form

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



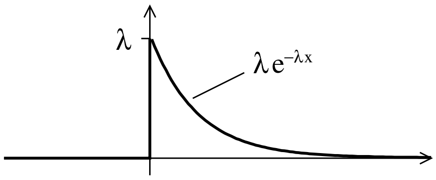
Denn die bekannte Verteilungsfunktion entsteht hieraus durch Integration wie folgt.

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

Beachte: 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 nur „fast sicher eindeutig“[3].

b) Die *Exponentialverteilung* $(X \sim \mathrm{Exp}(\lambda),\ \lambda > 0)$ besitzt eine Dichte der Form

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

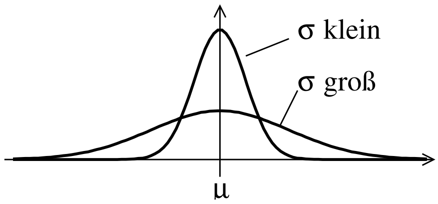


Die zugehörige Verteilungsfunktion entsteht wie folgt.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} \lambda e^{-\lambda t}dt\ \mathbb{I}_{[0,\infty)}(x) \\ &= -e^{-\lambda t}\Big|_{0}^{x}\ \mathbb{I}_{[0,\infty)}(x) = (1 - e^{-\lambda x})\ \mathbb{I}_{[0,\infty)}(x). \end{aligned}$$

c) Die *Normalverteilung* $(X \sim \mathrm{N}(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma > 0)$ besitzt die folgende Dichte.

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



Die Verteilungsfunktion hat keine geschlossene Darstellung, sie hat die Gestalt

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

Die Berechnung erfolgt daher numerisch. Ihre Werte können auch aus Tabellen der Standard-Normalverteilung mit Parametern $\mu = 0$ und $\sigma = 1$ wie folgt berechnet werden. Sei hierzu $Y \sim \mathrm{N}(0, 1)$ mit der Verteilungsfunktion $F_Y(y) = \Phi(y)$ und $X \sim \mathrm{N}(\mu, \sigma^2)$. Es gilt

$$F_X(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{\frac{t^2}{2}} dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Der folgende Zusammenhang erlaubt die Berechnung des Funktionswerts negativer Argumente aus positiven, so daß Tabellen auf die Angabe für posititve Argumente beschränkt werden können.

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

[3] siehe später

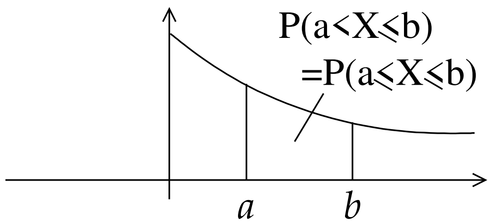
3.3.1 Berechung von Wahrscheinlichkeiten mit Dichten

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

- Es gilt

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

  Die Wahrscheinlichkeit $P(a < X \leq b)$ ergibt sich als die Fläche zwischen der x-Achse und dem Graphen der Dichte wie in der folgenden Abbildung.



- 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

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

- 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 $T = \{t_0, t_1, t_2, \ldots\}$ und Zähldichte $f_X(t_k) = p_k,\ k \in \mathbb{N}_0$. Dann heißt

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

(ex. für $|z| \leq 1$, da $\sum_{k=0}^{\infty} p_k$ Majorante mit Wert 1.)

erzeugende Funktion von X bzw. P^X (auch *Z-Transformation*, engl.: *probability generating function*).

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

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

(ex. für $s \geq 0$, da $e^{-sx} \leq 1 (x \geq 0)$ und $\int_0^{\infty} e^{-sx} f(x)dx \leq \int_0^{\infty} f(x)dx = 1$.)

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 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-stetige Zufallsvariablen mit $f_X(x) = f_Y(x) = 0$ für alle $x < 0$. X und Y besitzen dieselbe Verteilung genau dann, wenn

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

Beweis.

a) Eindeutigkeitssatz für Potenzreihen

b) Feller II, p. 408, Chapter XIII,1 oder Satz von Stone-Weierstraß

□

Also: Transformierte bestimmen die Verteilung eindeutig.

Satz 3.17 (Inversionsformeln).

a) $G_X(z)$ sei die erzeugende Funktion einer diskreten Zufallsvariablen X mit Träger $T = \{t_0, t_1, t_2, \ldots\}$. Dann gilt

$$P(X = t_k) = \frac{1}{k!} G_X^{(k)}(0) \qquad (k\text{-te Ableitung von } G_X \text{ an der Stelle } 0).$$

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

$$f_X(x) = \lim_{y\to\infty} \frac{1}{2\pi i} \int_{c-iy}^{c+iy} e^{sx} L_X(s) ds \quad \text{für genügend großes } c > 0.$$

Diese Formel setzt die Definition der Laplace-Transformierten im Komplexen voraus.

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

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

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

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

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

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

$$\begin{aligned} L_X(s) &= \int_0^{\infty} e^{-sx} \cdot \mathbb{I}_{[0,1]}(x) dx = \int_0^1 e^{-sx} dx \\ &= -\frac{1}{s} e^{-sx} \Big|_0^1 = -\frac{1}{s} e^{-s} + \frac{1}{s} = \frac{1 - e^{-s}}{s}, \quad s \geq 0 \end{aligned}$$

d) Exponentialverteilung ($X \sim \mathrm{Exp}(\lambda)$)

$$\begin{aligned} L_X(s) &= \int_0^{\infty} e^{-sx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(s+\lambda)x} dx \\ &= \frac{\lambda}{\lambda + s} \underbrace{\int_0^{\infty} (\lambda + s) e^{-(s+\lambda)x} dx}_{=1, \text{ da Int. über Dichte}} = \frac{\lambda}{\lambda + s}, \quad s \geq 0 \end{aligned}$$

Kapitel 4

Produkträume und Zufallsvektoren

Dieses Kapitel behandelt Zufallsexperimente, bei denen mehrere Ausgänge beobachtet werden können. Dies beinhaltet auch den Fall, daß dasselbe Experiment mehrfach wiederholt wird.

$$
\begin{array}{lll}
(\Omega, \mathfrak{A}, P) & \xrightarrow{X_1(\omega)} & (\mathbb{R}^1, \mathfrak{B}^1) \\
 & \xrightarrow{X_2(\omega)} & (\mathbb{R}^1, \mathfrak{B}^1) \\
 & \vdots & \\
 & \xrightarrow{X_n(\omega)} & (\mathbb{R}^1, \mathfrak{B}^1)
\end{array}
$$

Bei gemeinsamer Betrachtung aller Ausgänge wird sinnvollerweise eine Darstellung mit einem Zufallsvektor $X(\omega) = \big(X_1(\omega), \ldots, X_n(\omega)\big)$ verwendet. Die Frage stellt sich insbesondere nach einem geeigneten Wahrscheinlichkeitsraum auf der Bildseite der vektorwertigen Abbildung X.

$$
(\Omega, \mathfrak{A}, P) \xrightarrow{X=(X_1,\ldots,X_n)} (\mathbb{R}^n, \;\; ?, \;\; ?)
$$

Welche gemeinsame Verteilung ?

Welche $\sigma-$Algebra ?

4.1 Produkträume

Gegeben seien n Meßräume $(\Omega_i, \mathfrak{A}_i),\ i = 1, \ldots, n$. Häufig, aber nicht notwendigerweise, gilt $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}^1, \mathfrak{B}^1)$ für alle $i = 1, \ldots, n$. Setze

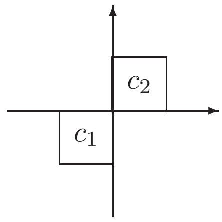
$$
\Omega = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_i \in \Omega_i\}.
$$

Um einen Wahrscheinlichkeitsraum zu erhalten, muß über Ω eine $\sigma - Algebra$ konstruiert werden. Ein Ansatz wäre, diese durch

$$\mathcal{E} = \{A_1 \times \ldots \times A_n \mid A_i \in \mathfrak{A}_i,\ i = 1, \ldots, n\}$$

zu definieren. $\mathcal{E}$ 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

$$\begin{aligned} C_1 &= [-1, 0] \times [-1, 0] \in \mathcal{E} \\ C_2 &= [0, 1] \times [0, 1] \in \mathcal{E}, \end{aligned}$$



aber

$$C_1 \cup C_2 \notin \mathcal{E},$$

weil die Vereinigung offensichtlich nicht als kartesisches Produkt zweier Menge darstellbar ist. Ein naheliegender Schritt ist, die von $\mathcal{E}$ erzeugte σ-Algebra zu verwenden.

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

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

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 der Meßräume $(\Omega_i, \mathfrak{A}_i),\ i = 1, \ldots, n$.

Beispiel 4.2. Sei $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}^1, \mathfrak{B}^1)$ und $\mathcal{E} = \{B_1 \times \ldots \times B_n \mid B_i \in \mathfrak{B}^1\}$. Dann heißt $\mathfrak{A}(\mathcal{E}) = \mathfrak{B}^n$ die *n-dimensionale Borelsche-σ-Algebra*.
Im Fall $n = 2$ enthält $\mathfrak{B}^n$ beispielsweise alle

- Rechtecke,
- alle abzählbaren Vereinigungen von Rechtecken,
- alle Linien (darstellbar durch abzählbare Schnitte und Vereinigungen von Rechtecken),
- „Dreiecksflächen“ ($\{x_i \geq 0 \mid \sum_{i=0}^{n} x_i \leq t\}$),
- Kreisflächen ($\{x_1^2 + x_2^2 \leq c\}$), usw...

4.2 Zufallsvektoren und Folgen von Zufallsvariablen

Definition 4.3. Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung

$$X = (X_1, \dots, X_n) : \ \Omega \to \mathbb{R}^n$$

heißt *Zufallsvektor*, wenn $X^{-1}(B) \in \mathfrak{A} \quad \forall\, B \in \mathfrak{B}^n$ (Meßbarkeit). Bezeichnung:

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

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

Lemma 4.4. Sei X ein Zufallsvektor. Durch

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

wird eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}^n, \mathfrak{B}^n)$ definiert, die Verteilung des Zufallsvektors X. $P^X = P^{(X_1,\dots,X_n)}$ heißt *gemeinsame Verteilung* von $(X_1, \dots, X_n)$.

Im folgenden werden gemeinsame Verteilungen mit n-dimensionalen Verteilungsfunktionen und Dichten beschrieben. Falls $X_1, \dots, X_n$ alle diskret sind, so besitzt auch der Zufallsvektor $X = (X_1, \dots, X_n)$ diese Eigenschaft. Die gemeinsame Verteilung kann dann wie in Kapitel 3 durch eine diskrete Zähldichte beschrieben werden.

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

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

Bezeichnung: $X \sim F_X$.

Die Beschreibung von Verteilungen mit n-dimensionalen Dichten ist oft einfacher.

Definition 4.6. Sei $F_X(x_1, \dots, x_n)$ die Verteilungsfunktion des Zufallsvektors X. Eine (uneigentlich Riemann-) integrierbare Funktion $f_X : \ \mathbb{R}^n \to \mathbb{R}^+$ heißt *Dichte* von X (bzw. F_X oder P^X), wenn

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

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 Verteilungsfunktion F_X“.

Wahrscheinlichkeiten können mit Hilfe von Dichten wie folgt berechnet werden. Sei $X = (X_1, \ldots, X_n)$ absolut-stetig mit Verteilung P^X, Verteilungsfunktion F_X und Dichte f_X. Dann gilt für alle $a_i \leq b_i \in \mathbb{R}$

$$\begin{aligned} P^X(\langle a_1, b_1\rangle \times \ldots \times \langle a_n, b_n\rangle) &= P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) \\ &= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_X(t_1, \ldots, t_n) dt_1 \ldots dt_n \end{aligned}$$

wobei „⟨⟩“ beliebig für „()“, „[]“, „[)“ oder „(]“ stehen.
Allgemein gilt für $B \in \mathfrak{B}^n$

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

Mit dem Hauptsatz der Differential- und Integralrechung folgt, daß

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

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 die Dichte

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

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

- Speziell gilt für die Gleichverteilung auf dem Einheitswürfel des $\mathbb{R}^n$

$$T = \{(x_1, \ldots, x_n) \mid 0 \leq x_i \leq 1,\ i = 1, \ldots, n\} = [0,1]^n$$

$$f(x_1, \ldots, x_n) = \mathbb{I}_T(x_1, \ldots, x_n) = \begin{cases} 1 & , 0 \leq x_i \leq 1, \quad i = 1 \ldots n \\ 0 & , \text{sonst} \end{cases}$$

$$F(x_1, \ldots, x_n) = \begin{cases} 0 & , \quad \exists\, i:\ x_i < 0 \\ \min\{x_1, 1\} \cdot \ldots \cdot \min\{x_n, 1\} & , \quad \text{sonst.} \end{cases}$$

- Die Dichte der Gleichverteilung auf der Einheitskugel des $\mathbb{R}^n$ lautet

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

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

Zur Erinnerung: Für $n \in \mathbb{N}$ gilt $\Gamma(n) = (n-1)!$. Ferner $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ und $\Gamma(x+1) = x\Gamma(x)$. Die Verteilungsfunktion ist schwierig zu berechnen. Man benötigt dazu Dirichlet-Integrale.

b) Die n-dimensionale Normalverteilung $N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^n$ und positiv definiter Matrix $\Sigma \in \mathbb{R}^{n\times n}$ hat die Dichte

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

c) $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}^+$ seien n-dimensionale Dichten (auch $n = 1$) und $\alpha_1, \ldots, \alpha_k \geq 0$ 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 $f_1, \ldots, f_n$. In der Tat liegt mit $f(x)$ eine Dichte vor, da $f(x) \geq 0$ und

$$\int f(x)dx = \int \sum \alpha_i f_i(x)dx = \sum \alpha_i \underbrace{\int f_i(x)dx}_{=1} = \sum \alpha_i = 1.$$

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

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

für alle $x_1, \ldots, x_n \in \mathbb{R}$. Dies bedeutet in ausführlicher Schreibweise

$$P(\{\omega \mid X_1(\omega) \leq x_1\} \cap \ldots \cap \{\omega \mid X_n(\omega) \leq x_n\})$$
$$= P(\{\omega \mid X_1(\omega) \leq x_1\}) \cdots P(\{\omega \mid X_n(\omega) \leq x_n\})$$

für alle $x_1, \ldots, x_n \in \mathbb{R}$. $X_1, \ldots, X_n$ sind also stochastisch unabhängig genau dann, wenn die Ereignisse

$$\{\omega \mid X_i(\omega) \leq x_i\}, \; i = 1, \ldots, n,$$

für alle $x_1, \ldots, x_n$ stochastisch unabhängig im Sinn von Definition 2.16 sind. Die dort geforderte Auswahl von beliebigen Indizes $i_1, \ldots, i_k$ für die Produktbeziehung wird erreicht, indem die Werte $x_j = \infty$ gesetzt werden, sofern $j \notin \{i_1, \ldots, i_k\}$.

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

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

Gilt nun

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

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

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

eine Dichte des Zufallsvektors $(X_1, \dots, X_n)$.

Beweis.

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

ist eine Dichte von X_i, da

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

Nun gilt

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

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

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$
$$= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(t_i) dt_i = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \prod_{i=1}^{n} f_{X_i}(t_i)\ dt_1 \dots dt_n.$$

Somit ist $\prod_{i=1}^{n} f_{X_i}(t_i)$ eine Dichte von $(X_1, \dots, X_n)$. □

Im Fall von Zähldichten vereinfacht sich Lemma 4.9 zu

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

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

Beweis. Zur Übung. □

Definition 4.11. Eine Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ heißt *stochastisch unabhängig*, wenn jeweils die ersten n Zufallsvariablen $X_1, \ldots, X_n$ 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* (engl: *iid*: „independent identically distributed".).

Beispiel 4.12.

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

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

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

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

Beispiel 4.13. Seien $X_1, \ldots, X_n$ stid $\sim F$ (Verteilungsfunktion).

a) Sei $Y = \max\{X_1, \ldots, X_n\}$.
Dies könnte z.B. interpretiert werden als das Maximum der „Laufzeiten" von parallelen, unabhängigen Prozessen $X_1, \ldots, X_n$.
Für alle $x \in \mathbb{R}$ gilt dann

$$P(Y \le x) = P(X_1 \le x, \ldots, X_n \le x) = \prod_{i=1}^{n} P(X_i \le x) = F^n(x).$$

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

Sei beispielsweise $X_i \sim \mathrm{R}(0,1)$, mit

$$F_{X_i}(x) = \begin{cases} 0 & , x \leq 0 \\ x & , 0 \leq x \leq 1 \\ 1 & , x \geq 1 \end{cases} \quad \Rightarrow \quad F^n(x) = \begin{cases} 0 & , x \leq 0 \\ x^n & , 0 \leq x \leq 1 \\ 1 & , x \geq 1 \end{cases}$$

b) Sei $Y = \min\{X_1, \dots, X_n\}$. Dann gilt für alle $x \in \mathbb{R}$

$$\begin{aligned} P(Y > x) &= P(X_1 > x, \dots, X_n > x) = P(X_1 > x) \cdots P(X_n > x) \\ &= (1 - P(X_1 \leq x)) \cdots (1 - P(X_n \leq x)) = (1 - F(x))^n . \end{aligned}$$

Also hat $Y = \min\{X_1, \dots, X_n\}$ die Verteilungsfunktion $1 - (1 - F(x))^n$.

Beispielsweise gilt für X_i stid $\sim \mathrm{Exp}(\lambda)$, $\lambda > 0$,

$$P(Y \leq x) = 1 - e^{-n\lambda x}\ , \quad x \geq 0.$$

Das Minimum von stochastisch unabhängigen $\mathrm{Exp}(\lambda)$-verteilten Zufallsvariablen ist also wieder exponentialverteilt, allerdings mit Parameter $n\lambda$.

Beispiel 4.14. Seien X_1 und X_2 stochastisch unabhängig, identisch jeweils $\mathrm{R}(0,1)$-verteilt. Eine gemeinsame Dichte lautet dann

$$f_{(X_1,X_2)}(x_1, x_2) = \mathbb{I}_{[0,1]}(x_1) \cdot \mathbb{I}_{[0,1]}(x_2) = \mathbb{I}_{[0,1]^2}(x_1, x_2), \quad x_1, x_2 \in \mathbb{R}.$$

Uns interessiert die Verteilung der Summe $Z = X_1 + X_2$. Hierzu wird zunächst die Verteilungsfunktion von Z bestimmt. Offensichtlich gilt

$$F_Z(z) = \begin{cases} 0, & \text{falls } z < 0 \\ 1, & \text{falls } z > 2 \end{cases}.$$

Es bleibt also, die Verteilungsfunktion für Argumente $0 \leq z \leq 2$ zu bestimmen. Hier gilt

$$F_z(z) = P(X_1 + X_2 \leq z) = \iint\limits_{0 \leq x_1 + x_2 \leq z} \mathbb{I}_{[0,1]^2}(x_1, x_2)\, dx_1 dx_2$$

$$= \iint\limits_{\substack{0 \leq x_1, x_2 \leq 1 \\ 0 \leq x_1 + x_2 \leq z}} 1\, dx_1 dx_2 = \begin{cases} \frac{z^2}{2}, & \text{falls } 0 \leq z \leq 1 \\ 1 - \frac{(2-z)^2}{2}, & \text{falls } 1 \leq z \leq 2 \end{cases}.$$

Die Gültigkeit dieser Formel macht man sich relativ leicht geometrisch klar.

Durch Differenzieren von F_Z ergibt sich eine Dichte f_Z der Zufallsvariablen Z wie folgt

$$f_Z(z) = \begin{cases} z & \text{falls } 0 \leq z \leq 1 \\ 2 - z & \text{falls } 1 \leq z \leq 2 \ . \\ 0, & \text{sonst} \end{cases}$$

Wegen der Gestalt ihrer Dichte heißt die zugehörige Verteilung *Dreieckverteilung* auf $[0, 2]$, siehe Abbildung 4.1.

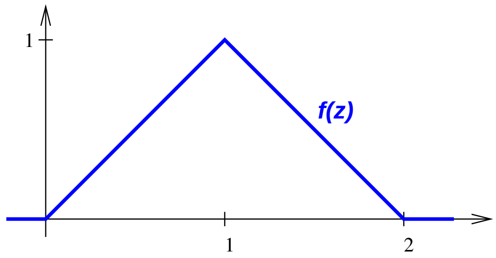


Abbildung 4.1: Dichte der Dreieckverteilung auf dem Intervall $[0,2]$.

Beispiel 4.15. X_1 und X_2 seien stochastisch unabhängige Zufallsvariable, identisch exponentialverteilt mit Parameter $\lambda > 0$. $\big(X_1, X_2 \text{ stid } \sim \operatorname{Exp}(\lambda)\big)$. Dann gilt

$$
\begin{aligned}
&P(X_1 + X_2 \leq z) \\
&= \iint\limits_{x_1+x_2\leq z} \lambda e^{-\lambda x_1} \cdot \mathbb{I}_{(0,\infty)}(x_1) \cdot \lambda e^{-\lambda x_2} \cdot \mathbb{I}_{(0,\infty)}(x_2) dx_1 dx_2 \\
&= \int_0^z \int_0^{z-x_2} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} dx_1 dx_2 = \int_0^z \lambda e^{-\lambda x_2} \left[1 - e^{-\lambda(z-x_2)}\right] dx_2 \\
&= \int_0^z \left(\lambda e^{-\lambda x_2} - \lambda e^{-\lambda z}\right) dx_2 = 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} \\
&= F_{X_1+X_2}(z), \quad z \geq 0.
\end{aligned}
$$

Durch Differenzieren nach z erhält man die Dichte

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

Es folgt also, daß die Verteilung der Summe von stochastisch unabhängigen, identisch $\operatorname{Exp}(\lambda)$-verteilten Zufallsvariablen absolut-stetig ist 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, $\alpha, \lambda > 0$ zwei Parameter,

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

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 sind

(i) $\alpha = 1$: Exp(λ)-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

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

Bezeichnung: Erl(n, λ)
Verteilung der Summe von n i.i.d. Exp(λ)-verteilten Zufallsvariablen (s. später).

Beispiel 4.16. Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen, identisch verteilten Zufallsvariablen und $B \in \mathfrak{B}^1$. Dann heißt

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

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

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

wobei $p := P(X_1 \in B)$. S ist also geometrisch verteilt mit dem Träger $\mathbb{N}$. Anders ausgedrückt, ist $S - 1 \sim Geo(p)$.
Konkret betrachte man z.B. eine unabhängige Folge von Münzwürfen, so daß $X_n,\ n \in \mathbb{N}$, stid $\sim Bin(1, p)$. Sei $B = \{1\}$, 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 $X \sim Geo(p)$.

Der folgende Satz besagt, daß die stochastische Unabhängigkeit erhalten bleibt, wenn man stochastisch unabhängige Zufallsvariablen disjunkt zu Vektoren zusammengruppiert oder Funktionen hiervon betrachtet.

Satz 4.17. $X_1, \dots, X_n$ seien stochastisch unabhängige Zufallsvariablen und $I, J \subseteq \{1, \dots, n\}$ 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.

Beweis. MaPf La. 2.1.6, 2.1.7, S. 74 $\square$

Beispielweise gilt für stochastisch unabhängige Zufallsvariablen $X_1, \dots, X_4$, daß $Y = X_1 + X_2$ und $Z = X_3 + X_4$ stochastisch unabhängig sind.

Kapitel 5

Transformationen von Zufallsvariablen und Verteilungen

Im folgenden werden allgemeine Hilfsmittel zur Berechnung der Verteilung von Funktionen von Zufallsvariablen behandelt. Zur Motivation:

(i) X_1, X_2 i.i.d. $\sim Exp(\lambda)$ (Laufzeiten)
Frage: $X_1 + X_2 \sim ?$ (Summe der Laufzeiten, s.o.)

(ii) X_1, X_2 i.i.d. $\sim Exp(\lambda)$
Frage: $(Y_1, Y_2) = (\min\{X_1, X_2\}, \max\{X_1, X_2\}) \sim ?$
(2-dim. Verteilung)

(iii) (U_1, U_2, U_3) gleichverteilt auf $B = \{(u_1, u_2, u_3) \mid 0 < u_1 < u_2 < u_3 < a\}$, wobei $a > 0$ fest.
Frage: $(Y_1, Y_2, Y_3) = \Big(\underbrace{U_1 \cdot U_2 \cdot U_3}_{Volumen}, \underbrace{2(U_1 \cdot U_2 + U_2 \cdot U_3 + U_1 \cdot U_3)}_{Oberfläche}, \underbrace{4(U_1 + U_2 + U_3)}_{Kantenlänge} \Big) \sim ?$
(3-dim. Verteilung)

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

$$f_X(x_1, \ldots, x_n) = 0 \quad \forall (x_1, \ldots, x_n) \in M^{\complement} \text{ für eine offene Menge } M \subseteq \mathbb{R}^n.$$

Des weiteren 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 ($\widetilde{T}$: Restriktion von T auf M).

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

(iii) Für die *Funktionaldeterminante* gilt

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

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

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

Beweis. z.B. Krickeberg (63) □

Im folgenden nehmen wir für die Funktion T direkt den Definitionsbereich M an und vermeiden hiermit die Notation $\widetilde{T}$.

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

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

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 inkl. Nullpunkt, da Q offenes Intervall). Also ist $Y = T(X)$ gleichverteilt auf dem Einheitskreis (vgl. Beispiel 4.7).

Beispiel 5.3 (Rayleigh-Verteilung). Seien X, Y zwei stochastisch unabhängige, identisch normalverteilte Zufallsvariablen, $X, Y \sim \mathrm{N}(0, \sigma^2)$. Die gemeinsame Dichte ist somit gegeben durch das Produkt der Randdichten,

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

Die Abbilung

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

mit

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

transformiert auf Polarkoordinaten. Die zugehörige Umkehrabbildung ist $T^{-1}: \ (r,\varphi) \mapsto (r\cos\varphi, r\sin\varphi)$. Zur Anwendung des Transformationssatz wird

$$\det\left(\frac{\partial T_i^{-1}}{\partial z_j}\right)_{1\le i,j\le 2} = \det\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = r\left(\cos^2\varphi + \sin^2\varphi\right) = r \neq 0$$

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

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

R, Φ sind somit stochastisch unabhängig mit Dichten

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

($R \sim \mathrm{Ray}(\sigma^2)$ heißt *Rayleigh-verteilt*), und

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

(Φ ist also rechteckverteilt auf $[0, 2\pi)$).

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 für $T(x_1,x_2)=(x_1,x_1+x_2)$ mit $T^{-1}(y_1,y_2)=(y_1,y_2-y_1)$, daß

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

Also gilt

$$f_Y(y_1,y_2)=f_{T(X)}(y_1,y_2)=f_X(y_1,y_2-y_1).$$

Die Dichte von Y ist durch die zweite Randverteilung der gemeinsamen Dichte gegeben und folgt durch Integration über die erste Komponente,

$$f_Y(y)=\int_{-\infty}^{\infty}f_X(y_1,y-y_1)dy_1$$

□

Speziell folgt aus Lemma 5.4, daß im Fall der stochastischen Unabhängigkeit von X_1 und X_2, also $f_{(X_1,X_2)}(x_1,x_2)=f_{X_1}(x_1)\cdot f_{X_2}(x_2)$, 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* (engl.: *convolution*) der Verteilungen von X_1 bzw. X_2.
Bezeichnung: $P^{X_1+X_2}=P^{X_1}*P^{X_2}$.

Beispiel 5.5.

a) X_1,X_2 seien stochastisch unabhängig, $X_1\sim\Gamma(\alpha,\lambda)$, $X_2\sim\Gamma(\beta,\lambda)$ mit $\alpha,\beta,\lambda>0$. Dann gilt

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

d.h., $\Gamma(\alpha,\lambda)*\Gamma(\beta,\lambda)=\Gamma(\alpha+\beta,\lambda)$. Die Γ-Verteilung ist also *faltungsstabil*.

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

$$\begin{aligned}f_{X_1+X_2}(y)&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-t-\mu_2)^2}{2\sigma_2^2}}dt\\&=\ldots=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}\end{aligned}$$

□

c) Sei $X \sim \mathrm{N}(0,1)$. Dann gilt $X^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$, da für $x > 0$ gilt,

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

Eine Dichte von X^2 ist also

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

Dies ist die 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 einer 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 Beispiel 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 Parametern n, λ). Zur Interpretation stelle man sich ein X_i als Bedienzeit für eine Anforderung an einen Server vor. Dann ist S_n die Gesamtbedienzeit für n Anforderungen.

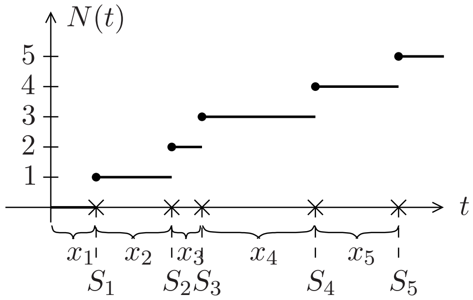
Umgekeht kann man auch fragen, mit welcher Wahrscheinlichkeit im Intervall $[0, t]$ genau n Anforderungen bedient werden, also wieviele Anfragen mit $\mathrm{Exp}(\lambda)$-verteilten Bedienzeiten der Server bis zur Zeit $t > 0$ (fest) bearbeitet.

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

$$N(t) = \max\left\{n \in \mathbb{N}_0 \mid \sum_{i=1}^n X_i \le t\right\} = \left|\left\{n \in \mathbb{N} \mid \sum_{i=1}^n X_i \le t\right\}\right|$$

heißt *Poisson-Prozess* mit Parameter $\lambda > 0$ (Bezeichnung: $\mathrm{PP}(\lambda)$).

Graphisch:



Interpretation: Anforderungen kommen mit $\mathrm{Exp}(\lambda)$-verteilten Zwischenankunftszeiten an. $N(t)$ zählt die Anzahl der Kunden, die bis zur Zeit t angekommen sind.

Die X_N heißen auch *Zwischenankunftszeiten* oder *Verweilzeiten* (*sojourn times*, *inter-arrival 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

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

$$= \frac{\lambda^k}{(k-1)!}\left(\int_0^t \underbrace{x^{k-1}}_{u'}\underbrace{e^{-\lambda x}}_{v}dx + \int_0^t \underbrace{\frac{x^k}{k}}_{u}\underbrace{\left(-\lambda e^{-\lambda x}\right)}_{v'}dx\right)$$

$$= \frac{\lambda^k}{(k-1)!}\underbrace{\frac{x^k}{k}}_{u}\underbrace{e^{-\lambda x}}_{v}\Bigg|_0^t = \frac{\lambda^k}{k!}t^k e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad \square$$

Es gilt sogar für $N_{(s,t]} = N(t) - N(s),\ 0 \leq s < t$, den *Zuwachs* im Intervall $(s,t]$:

- $N_{(s,t]} \sim \mathrm{Poi}(\lambda(t-s))$
- $N_{(s_i,t_i]},\ i \in \mathbb{N}$ 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 wird die Faltung 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.

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

Beispiel 5.9. Seien $X_1, \ldots, X_n$ stid $\sim \mathrm{Geo}(p)$, mit $0 < p < 1$. 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: $\overline{\mathrm{Bin}}(n,p)$). Es gilt also

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

Interpretation mit Beispiel 3.6: $X_1 + \ldots + X_n$ entspricht der Wartezeit bis zum n-ten Treffer (ohne die Treffer mitzuzählen).

Beweis. (mit Vollständiger Induktion)
$n = 1$:

$$P(X_1 = k) = \binom{k}{0}(1-p)^k p^1 = (1-p)^k \cdot p$$

$n \to n+1$:

$$\begin{aligned} P\left(\sum_{i=1}^{n+1} X_i = k\right) &= \sum_{j=0}^{k} P\left(\sum_{i=1}^{n} X_i = j\right) P\left(X_{n+1} = k-j\right) \\ &= \sum_{j=0}^{k} \binom{n+j-1}{n-1}(1-p)^j p^n (1-p)^{k-j} p \\ &= \sum_{j=0}^{k} \binom{n+j-1}{n-1}(1-p)^k p^{n+1} \\ &= (1-p)^k p^{n+1} \underbrace{\sum_{j=0}^{k} \binom{n+j-1}{n-1}}_{=\binom{n+k}{n}} \end{aligned}$$

□

Lemma 5.10. Es gilt

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

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

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

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

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 $f_{X_i}(x) = 0,\ i = 1, 2$, 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

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

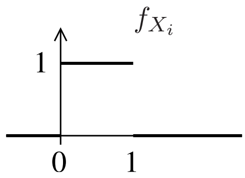
Beweis. Nutzt den Transformationssatz für Dichten mit $T(x,y) = (x, x\cdot y)$ bzw. $T(x,y) = \left(x, \frac{x}{y}\right)$. $\square$

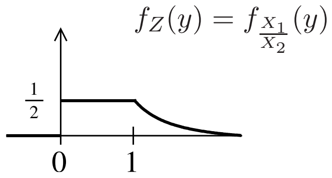
Beispiel 5.12. X_1, X_2 s.u. $\sim \mathbb{R}(0,1)$, d.h. $f_{X_i}(x) = \mathbb{I}_{(0,1)}(x)$. $Z = \frac{X_1}{X_2}$ besitzt die Dichte

$$f_Z(y) = \int_0^\infty t\, \mathbb{I}_{(0,1)}(yt)\, \mathbb{I}_{(0,1)}(t) dt \qquad , y \geq 0,$$

aus $0 < y \cdot t < 1$ und $0 < t < 1 \quad \Rightarrow \quad 0 < t < \min\left\{\frac{1}{y}, 1\right\}$, somit ist

$$\begin{aligned} f_Z(y) &= \int_0^{\min\{\frac{1}{y},1\}} t\, dt = \frac{t^2}{2}\Big|_0^{\min\{\frac{1}{y},1\}} \\ &= \begin{cases} \frac{1}{2} & , 0 \leq y \leq 1 \\ \frac{1}{2y^2} & , y \geq 1 \end{cases} \end{aligned}$$





Beispiel 5.13. X_1, X_2 s.u., $X_1 \sim \Gamma(\alpha, \lambda)$, $X_2 \sim \Gamma(\beta, \lambda)$, $\alpha, \beta, \lambda > 0$

a) $Y = \frac{X_1}{X_2}$ besitzt eine Dichte

$$f_Y(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}} \mathbb{I}_{(0,\infty)}(x)$$

b) $Z = \frac{X_1}{X_1+X_2}$ besitzt die Dichte

$$f_Z(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x)$$

Z heißt Beta-verteilt mit Parametern $\alpha, \beta > 0$. Bezeichung: $Z \sim \text{Beta}(\alpha, \beta)$. Beachte: $\alpha = \beta = 1$, dann ist

$$\begin{aligned} f_Z(x) &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^0 (1-x)^0 \mathbb{I}_{(0,1)}(x) \\ &= \mathbb{I}_{(0,1)}(x). \end{aligned}$$

Dieses ist Dichte einer $\text{R}(0,1)$-Verteilung, also $\text{Beta}(1,1) = \text{R}(0,1)$.

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 dafür ist eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit der Verteilung $P(X = i) = 1/6$, $i = 1, \ldots, 6$. Betrachte nun die folgenden Problemstellungen:

   - Angenommen, es werde bei jedem Würfelwurf die Anzahl der geworfenen Augen in EURO ausgezahlt. Dann ist die „mittlere" oder zu „erwartende" Auszahlung ganz intuitiv gegeben durch
   $$E(X) = \frac{1}{6} \cdot 1 + \ldots \frac{1}{6} \cdot 6 = 3,5.$$
   Die „mittlere" Auszahlung beträgt also 3,5 EURO
   - Angenommen, es werde bei jedem Würfelwurf das Quadrat der geworfenen Augen in EURO, $g(X) = X^2$, ausgezahlt. Dann ist die „zu erwartende" Auszahlung
   $$E(g(X)) = \frac{1}{6} \cdot 1^2 + \ldots + \frac{1}{6} 6^2 = \frac{1}{6} \cdot 91 = 15,1\overline{6}.$$

b) Unter dem Namen „*Petersburger Paradoxon*" ist folgende Überlegung von *Nikolaus Bernoulli* (1695-1726) bekannt. Die Frage ist, ob es für ein vereinfachtes Roulettespiel, bei dem die Farben „Rot" und „Schwarz" jeweils mit der Wahrscheinlichkeit 1/2 auftreten, eine Gewinnstrategie gibt. Dabei wird solange gespielt, bis das erste Mal gewonnen wird. Solange man verliert, verdoppelt man den Einsatz in jedem Spiel. Im ersten Spiel wird hierbei 1 EUR gesetzt.

   Der Einsatz im n-ten Spiel beträgt also 2^{n-1} EUR, $n = 1, 2, 3, \ldots$. Fällt beim n-ten Spiel erstmalig rot, so beläuft sich der Gesamteinsatz bis dahin auf $1 + 2 + 4 + \ldots + 2^{n-1} = 2^n - 1$ EUR. Im Gewinnfall werden 2^n EURO ausgezahlt, womit als Gewinn genau 1 EUR übrigbleibt.

Mathematisches Modell:
$\{X_n\}_{n\in\mathbb{N}_0}$ sei eine Folge von stid. Zufallsvariablen, $X_n \sim \text{Bin}(1, 1/2)$,
$\{X_n = 1\}$ bedeutet, daß im n-ten Spiel "rot" auftritt.
$S = \min\{n \in \mathbb{N} \mid X_n = 1\}$: Zeitpunkt des ersten Auftretens von "rot".
Es gilt $S - 1 \sim \text{Geo}\left(\frac{1}{2}\right)$ mit $P(S = k) = (1-p)^{k-1}p$, $k \in \mathbb{N}$, p=1/2.

Die Auszahlung beträgt $A = 1$, falls $S < \infty$. Für die zugehörige Wahrscheinlichkeit gilt $P(A = 1) = P(S < \infty) = \sum_{k=0}^{\infty} P(S = k) = 1$.
Die erwartete Auszahlung beträgt hiermit

$$E(A) = 1 \cdot P(A = 1) = 1.$$

Es liegt also eine sichere Gewinnstrategie vor, die jedoch unendlich viel Kapital und unendlich viel Zeit erfordert.

Interessant ist eher die erwartete Auszahlung bei begrenztem Kapital. Angenommen das maximale Kapital beträgt $2^L - 1$ EUR. Das bedeutet, daß höchstens L Spiele spielbar sind. Also ist die Auszahlung $A = 1$ genau dann, wenn $S \leq L$ und $A = -(2^L - 1)$ genau dann, wenn $S \geq L + 1$. Für die Verteilung der Auszahlung gilt

$$P(A = 1) = P(S \leq L) = \sum_{k=1}^{L} \frac{1}{2}\left(\frac{1}{2}\right)^{k-1} = \sum_{k=1}^{L} \frac{1}{2^k} = 1 - \frac{1}{2^L}$$

und

$$P\left(A = -\left(2^L - 1\right)\right) = P(S \geq L + 1) = 1 - P(A = 1) = \frac{1}{2^L}.$$

Für die erwartete Auszahlung ergibt sich also

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

Unter der realistischen Annahme, daß das verfügbare Kapital beschränkt ist, ist der mittlere Gewinn bei diesem Spiel null. Man bedenke allerdings, welch riesigem möglichen Verlust der geringe Gewinn von 1 EUR gegenübersteht.

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 der Definition des Erwartungswerts für den Fall unendlich vieler Trägerpunkte oder bei absolut-stetigen Zufallsvariablen, daß die zugehörige Reihe bzw. das Integral existieren müssen. Dies wird in den Voraussetzungen der folgenden Definition gefordert.

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 folgt für die Identität $g(x) = x$:

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) \text{ bei diskreten Zufallsvariablen und}$$
$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \text{ bei absolut-stetigen Zufallsvariablen}$$

Für beliebige, nicht notwendig diskrete oder absolut-stetige 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. Für den Fall differenzierbarer Verteilungsfunktionen F ($F'(x) = f(x)$ ist dann eine Dichte). Es gilt sowohl

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

als auch

$$\begin{aligned} \int_{0}^{\infty} (1 - F(x)) \cdot 1\, dx &= (1 - F(x)) \cdot x \Big|_{0}^{\infty} - \int_{0}^{\infty} -f(x) \cdot x\, dx \\ &= \int_{0}^{\infty} f(x) \cdot x\, dx. \end{aligned}$$

Insgesamt folgt

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

Beispiel 6.3. Erwartungswerte von speziellen Zufallsvariablen.

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

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

$\Rightarrow \quad E(X) = \frac{1-p}{p}, \quad 0 < p \leq 1.$

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

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot \lambda e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}, \end{aligned}$$

$\Rightarrow \quad E(X) = \frac{1}{\lambda}, \quad \lambda > 0.$

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

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

$\Rightarrow \quad E(X) = \mu$

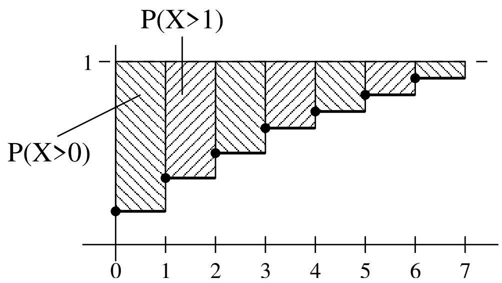
Der Parameter μ der $\mathrm{N}(\mu, \sigma^2)$-Verteilung ist also gerade der zugehörige Erwartungswert.

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

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



$P(X>0)=1-P(X=0)$

□

Erwartungswerte von Funktionen von Zufallsvektoren werden wie folgt berechnet.

Satz 6.5. 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 $f_{(X_1,\ldots,X_n)}$. Falls $\sum_{i=1}^{\infty}|g(t_i)|\,f(t_i)<\infty$, so gilt

$$\begin{aligned} E\big(g(X_1,\ldots,X_n)\big) &= \sum_{i=1}^{\infty} g(t_i)f(t_i) \\ &= \sum_{i=1}^{\infty} g(t_i)P\big((X_1,\ldots,X_n)=t_i\big) \end{aligned}$$

b) Sei $(X_1,\ldots,X_n)$ absolut-stetig mit Dichte f.
Falls $\int\ldots\int|g(x_1\ldots,x_n)|\,f(x_1,\ldots,x_n)\,dx_1\ldots dx_n<\infty$, so gilt

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

Beispiele für Transformationen g, die im folgenden speziell behandelt werden, sind:

$$\begin{array}{lcl} (X,Y) & \longmapsto & X+Y \\ (X_1,X_2) & \longmapsto & aX_1+bX_2\ , \quad a,b\ \in\ \mathbb{R} \\ (X,Y) & \longmapsto & X\cdot Y \end{array}$$

Satz 6.6 (Eigenschaften des Erwartungswertes). X,Y seien Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega,\mathfrak{A},P)$, deren Erwartungswert existiert. Dann gilt:

a) $E(aX+bY)=aE(X)+bE(Y)\ \forall a,b\in\mathbb{R}$ (*Linearität*).

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

c) Für $X=\mathbb{I}_A,\ A\in\mathfrak{A}$, gilt $E(X)=E(\mathbb{I}_A)=P(A)=:\int\mathbb{I}_A dP$

d) $P(|X|>c)\le\frac{E(|X|)}{c}\qquad\forall c>0$ (*Markoff-Ungleichung*)

e) X,Y stochastisch unabhängig $\Rightarrow\ E(X\cdot Y)=E(X)\cdot E(Y)$

Beweis.

a) Wir führen den Beweis nur für absolut-stetige Zufallsvariable.

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

b) Die Aussage folgt aus den Monotonieeigenschaften der Summation und Integration.

c) Die Aussage folgt sofort aus der Definition der Indikatorfunktion,

$$\mathbb{I}_A(w) = \begin{cases} l, & \text{falls } w \in A \\ 0, & \text{falls } w \notin A \end{cases}.$$

$\mathbb{I}_A$ ist also eine diskrete Zufallsvariable mit den Werten 0 und 1, ihr Erwartungswert beträgt $E(\mathbb{I}_A) = 1 \cdot P(A) + 0 \cdot P\left(A^{\complement}\right) = P(A)$.

d) Für alle $c > 0$ gilt

$$c \cdot \mathbb{I}_{\{|X|>c\}} = \begin{cases} c, & \text{falls } |X| > c \\ 0, & \text{falls } |X| \leq c \end{cases}.$$

Also ist $E\left(c \cdot \mathbb{I}_{\{|X|>c\}}\right) = c \cdot P(|X| > c) \leq E(|X|)$, woraus $P(|X| > c) \leq E(|X|)/c$ folgt.

e) Für diskrete, stochastisch unabhängige Zufallsvariablen X, Y gilt

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

Für absolut-stetige Zuvallsvariablen X, Y gilt analog, daß

$$\begin{aligned} E(X \cdot Y) &= \iint x \cdot y \cdot f_{(X,Y)}(x,y)\, dy\, dx \\ &= \iint x \cdot y \cdot f_X(x) f_Y(y)\, dy\, dx \\ &= \int x f_X(x)\, dx \int y f_Y(y)\, dy \;=\; E(X) \cdot E(Y). \end{aligned}$$

□

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 6.6 e) folgt aus der stochastischen Unabhängigkeit der Zufallsvariablen X und Y deren Unkorreliertheit. Die Umkehrung hiervon ist im allgemeinen falsch.

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* (engl.: k^{th} *moment*) von X.

   Speziell ergibt sich im Fall $k = 1$ der Erwartungswert von X.

b) $E\left((X - EX)^k\right)$ heißt k-tes *zentrales Moment* (*central moment*) von X.

   Speziell für $k = 2$: $\operatorname{Var}(X) = E\left((X - EX)^2\right)$ heißt *Varianz* (engl.: *variance*) von X. Bezeichnung auch $\mathrm{V}(X)$.

   $\sqrt{\operatorname{Var}(X)}$ heißt *Standardabweichung* (engl.: *standard deviation*) von X.

c) $\operatorname{Cov}(X, Y) = E((X - EX)(Y - EY))$ heißt *Kovarianz* (engl.: *covariance*) von X und Y. $\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}$ heißt *Korrelation* (engl.: *correlation*) von X und Y.

Lemma 6.9. $X, Y, X_1, \ldots, X_n$ seien Zufallsvariablen, alle auftretende Momente sollen existieren. Es gelten die folgenden Aussagen.

a) $\operatorname{Cov}(X, Y) = E(X \cdot Y) - E(X)E(Y)$, insbesondere gilt

   $\operatorname{Cov}(X, X) = \operatorname{Var}(X) = E\left(X^2\right) - (E(X))^2$.

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

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

   Insbesondere gilt $\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$, falls $X_1, \ldots, X_n$ paarweise unkorreliert sind.

d) $|\operatorname{Cov}(X, Y)| \leq \sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)} \qquad$ (*Cauchy-Schwarz-Ungleichung*)

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

Beweis. Zur Vereinfachung der Notation wird bei dem Erwartungswert $\mathrm{E}(X)$ gelegentlich auf die Klammersetzung verzichtet.

a)

$$
\begin{aligned}
\operatorname{Cov}(X,X) &= E\big((X-EX)(X-EX)\big) = E\big((X-EX)^2\big)\\
&= \operatorname{Var}(X)\\
\operatorname{Cov}(X,Y) &= E\big((X-EX)(Y-EY)\big)\\
&= E\big(XY-(EX)Y-X(EY)+(EX)(EY)\big)\\
&= E(XY)-E\big((EX)Y\big)-E\big(X(EY)\big)+E\big((EX)(EY)\big)\\
&= E(XY)-(EX)(EY)-(EX)(EY)+(EX)(EY)\\
&= E(XY)-(EX)(EY)
\end{aligned}
$$

Hieraus folgt, daß $\operatorname{Var}(X) = E\left(X^2\right) - (E(X))^2$.

b)

$$
\begin{aligned}
\operatorname{Var}(aX+b) &= E\left(\big(aX+b-E(aX+b)\big)^2\right)\\
&= E\Big(\big(aX+b-aEX-\underbrace{E(b)}_{b}\big)^2\Big)\\
&= E\left(\big(a(X-EX)\big)^2\right) = E\left(a^2(X-EX)^2\right)\\
&= a^2\operatorname{Var}(X)
\end{aligned}
$$

c)

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= E\left(\left(\sum_{i=1}^n X_i\right)^2\right) - \left(E\left(\sum_{i=1}^n X_i\right)\right)^2\\
&= E\left(\sum_{i,j=1}^n X_iX_j\right) - \left(\sum_{i=1}^n E(X_i)\right)^2\\
&= \sum_{i,j=1}^n E(X_iX_j) - \sum_{i,j=1}^n E\left(X_i\right)E\left(X_j\right)\\
&= \sum_{i=1}^n \left(E\left(X_i^2\right) - \big(E(X_i)\big)^2\right) + \sum_{i\neq j}\left(E(X_iX_j)-E(X_i)E(X_j)\right)\\
&= \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i\neq j}\operatorname{Cov}(X_i,X_j)\\
&= \sum_{i=1}^n \operatorname{Var}(X_i) + 2\cdot\sum_{i<j}\operatorname{Cov}(X_i,X_j)
\end{aligned}
$$

d) Der Beweis wird nur für den Fall der Existenz von Dichten geführt. Er benutzt in der zweiten Zeile die aus der Analysis bekannte Cauchy-Schwarz-Ungleichung.

$$\begin{aligned}
|E(XY)| &= \left|\iint xyf(x,y)\,dx\,dy\right| \\
&\leq \iint |xy| f(x,y)\,dx\,dy = \iint |x||y| f(x,y)\,dx\,dy \\
&\leq \left(\iint x^2 f(x,y)\,dx\,dy\right)^{\frac{1}{2}} \left(\iint y^2 f(x,y)\,dx\,dy\right)^{\frac{1}{2}} \\
&= \left[\int x^2 \Bigg(\underbrace{\int f(x,y)\,dy}_{f_X(x)}\Bigg) dx\right]^{\frac{1}{2}} \left[\int y^2 \Bigg(\underbrace{\int f(x,y)\,dx}_{f_Y(y)}\Bigg) dy\right]^{\frac{1}{2}} \\
&= \left(\int x^2 f_X(x)\,dx\right)^{\frac{1}{2}} \cdot \left(\int y^2 f_Y(y)\,dy\right)^{\frac{1}{2}} \\
&= \left(E\left(X^2\right)\right)^{\frac{1}{2}} \left(E\left(Y^2\right)\right)^{\frac{1}{2}}
\end{aligned}$$

Es folgt also $|E(XY)| \leq \sqrt{E\left(X^2\right) E\left(Y^2\right)}$. Die Behauptung erhält man durch Ersetzen von X durch $X - EX$ und Y durch $Y - EY$ wie folgt.

$$|E\big((X-EX)(Y-EY)\big)| \leq \sqrt{E\big((X-EX)^2\big)E\big((Y-EY)^2\big)},$$

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

□

Die in Definition 6.8 eingeführten Größen können wie folgt interpretiert werden.

- $E(X)$ ist der erwartete oder mittlere Wert einer Zufallsvariablen.
- $\operatorname{Var}(X)$ ist ein Streuungsmaß, es gibt die mittlere quadratische Abweichung vom Erwartungswert an.
- $\operatorname{Cov}(X,Y)$ kann als Korrekturterm bei der Varianzberechnung von Summen interpretiert werden. Es gilt nämlich $\operatorname{Var}(X+Y) = \operatorname{Var}(X)+\operatorname{Var}(Y)+2\operatorname{Cov}(X,Y)$.
- $\operatorname{Corr}(X,Y)$ ist eine Maßzahl für den linearen Zusammenhang von Zufallsvariablen. Dies erkennt man aus der folgenden Äquivalenz.

  $$\exists\, a,b \in \mathbb{R} \;:\; P(X = aY + b) = 1 \Leftrightarrow |\operatorname{Corr}(X,Y)| = 1$$

  Man beachte hierbei, daß wegen der Cauchy-Schwarz-Ungleichung allgemein $|\operatorname{Corr}(X,Y)| \leq 1$ gilt, bei linearem Zusammenhang wird die Schranke also angenommen.

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

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

Beweis. Nach Definition 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 die erzeugende Funktion oder Transformierte von $X_1 + X_2$ wegen Satz 6.6 e)

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

□

Beispiel 6.11.

a) $X \sim \text{Bin}(n, p)$, dann ist $G_X(z) = (1 - p + pz)^n, \ |z| \leq 1$.
Seien $X_1 \sim \text{Bin}(n_1, p), \ X_2 \sim \text{Bin}(n_2, p)$ stochastisch unabhängig, dann ist

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

Dies ist die erzeugende Funktion einer Binomialverteilung mit Parametern $n_1 + n_2$ und p. Also gilt $X_1 + X_2 \ \sim \ \text{Bin}(n_1 + n_2, p)$.

b) $X \sim \text{Exp}(\lambda)$, dann ist $L_X(s) = \frac{\lambda}{s+\lambda}, \ s \geq 0$.
Seien $X_1 \sim \text{Exp}(\lambda), \ X_2 \sim \text{Exp}(\mu)$ stochastisch unabhängig, $\lambda \neq \mu$, dann ist

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

Wie man leicht nachrechnet, hat die Dichte

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

gerade die obige Laplace-Transformierte. Wegen des Eindeutigkeitssatzes für Laplace-Transformierte hat $X_1 + X_2$ also die oben angegebene Dichte f.

Satz 6.12 (Transformierte und Momente).

a) $G(z) = \sum_{k=0}^{\infty} P(X = k)z^k \ |z| \leq 1$, sei erzeugende Funktion einer diskreten Zufallsvariablen X mit Träger $\mathbb{N}_0$. Dann gilt

$$E(X) = G'(1)$$
$$\underbrace{E(X \cdot (X-1) \cdots (X-k+1))}_{k\text{-tes } \textit{faktorielles Moment}} = G^{(k)}(1)$$

Also: $E\left(X^2\right) = G'(1) + G''(1)$

Beachte: $G^{(i)}(1) = \lim_{z \uparrow 1} G^{(i)}(z)$, falls $G(z)$ nicht existiert für $z > 1$.

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

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

Beweis. Der Beweis wird nur für den Teil a) geführt. Bezeichne hierzu $p_k = P(X = k)$. Es gilt

$$G'(z) = \sum_{k=0}^{\infty} k\, z^{k-1}\, p_k, \quad |z| < 1. \quad \text{(durch gliedweises Differenzieren im Inneren)}$$

$$E(X) \text{ existiert } \Rightarrow \sum_{k=0}^{\infty} p_k\, k\, z^{k-1} \text{ konvergent für } z = 1.$$

Mit dem Abelschen Grenzwertsatz folgt, daß $G'(z)$ linksseitig stetig in 1 ist. Also gilt

$$G'(1) = \lim_{z \to 1+} G'(z) = \sum_{k=0}^{\infty} k\, p_k = E(X).$$

Analog gilt für höhere Ableitungen

$$G''(z) = \sum_{k=0}^{\infty} p_k\, k\,(k-1)\, z^{k-2}$$
$$\Rightarrow \quad G''(1) = \sum_{k=0}^{\infty} p_k\, k\,(k-1) = E\big(X(X-1)\big)$$

Speziell: $E\big(X(X-1)\big) = E(X^2) - E(X) \ \Rightarrow \ E(X^2) = G''(1) + G'(1)$ □

Beispiel 6.13 (Berechnung von Momenten).

a) $X \sim \text{Bin}(n,p),\ 0 \leq p \leq 1$. Die Berechung des Erwartungswertes kann alternativ mit den folgenden drei Methoden durchgeführt werden.

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

(ii) $X \sim \sum_{i=1}^{n} X_i$ mit X_i stid $\sim \text{Bin}(1,p),\ E(X_i) = p$,

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

(iii) Über die erzeugende Funktion

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

Die Varianz wird mit der Methode (ii) bestimmt. Zunächst berechnen wir

$$E\left(X_i^2\right) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

und hieraus

$$\text{Var}(X_i) = E(X_i^2) - (EX_i)^2 = p - p^2 = p(1-p).$$

Insgesamt folgt

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

b) $X \sim \text{Exp}(\lambda),\ \lambda > 0$. Dann gilt

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

Also ist

$$E\left(X^k\right) = k!\lambda^{-k},\ k \in \mathbb{N}_0, \quad E(X) = \frac{1}{\lambda}, \quad E\left(X^2\right) = \frac{2}{\lambda^2}$$
$$\text{Var}(X) = E\left(X^2\right) - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

c) $X \sim \mathbb{N}(0,1)$, also normalverteilt mit Parametern $\mu = 0$, $\sigma^2 = 1$. Dann ist $E(X) = 0$ und

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

Es folgt: $\operatorname{Var}(X) = E\left(X^2\right) - (EX)^2 = 1 - 0 = 1$.

$Y = \sigma X + \mu$ besitzt die Dichte

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

d.h. $Y \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Also gilt

$$E(Y) = E(\sigma X + \mu) = \sigma E(X) + \mu = \mu \quad \text{und}$$
$$\operatorname{Var}(Y) = \operatorname{Var}(\sigma X + \mu) = \sigma^2 \operatorname{Var}(X) = \sigma^2$$

Das bedeutet, daß der erste Parameter einer Normalverteilung den Erwartungswert und der zweite die Varianz angibt.

Beispiel 6.14 (Average Case Analyse von Quicksort).

QS: Gegeben sind Elemente $1, \ldots, n$,
Wähle zufällig ein Trennelement d.
Vergleiche alle anderen Elemente mit d und bilde linkes Teilfeld L von kleineren Elementen und rechtes Teilfeld R von größeren Elementen.
$\longrightarrow \quad (L, d, R)$
Wende QS rekursiv auf L und R an.

Komplexitätsmaß: Anzahl der Vergleiche bei QS.
Seien $s_1, \ldots, s_n,\ s_1 < \ldots < s_n$ das Ergebnis von QS.

$$\text{Zufallsvariablen} \quad X_{ij} = \begin{cases} 1, & s_i \text{ und } s_j \text{ werden bei QS verglichen} \\ 0, & \text{sonst} \end{cases}$$

Gesamtaufwand: $T = \sum_{i<j} X_{ij}$

s_i und s_j werden verglichen, wenn s_i oder s_j von den Elementen $s_i, s_{i+1}, \ldots, s_{j-1}, s_j$ als Trennelement ausgewählt wurde. Also gilt bei rein zufälliger Wahl

$$P\left(X_{ij} = 1\right) = \frac{2}{j - i + 1}.$$

Insgesamt ergibt sich für den Aufwand

$$\begin{aligned} E(T) = E\left(\sum_{i<j} X_{ij}\right) &= \sum_{i<j} \frac{2}{j-i+1} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n} \sum_{j=2}^{n-i+1} \frac{2}{j} \\ &\leq 2 \sum_{i=1}^{n} \underbrace{\sum_{j=2}^{n} \frac{1}{j}}_{H(n)} = 2\, n\, H(n) = 2\, n\, \ln(n) + O(n), \end{aligned}$$

mit $H(n) = \ln(n) + \gamma + \frac{1}{2n} + O\left(\frac{1}{n^2}\right)$.

$$\gamma = \lim_{n \to \infty} \left(H(n) - \ln n\right) = 0.5772...$$

bezeichnet hierbei die *Eulersche Konstante*.

Kapitel 7

Bedingte Verteilungen und Erwartungswerte

7.1 Diskreter Fall

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

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

Die *bedingte Zähldichte* von X unter $Y = y$ wird durch

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

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

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

Der Satz von der totalen Wahrscheinlichkeit (Satz 2.15 a)) 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. X und N seien Zufallsvariable mit $f_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, \ldots, n$, also $P^{X|N=n} = \text{Bin}(n,p)$ und $N \sim Poi(\lambda)$, $\lambda > 0$.
Zur Interpretation stelle man sich bespielsweise ein zweistufiges Experiment vor, in dem zunächst die Anzahl der Münzwürfe gemäß einer Poissonverteilung ermittelt wird und

anschließend entsprechend oft eine Münze geworfen wird, wobei die Anzahl des Auftretens von „Kopf“ gezählt wird. Für die Wahrscheinlichkeit, daß genau k-mal „Kopf“ auftritt, gilt

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

also gilt $X \sim Poi(\lambda p)$.

Dieses Ergebnis hat eine weitere interessante Interpretation. Wir wissen, daß beim Poissonprozeß die Anzahl der in einem festen Intervall auftretenden Ereignisse poissonverteilt ist. Wird nun in weiteren unabhängigen Münzwürfen mit Trefferwahrscheinlichkeit p entschieden, ob jedes Ereignis mitgezählt werden soll oder nicht, so liegt genau die oben beschriebene Situation vor. Es folgt, daß die Anzahl der verbleibenden Ereignisse wieder poissonverteilt ist mit einem um den Faktor p reduzierten Parameter. Hieraus läßt sich herleiten, daß dieses sogenannte *Ausdünnen* (engl. thinning) von Poissonprozessen wieder zu einem Poissonprozeß führt.

Der Erwartungswert bzgl. der bedingten Verteilung wird wie folgt berechnet.

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

heißt *bedingter Erwartungswert* von $g(X)$ unter $Y=y$ (sofern existent).

Der Erwartungswerts läßt sich mit Hilfe der bedingten Erwartungswerte wie folgt berechnen.

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

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

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

Beachte: $f_{X|Y}(x|y)$ ist eine Dichte in der Variablen x für alle festen $y \in \mathbb{R}$. Die zugehörige Verteilung heißt *bedingte Verteilung* von X unter $Y = y$. Insbesondere gilt für alle $y \in \mathbb{R}$ und alle Ereignisse A, daß

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

Speziell heißt

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

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

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

Ebenso definiert

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

den *bedingten Erwartungswert* von $g(X)$ unter $Y = y$.
Analog zum diskreten Fall läßt sich der Erwartungswert von X durch Integration über die bedingten Erwartungswerte wie folgt berechnen.

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

7.3 Gemischte Fälle

Gemischte Fälle, bei denen eine Zufallsvariable absolut-stetig und eine andere diskret ist, können analog mit Hilfe von elementaren bedingten Wahrscheinlichkeiten, dem Satz von der totalen Wahrscheinlichkeit und Integration über bedingte Dichten behandelt werden. Die folgenden Beispiele zeigen einige typische Vorgehensweisen.

Beispiel 7.2 (Wartezeit in einem Wartesystem). Sei N die zufällige Anzahl von Kunden in einer Warteschlange, es gelte $N \sim \text{Geo}(p)$.
$X_n,\ n \in \mathbb{N}$, sei die Bedienzeit von Kunde n, X_n seien stid $\sim \text{Exp}(\lambda)$.

Gesucht ist die Verteilung der Gesamtwartezeit eines neu ankommenden Kunden, d.h.

$$W = \sum_{i=1}^{N+1} X_i.$$

Bekannt ist, daß $P^{W|N=n} = \text{Erl}(n+1, \lambda)$. 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 $W \sim \text{Exp}(\lambda p)$ und für den Erwartungswert gilt $E(W) = \frac{1}{\lambda p}$.

Beispiel 7.3 (Ankünfte eines Poisson-Prozesses in einem zufälligen Intervall). Sei $N(t),\ t \ge 0$, ein Poisson-Prozess mit Parameter $\lambda > 0$. Dann gilt bekanntlich $N(t) \sim \text{Poi}(\lambda t)$.

Sei ferner Y eine exponentialverteilte Zufallsvariable mit Parameter $\mu > 0$, also $Y \sim \text{Exp}(\mu)$. Gesucht ist die Verteilung der Anzahl von Ankünften in dem zufälligen Intervall $[0, Y]$, also die Verteilung der Zufallsvariablen

$$N(Y) = \max\{n \in \mathbb{N}_0 \mid \sum_{i=1}^{n} X_i \le Y\},$$

wobei die X_i stid $\sim \text{Exp}(\lambda)$, auch stochastisch unabhängig von Y sind.

Bekannt ist nun, dass $P^{N(Y)|Y=t} = \text{Poi}(\lambda t)$ für alle $t > 0$. Hiermit folgt für alle $k \in \mathbb{N}_0$

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

Dies ist die Zähldichte einer geometrischen Verteilung mit Parameter $\mu/(\lambda+\mu)$, so daß $N(Y) \sim \text{Geo}\left(\frac{\mu}{\lambda+\mu}\right)$ gilt.

7.4 Der allgemeine Fall

Die oben beschriebenen Fälle lassen sich allgemein wie folgt beschreiben. Für alle Ereignisse A und B gilt

$$
P(X \in A, Y \in B) = \int_B P^{X|Y=y}(A) dP^Y(y). \tag{7.1}
$$

Eine Funktion $P^{X|Y=y}(A)$, die Wahrscheinlichkeitsverteilung in A für jedes y und eine Zufallsvariable in y für jedes A ist, heißt bedingte Verteilung von X unter $Y = y$, wenn (7.1) gilt.

Kapitel 8

Grenzwertsätze

Zur Motivation drei Problemstellungen:

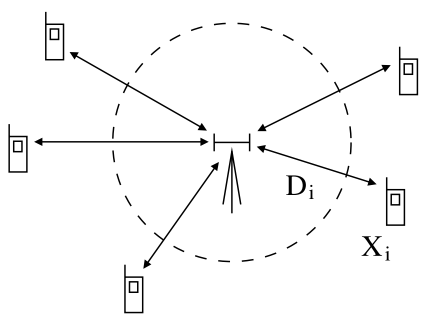
a) Man betrachte den unendlichen Münzwurf, wobei die Wahrscheinlichkeit p für das Auftreten von „Kopf“ unbekannt ist. Das unbekannte p soll in einem Experiment geschätzt werden. Es liegt nahe, hierzu die Münze n mal zu werfen, die Anzahl k des Aufretens von Kopf zu zählen und den Quotienten k/n als Schätzer für p zu verwenden.

   Unser Modell für diese Situation ist wie folgt. $X_i \sim \mathrm{Bin}(1, p)$ seien stochastisch unabhängige, binomialverteilte Zufallsvariablen. Der vorgeschlagene Schätzer läßt sich beschreiben als $\frac{1}{n}\sum_{i=1}^{N} X_i$. Die entscheidende Frage ist, ob die relative Häufigkeit des Auftretens von Kopf gegen die wahre Wahrscheinlichkeit p konvergiert. Gilt also

$$\frac{\text{Anzahl der Würfe mit Kopf}}{\text{Anzahl aller Würfe}} = \frac{\sum_{i=1}^{n} X_i}{n} \overset{?}{\longrightarrow} p \ (n \longrightarrow \infty) \quad ?$$

   Das *Gesetz großer Zahlen* wird hierauf eine positive Antwort geben.

b) Vereinfachtes UMTS-Modell (CDMA-Netz, CDMA = Code Division Multiple Access).



   Die Interferenzleistung von Mobilstationen „außerhalb“ der eigenen Zelle limitiert die Kapazität. Die gesamte Interferenz von n Mobilstationen wird beschrieben durch

$$S = \sum_{i=1}^{n} X_i \cdot D_i^{-\gamma},$$

wobei $\gamma \in [2, 5]$ ein Parameter ist, der an die speziell betrachtete Umgebung angepaßt wird. Hierbei bedeutet X_i die zufällige Sendeleistung der Station i, die von der gewünschten Datenrate und Übertragungsqualität abhängt, und D_i den zufälligen Abstand der Mobilstation i von der betrachteten Basisstation.

Wenn X_i und D_i stochastisch unabhängig mit bekannten Vertei,ungen angenommen werden, so läßt sich grundsätzlich $E(S)$ und $\mathrm{Var}(S)$ bestimmen. Die Frage nach der asymptotischen Verteilung von S für große n läßt sich dann wie folgt beantworten.

$$\frac{S - E(S)}{\sqrt{Var(S)}} \overset{as}{\sim} N(0, 1)$$

und zwar unabhängig von der Verteilung von X_i und D_i. Dies wird aus dem *Zentralen Grenzwertsatz* folgen.

c) In einem Wartesystem sei $X_i \sim \mathrm{Poi}(\lambda)$ die Anzahl der ankommenden Kunden im i-ten Zeitintervall (jeweils der Länge 1). Die Anzahl der bis zur Zeit n angekommenen Kunden ist somit $Y_n = \sum_{i=1}^{n} X_i \sim \mathrm{Poi}(n\lambda)$.
Für große n ist die numerische Berechnung der Verteilungsfunktion sehr aufwendig. Man interessiert sich deshalb für die Verteilung, welche sich asymptotisch für große n ergibt. Der *Zentrale Grenzwertsatz* besagt, daß mit den Konstanten $a_n = \mathrm{E}(Y_n)$ und $b_n = \sqrt{\mathrm{Var}(Y_n)}$ folgt

$$\frac{Y_n - a_n}{b_n} \overset{as}{\sim} N(0, 1).$$

Dies gilt für beliebige Verteilung der X_i, insbesondere auch für diskrete X_i, wie z.B. $X_i \sim \mathrm{Bin}(k, p)$.

Zunächst müssen geeignete Konvergenzbegriffe definiert werden:

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

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

$$P\left(\left\{\omega \,\middle|\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) \overset{kurz}{=} P\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

Bezeichnung:

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

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

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

Bezeichnung:

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

c) *schwach konvergent* oder *verteilungskonvergent* (engl.:*convergent in distribution*) gegen X, wenn

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

in allen Stetigkeitspunkten x von F, wobei $F_n(x)$ Verteilungsfunktion von X_n und $F(x)$ Verteilungsfunktion von X bezeichnet.

Bezeichnung:

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

Lemma 8.2. Es gilt

$$X_n \underset{n\to\infty}{\longrightarrow} X \text{ P-f.s.} \quad \Longleftrightarrow \quad P\big(\limsup_{n\to\infty}\{|X_n - X| > \varepsilon\}\big) = 0 \quad \forall\, \varepsilon > 0$$

Also gilt:

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

Beweis.

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

Es folgt:

$$\underbrace{P\left(\bigcup_k \left\{\omega \mid |X_n(\omega) - X(\omega)| > \frac{1}{k} \;\text{ für } \infty \text{ viele } n\right\}\right)}_{=\left\{\omega \,\mid\, \exists\, k\, \forall\, n_0\, \exists\, n \geq n_0 \,:\, |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}} \leq \sum_k P\big(\{\cdots\}\big) = 0$$

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

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

„$\Rightarrow$“ : klar

$\square$

Satz 8.3 (Zusammenhänge zwischen den Konvergenzarten).

a) $X_n \xrightarrow[n\to\infty]{} X$ P-f.s. $\implies$ $X_n \xrightarrow[n\to\infty]{} X$ P-stoch $\implies$ $X_n \xrightarrow{\mathcal{D}} X$ $(n \to \infty)$.
Die Umkehrungen sind im Allgemeinen falsch.

b) „Schnelle" stochastische Konvergenz impliziert fast sichere Konvergenz:

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

Beweis. a) Wir führen den Beweis nur für die erste Implikation. Der Beweis der zweiten wird in Mathar, Pfeiffer, Lemma 2.3.2, Seite 137 gegeben.

$\forall\, \varepsilon > 0$ gilt:

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

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

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

Die Behauptung folgt nun durch Anwendung von Lemma 8.2.

□

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

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

Beweis. Mit der Markoff-Ungleichung (Satz 6.6) folgt

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

□

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

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

Beweis. O.B.d.A. sei $EX_i = 0$ für alle i, die X_i seien paarweise unkorreliert.
Setze $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
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, wegen Satz 8.3.b).
Betrachte zunächst die Teilfolge $\overline{X}_{n^2}$. Mit der Tschebyscheff-Ungleichung folgt

$$\begin{aligned} P\left(\left|\overline{X}_{n^2}\right| > \varepsilon\right) &\leq \frac{\mathrm{Var}\left(\overline{X}_{n^2}\right)}{\varepsilon^2} = \frac{\mathrm{Var}\left(\frac{1}{n^2}\sum_{i=1}^{n^2} X_i\right)}{\varepsilon^2} \\ &= \frac{\frac{1}{n^4}\sum_{i=1}^{n^2}\mathrm{Var}(X_i)}{\varepsilon^2} \leq \frac{\frac{1}{n^4}n^2 M}{\varepsilon^2} = \frac{M}{n^2\varepsilon^2} \qquad \forall\, n \in \mathbb{N},\ \varepsilon > 0 \end{aligned}$$

Also gilt

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

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

$$\mathrm{Var}\left(k\overline{X}_k - n^2\overline{X}_{n^2}\right) = \mathrm{Var}\left(\sum_{i=n^2+1}^{k} X_i\right) \leq (k-n^2)M.$$

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 gilt mit Satz 8.3.b).

$$\frac{k}{n(k)^2}\overline{X}_k - \overline{X}_{n(k)^2} \underset{k\to\infty}{\longrightarrow} 0 \quad \text{P-f.s.}$$

Insgesamt folgt:

$$\begin{aligned} &\exists F_1 \in \mathfrak{A},\ P(F_1) = 1\ \forall \omega \in F_1 :\ \overline{X}_{n^2} \overset{n\to\infty}{\longrightarrow} 0 \\ &\exists F_2 \in \mathfrak{A},\ P(F_2) = 1\ \forall \omega \in F_2 :\ \frac{k}{n(k)^2}\overline{X}_k - \overline{X}_{n(k)^2} \overset{k\to\infty}{\longrightarrow} 0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \frac{k}{n(k)^2} \overset{k\to\infty}{\longrightarrow} 1 \end{aligned}$$

Zusammengenommen folgt: $\overline{X}_n \underset{n\to\infty}{\longrightarrow} 0 \quad \forall\, \omega \in F_1 \cap F_2$, mit

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

d.h. $\overline{X}_n \underset{n\to\infty}{\longrightarrow} 0$ P-fast sicher. $\square$

Bemerkung 8.6.

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

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

Die gefolgerte Aussage heißt „*Schwaches Gesetz großer Zahlen*“ (engl.: *Weak Law of Large Numbers* (*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 $E(X_1) = \mu$ gilt

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

Beweis. aufwendig (s. Shiryayev) $\square$

Beispiel 8.7 (Anwendung des SGGZ auf Wahrscheinlichkeiten). Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von stochastisch unabhängigen, identisch verteilten Zufallsvariablen, $A \in \mathfrak{B}^1, P(X_n \in A) = p \quad \forall\, n \in \mathbb{N}$.
$\sum_{i=1}^{n} \mathbb{I}_A(X_i)$ beschreibt die Anzahl des Auftretens des Ereignisses $\{X_i \in A\}$ 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

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} Y_i &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_A(X_i) \\ &= \frac{\text{Anz. des Auftretens von } \{X_i \in A\} \text{ bis zum } n\text{-ten Versuch}}{\text{Anzahl aller Versuche}} \\ &= \text{rel. Häufigkeit des Auftretens von } \{X_i \in A\} \text{ in } n \text{ Versuchen} \\ &\underset{n\to\infty}{\longrightarrow} p \quad \text{P-fast sicher.} \end{aligned}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n \underset{n\to\infty}{\longrightarrow} \mu \quad \text{P-fast sicher}$$

d.h. $\overline{X}_n$ ist ein „stark konsistenter Schätzer“ für μ. Des weiteren gilt

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

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

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + (\overline{X}_n)^2) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n\sum_{i=1}^{n} X_i\right) + (\overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2(\overline{X}_n)^2 + (\overline{X}_n)^2 \\
&= \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\to E(X_1^2)} - \underbrace{(\overline{X}_n)^2}_{\to\mu^2} \underset{n\to\infty}{\longrightarrow} \underbrace{EX_1^2 - (EX_1)^2}_{\mathrm{Var}(X)} \text{ P-fast sicher}
\end{aligned}$$

wegen des starken Gesetzes großer Zahlen. Also:

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \underset{n\to\infty}{\longrightarrow} \sigma^2 \ \text{ P-f.s.,}$$

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 existierender Varianz $\sigma^2 = \mathrm{Var}(X_n) > 0$. Dann gilt

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

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

Beweis. z.B. „Casella und Berger“, Seite 216 ff. □

Für $S_n = \sum_{i=1}^n X_i$ gilt $E(S_n) = n\mu,\ \mathrm{Var}(S_n) = n\sigma^2$. Dann ist

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

Mit dem Zentralen Grenzwertsatz gilt: Die standarisierte Summe Y_n von Zufallsvariablen ist asymptotisch $N(0{,}1)$-verteilt.

Beispiel 8.10. $\{X_n\}_{n\in\mathbb{N}}$ sei eine Folge stochastisch unabhängiger, identisch $\mathrm{Bin}(1,p)$-verteilter Zufallsvariablen.
Somit ist $E(X_n) = p$ und $\mathrm{Var}(X_n) = p(1-p)$.
Mit dem zentralen Grenzwertsatz folgt

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

Also ist

$$\begin{aligned} P\left(\sum_{i=1}^n X_i \le z\right) &= \sum_{k=0}^{\lfloor z \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{(aufwendig für große } n\text{)} \\ &= P\left(\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \le \frac{z-np}{\sqrt{np(1-p)}}\right) \approx \phi\left(\frac{z-np}{\sqrt{np(1-p)}}\right), \end{aligned}$$

wobei

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

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

Anwendung: Lieferung von 100 Teilen, Wahrscheinlichkeit dafür, dass ein einzelnes Teil defekt ist, sei $p = 0,08$.
Gesucht: Wahrscheinlichkeit, dass mehr als 9% der Teile in der Lieferung defekt sind:

$$\begin{aligned} P\left(\sum_{i=1}^n X_i > 0,09n\right) &= 1 - P\left(\sum_{i=1}^n X_i \le 0,09n\right) \\ &\approx 1 - \phi\left(\frac{0,09n - 0,08n}{\sqrt{n \cdot 0,08 \cdot 0,92}}\right) \\ &= 1 - \phi\left(\frac{0,01}{\sqrt{0,08 \cdot 0,92}}\sqrt{n}\right) \\ &\overset{n=100}{=} 1 - \phi(0,3686) = 1 - 0,64 = 0,36 \end{aligned}$$

Kapitel 9

Schätzfunktionen und Konfidenzintervalle

Zur Motivation der folgenden Begriffe zunächst drei Beispiele und Problemstellungen

a) Eine Münze werde in unabhängigen Versuchen geworfen. Die Wahrscheinlichkeit p, dafür daß „Kopf" fällt, sei nicht bekannt. Ziel ist es, eine gute Schätzung für p zu bestimmen.
   Mathematisches Modell: Zufallsvariablen $X_1, \ldots, X_n$ seien stochastisch unabhängig, identisch $\mathrm{Bin}(1, p)$-verteilt. Dann ist

$$\hat{p} = \hat{p}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n \underset{n \to \infty}{\longrightarrow} p \quad \text{P-fast sicher}$$

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

   Beispiel: $(1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1) \;\Rightarrow\; \frac{1}{14} \sum_{i=1}^{14} x_i = \frac{8}{14}$

b) $X_1, \ldots, X_n$ seien stochastisch unabhängig, identisch $\mathrm{Exp}(\lambda)$-verteilte Zufallsvariablen, mit Parameter $\lambda > 0$. (Die Familie der Exponentialverteilungen ist eine parametrische Familie.)

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

   ist ein nach dem SGGZ konsistenter 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. bei der Fragestellung nach der Anzahl M der regelmäßigen Besucher einer Web-Seite.

   Um M zu schätzen, geht man intuitiv wie folgt vor:

- Speichere die Adressen von n Besuchern (z.B. in einem log-file). Diese Besucher bezeichnen wir als „markiert“.
- Zu einem späteren Zeitpunkt: merke m Adressen und bestimme den Anteil x der bereits markierten Besucher.

Unter entsprechenden Verteilungsannahmen sollte $\frac{n}{M} \approx \frac{x}{m}$ gelten. Ein „vernünftiger“ Schätzer für M ist somit

$$\hat{M} = \left\lfloor \frac{n\,m}{x} \right\rfloor .$$

Kann dieses intuitive Vorgehen exakt begründet werden?

9.1 Methoden zur Bestimmung von Schätzern

Definition 9.1. $X_1, \ldots, X_n$ seien Zufallsvariablen mit gemeinsamer Verteilung $P_\vartheta^{(X_1,\ldots,X_n)}$, $\vartheta \in \Theta$. ϑ bezeichnet hierbei den Parameter, von dem die Verteilung abhängt, Θ heißt *Parameterraum*. $g(\vartheta): \ \Theta \to \mathcal{Y}$ sei eine Funktion.
Jede (meßbare) Abbildung $h(X_1, \ldots, X_n)$ mit Wertebereich $\mathcal{Y}$ heißt *statistische Schätzfunktion* oder *(Punkt-) Schätzer* (engl.: *point estimator*) von $g(\vartheta)$.

Beispiel 9.2.

a) Es seien $X_1, \ldots, X_n$ stid $\sim N(\mu, \sigma^2)$. Wähle $\vartheta = (\mu, \sigma^2)$ mit dem Parameterraum $\Theta = \mathbb{R} \times \mathbb{R}^+$.
Schätze zunächst den ersten Parameter (Erwartungswert):

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

Bereits bekannter Schätzer für $g_1(\vartheta)$:

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

Schätze den zweiten Parameter (Varianz):

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

Bekannter Schätzer für $g_2(\vartheta)$:

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

b) Seien $X_1, \ldots, X_n$ stid $\sim \mathrm{Exp}(\lambda)$. Setze $\vartheta = \lambda$ mit Parameterraum $\Theta = \mathbb{R}^+$.
Schätze den Erwartungswert

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

Ein geeigneter Schätzer ist aus Beispiel 6.3 b) bereits bekannt, nämlich

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

Schätze nun den Parameter λ direkt:

$$g_2(\vartheta) = g_2(\lambda) = \lambda$$

Es stellt sich die Frage, ob

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

ein „vernünftiger“ Schätzer für λ ist.

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

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

$$L\left(\hat{\vartheta}(x_1, \dots, x_n) \middle| x_1, \dots, x_n\right) = \sup_{\vartheta \in \Theta} L(\vartheta|x_1, \dots, x_n) \text{ für alle } x_1, \dots, x_n.$$

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

Für diskrete Verteilungen verfolgt man hierbei die Idee, den Parameter θ so zu wählen, daß die gemeinsame Zähldichte bei dem Beobachtungsvektor $(x_1, \dots, x_n)$ eine Maximalstelle hat, der Beobachtungsvektor also mit maximaler Wahrscheinlichkeit angenommen wird.
Die letzte Aussage ist für absoult-stetige Verteilungen nicht richtig, da jedes Elementarereignis (einelementiges Ereignis) nur mit Wahrscheinlichkeit Null auftritt. Analog jedoch wählt man den Parameter so, daß die unterliegende Dichte für die beobachteten Werten $(x_1, \dots, x_n)$ maximal ist.

Bemerkung.

a) $\hat{\vartheta}$ zu bestimmen, ist eine Maximierungsaufgabe, die oft durch Differenzieren und Nullstellenbestimmung der ersten Ableitung gelöst werden kann.

b) Oft ist es günstiger, statt $L(\vartheta|x)$ die *Log-Likelihood-Funktion* $\log L(\vartheta|x)$ zu betrachten. Die Maximalstelle $\hat{\vartheta}$ ändert sich dadurch nicht, da der Logarithmus eine monotone Funktion ist. Das Maximum von $L(\vartheta|x)$ und $\log L(\vartheta|x)$ wird also an der gleichen Stelle angenommen.

Beispiel 9.4.

a) Seien $X_1, \ldots, X_n$ stid $\sim N(\mu, \sigma^2)$. Gesucht ist ein Schätzer für $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$. D.h. die Familie der Verteilung ist bekannt, der Parameter der Verteilung aber nicht. Nach Definition 9.3 gilt

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

Setze nun $\tau = \sigma^2$ und bestimme die Nullstellen der partiellen Ableitungen:

$$\begin{aligned} \frac{\partial \log L}{\partial \mu} &= \frac{2}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \overset{!}{=} 0 \\ \frac{\partial \log L}{\partial \tau} &= -\frac{n}{2} \frac{2\pi}{2\pi\tau} + \frac{1}{2\tau^2} \sum_{i=1}^{n} (x_i - \mu)^2 \overset{!}{=} 0 \end{aligned}$$

Aus der ersten Gleichung folgt

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

Dies eingesetzt in die zweite liefert

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 .$$

Nach der Überprüfung, daß für diese Werte tatsächlich ein Maximum vorliegt, ergibt sich als Maximum-Likelihood-Schätzer

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

b) Seien $X_1, \ldots, X_n$ stid $\sim \operatorname{Exp}(\lambda)$. Gesucht ist ein Schätzer für $\vartheta = \lambda \in \mathbb{R}^+$. Es gilt

$$\begin{aligned} f(x_1, \ldots, x_n|\lambda) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \\ &= L(\lambda|x_1, \ldots, x_n), \qquad x_i \geq 0. \\ \Rightarrow \quad \log L(\lambda|x_1, \ldots, x_n) &= n \log \lambda - \lambda \sum_{i=1}^{n} x_i. \end{aligned}$$

Für die Nullstelle der ersten Ableitung nach λ ergibt sich

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

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

c) Seien $X_1, \dots, X_n$ stid $\sim \mathrm{Bin}(1, p)$. Gesucht ist ein Schätzer für den Parameter $p \in [0, 1]$. Es gilt

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

Das Maximum wird durch die Nullstelle der ersten Ableitung nach p bestimmt,

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

Also ist $\hat{p} = \overline{x}$ ein Maximum-Likelihood-Schätzer.

9.1.1 Bayes-Methode

Eine weitere allgemeine Methode zur Bestimmung von Schätzern ist die *Bayes-Methode*, die auf dem folgenden Konzept beruht.

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

Die Berechnung von $f(\vartheta|x)$ geschieht wie in Kapitel 7. Hier wird die Berechnung im Fall von Dichten durchgeführt. Für Zähldichten oder gemischte Fälle verlaufen die Rechnungen analog.

Gegeben sind $f(x|\vartheta)$ und $\pi(\vartheta)$. Dann gilt

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

Also folgt

$$f(\vartheta|x) = \frac{f(x,\vartheta)}{f(x)} = \frac{f(x|\vartheta)\pi(\vartheta)}{\underbrace{\int f(x|\vartheta)\pi(\vartheta)d\vartheta}_{\text{normierende Konstante}}}.$$

Ein naheliegender Schätzer für ϑ ist der Erwartungswert der zu $f(\vartheta|x)$ gehörigen Verteilung.

Definition 9.5. Es gelten die obigen Bezeichnungen. Die Zufallsvariable $\hat{\vartheta}(x)$ besitze die (Zähl-) Dichte $f(\vartheta|x)$. Dann heißt $E\left(\hat{\vartheta}(x)\right)$ *Bayes-Schätzer* von ϑ.

Beispiel 9.6. Sei $X \sim \mathrm{Bin}(n,p)$, $n \in \mathbb{N}$ fest.
Die a-priori Verteilung für p sei $\mathrm{Beta}(\alpha,\beta)$, $\alpha,\beta > 0$ mit Dichte

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

$$\left(\text{Es gilt: } X \sim \mathrm{Beta}(\alpha,\beta) \quad \Rightarrow \quad E(X) = \frac{\alpha}{\alpha+\beta}.\right)$$

Sei $x \in \{0,1,\ldots,n\}$. Es gilt

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

Der Schritt von der zweiten zur dritten Zeile läßt sich wie folgt ohne größere Rechnung durchführen. Man sieht der Dichte $f(p \mid x)$ an, daß sie die Form einer Beta-Dichte hat. Es brauchen lediglich die richtigen Parameter aus den Exponenten von p und $1-p$ identifiziert werden. Diese sind $x+\alpha$ und $n-x+\beta$, so daß die Darstellung der Dichte in der dritten Zeile durch Einsetzen in die allgemeine Gestalt der Beta-Dichte folgt. Insgesamt ergibt sich als a-posteriori-Verteilung eine $\mathrm{Beta}(x+\alpha, n-x+\beta)$-Verteilung. Der Bayes-Schätzer lautet damit

$$\hat{p}(x) = \frac{x+\alpha}{n+\alpha+\beta}.$$

Speziell folgt für $\alpha = \beta = 1$, daß

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

Man beachte, daß Beta$(1,1) = R(0,1)$ gilt, also in diesem Fall durch die a-priori-Verteilung kein Parameter p besonders bevorzugt wird.

Der Maximum-Likelihood-Schätzer für p lautet bekanntlich $\hat{p}_{ML}(x) = \frac{x}{n}$. Die beiden Konzepte zur Schätzung liefern also verschiedene Ergebnisse. Nötig ist die Entwicklung von weiteren Gütekriterien, um Schätzfunktionen vergleichen zu können. Dies geschieht im folgenden.

9.2 Gütekriterien für Schätzer

Definition 9.7. $H = H(X_1, \ldots, X_n)$ sei ein Schätzer für $g(\vartheta) \in \mathbb{R}$.

$$\mathrm{MSE}(\vartheta) = E_\vartheta\big(H - g(\vartheta)\big)^2 = E_\vartheta\big(H(X_1, \ldots, X_n) - g(\vartheta)\big)^2$$

heißt *mittlerer quadratischer 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 wurde gesetzt

$$\begin{aligned} \mathrm{Var}_\vartheta(H) &= E_\vartheta(H - E_\vartheta H)^2 \quad \text{und} \\ \mathrm{Bias}_\vartheta(H)) &= E_\vartheta(H) - g(\vartheta). \end{aligned}$$

$\mathrm{Var}_\vartheta(H)$ beschreibt die *Präzision* oder die *Variablilität* des Schätzers und $\mathit{Bias}_\vartheta(H)$ die *Schiefe* oder *Genauigkeit* des Schätzers. Die mittlere quadratische Abweichung läßt sich also in die Varianz und das Quadrat des Bias additiv zerlegen.
Wichtig ist der Fall, daß $\mathrm{Bias}_\vartheta = 0 \ \forall\ \vartheta \in \Theta$.

Definition 9.8. Ein Schätzer H heißt *erwartungstreu* (engl.: *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 H

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

so daß die Varianz des Schätzers die mittlere quadratische Abweichung vom zu schätzenden Wert $g(\vartheta)$ mißt.

Beispiel 9.9. Seien $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ stid.

a) Zur Überprüfung, ob der Schätzer $\hat{\mu} = H_1(X_1, \dots, X_n) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i = \overline{X}$ erwartungstreu für μ ist, wird sein Erwartungswert berechnet.

$$E_\vartheta(\hat{\mu}) = E_\vartheta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E_\vartheta(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \mu \quad \forall\, \vartheta = (\mu, \sigma^2) \in \Theta.$$

Also ist $\overline{X}$ ein erwartungstreuer Schätzer für den Parameter μ.

b) Analog gehen wir bei der Untersuchung von $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ auf Erwartungstreue für σ^2 vor.

In einer Vorüberlegung wird zunächst $\mathrm{E}(Z^2)$ einer $\mathrm{N}(\mu, \sigma^2)$-verteilten Zufallsvariablen bestimmt. Bekanntlich gilt

$$\sigma^2 = \mathrm{Var}(Z) = \mathrm{E}(Z^2) - (\mathrm{E}Z)^2 = \mathrm{E}(Z^2) - \mu^2,$$

so daß $\mathrm{E}(Z^2) = \mu^2 + \sigma^2$ folgt.

Wegen der Faltungsstabilität der Normalverteilung und den bekannten Rechenregeln für Erwartungswerte und Varianz folgt ferner, daß

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

Zusammen mit

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - (\overline{X})^2$$

folgt für den Erwartungswert

$$\begin{aligned} E_\vartheta\left(\widehat{\sigma^2}\right) &= \frac{1}{n}\sum_{i=1}^n \underbrace{E\left(X_i^2\right)}_{=\sigma^2+\mu^2} \quad - \quad \underbrace{E\left(\overline{X}^2\right)}_{=\frac{\sigma^2}{n}+\mu^2} \\ &= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \\ &= \frac{n-1}{n}\sigma^2. \end{aligned}$$

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

$$\lim_{n\to\infty} E_\vartheta\left(\widehat{\sigma^2}\right) = \sigma^2 \quad \text{für alle } (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+.$$

Multipliziert man $\widehat{\sigma^2}$ mit dem Faktor $n/(n-1)$, so folgt:

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

ist ein erwartungstreuer Schätzer für σ^2. Denn

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

für alle $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

Beispiel 9.10 (Gütevergleich von zwei Schätzern). Sei $X \sim \mathrm{Bin}(n,p)$. Für p wurden zwei Schätzer vorgestellt

Maximum-Likelihood-Schätzer: $\hat{p}_{\mathrm{ML}} = \dfrac{x}{n}$

Bayes-Schätzer zur a-priori-Verteilung $\mathrm{Beta}(\alpha, \beta)$: $\hat{p}_{\mathrm{B}} = \dfrac{x+\alpha}{n+\alpha+\beta}$

Zur Beantwortung der Frage, welcher Schätzer geeigneter ist, werden beide mit Hilfe des MSE, des mittleren quadratischen Fehlers, verglichen.

$$E_p\left(\hat{p}_{\mathrm{ML}} - p\right)^2 = \mathrm{Var}\left(\hat{p}_{\mathrm{ML}}\right) = \mathrm{Var}\left(\frac{1}{n}X\right) \underset{\text{da } \hat{p}_{\mathrm{ML}} \text{ E-treu}}{=} \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

Für den Bayes-Schätzer gilt

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

Wähle nun α und β so, daß der mittlere quadratische Fehler von $\hat{p}_B$ konstant ist, d.h. kein Wert von p bei der Schätzung bevorzugt wird. Dies ist erfüllt, wenn

$$\alpha = \beta = 1/2 \cdot \sqrt{n}$$

gewählt werden. Mit der zugehörigen a-priori Verteilung folgt dann nach einer kurzen Rechnung

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

Die zugehörigen mittleren quadratischen Abweichungen sind als Funktionen von p in Abbildung 9.1 für $n=5$ und $n=300$ dargestellt.

Für kleine Stichproben sollte der Bayes-Schätzer bevorzugt werden, solange es keine starken Indizien für $p \approx 0$ oder $p \approx 1$ gibt. Für große Stichproben sollte der Maximum-Likelihood-Schätzer bevorzugt werden, solange es keine fundierten Hinweise für $p \approx \frac{1}{2}$ gibt.

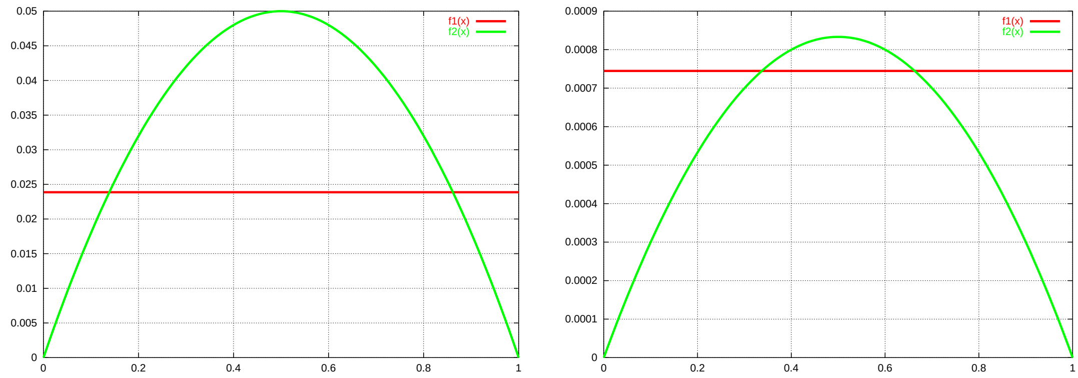


Abbildung 9.1: MSE von $\hat{p}_{\mathrm{B}}$ (rote Kurven, konstant) und MSE von $\hat{p}_{\mathrm{ML}}$ (grüne Kurven) für $n = 5$ (links) und $n = 300$ (rechts).

9.3 Konfidenzintervalle

Im vorherigen Abschnitt wurden Punktschätzer behandelt. Ziel war es, Schätzfunktionen $\hat{\vartheta}$ zu bestimmen, die unter den genannten Optimalitätskriterien möglichst nahe am tatsächlichen Parameter ϑ liegen.
Im folgenden werden Intervalle mit zufälliger oberer und unterer Grenze bestimmt, die den wahren Parameter ϑ mit vorgegebener Wahrscheinlichkeit überdecken. Die Idee zur Konstruktion solcher Intervalle wird an folgendem Beispiel klar.
Seien $X_1, \ldots, X_n$ stid $\sim \mathrm{Bin}(1, p)$ (binomialverteilt mit Parametern $n = 1$ und $p \in [0, 1]$). Mit dem schwachen Gesetz großer Zahlen folgt:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall\, \varepsilon > 0.$$

Äquivalent umgeformt bedeutet dies

$$P\left(\overline{X} - \varepsilon \leq p \leq \overline{X} + \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 1 \qquad \forall\, \varepsilon > 0$$

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 nun die Kenntnisse über die Verteilung von $\overline{X}$, um ε und die zugehörige Überdeckungswahrscheinlichkeit P in Abhängigkeit vom *Stichprobenumfang* n zu quantifizieren.

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

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

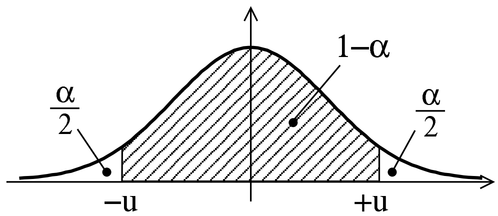
Beispiel 9.12. Seien $X_1, \ldots, X_n$ stid $\sim \mathrm{N}\left(\mu, \sigma^2\right)$, der Parameter σ^2 (die Varianz) sei fest vorgegeben und bekannt.
Bekanntlich ist das arithmetische Mittel $\overline{X}$ normalverteilt mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$. (Man verifiziere dies auf Grund der Faltungsstabilität der Normalverteilung.) Also folgt

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

Bestimmt wird nun ein $u \in \mathbb{R}$ so, daß

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

Das Prinzip hierfür wird in der nachfolgenden Skizze anhand der Dichte der $\mathrm{N}(0,1)$-Verteilung erläutert.



$u = u_{1-\frac{\alpha}{2}}$ ist also das $(1-\frac{\alpha}{2})$-Fraktil der N(0,1)-Verteilung. Auflösen nach μ ergibt die folgende Gleichung

$$P\left(\left|\overline{X}-\mu\right| \leq \frac{\sigma}{\sqrt{n}}\, u_{1-\frac{\alpha}{2}}\right) = P\left(\overline{X} - \frac{\sigma}{\sqrt{n}}\, u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}}\, u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Also ist

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

ein $(1-\alpha)$-Konfidenzintervall für μ.
Wie erwartet fällt die Länge des Konfidenzintervalls mit wachsendem n und wächst mit wachsendem Niveau $1-\alpha$.

Einseitige Konfidenzintervalle können analog aus aus der folgenden Gleichung konstruiert werden.

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

Wählt man $u = u_{1-\alpha}$, so ist

$$P\left(\overline{X}-\mu \leq \frac{\sigma}{\sqrt{n}}\, u_{1-\alpha}\right) = P\left(\mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}}\, u_{1-\alpha}\right) = 1-\alpha.$$

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

Tabelle 9.1: Fraktile der N(0, 1)-Verteilung.

Einige Fraktile der Standardnormalverteilung sind in Tabelle 9.1 angegeben. Ausführliche Tabellen finden sich im Buch von *G. Casella, R.L. Berger: Statistical Inference, Duxbury Press, Belmont California, 1990.*

Beispiel 9.13 (Approximatives Konfidenzintervall für Wahrscheinlichkeiten). Seien $X_1, \ldots, X$ stid $\sim \mathrm{Bin}(1, p)$. Dann ist

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{as}}{\sim} \mathrm{N}(0,1) \quad \text{(siehe Bsp. 8.10).}$$

Also gilt

$$P\left(\frac{|\overline{X} - p|}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\alpha/2}\right) \approx 1 - \alpha,$$

wobei $u_{1-\alpha/2}$ das Fraktil der N(0,1)-Verteilung ist. Ein zugehöriges Konfidenzintervall kann wie folgt bestimmt werden.

$$|\overline{X} - p| = u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \quad \Longleftrightarrow \quad \left(\overline{X} - p\right)^2 = u^2_{1-\alpha/2}\frac{p(1-p)}{n}$$

ist eine quadratische Gleichung in p mit zwei Lösungen $p_L\left(\overline{X}\right) \leq p_U\left(\overline{X}\right)$. Folglich ist

$$\left[p_L\left(\overline{X}\right),\ p_U\left(\overline{X}\right)\right]$$

ein approximatives $(1-\alpha)$-Konfidenzintervall für p. Die explizite Lösung der quadratischen Gleichung findet sich im o.a. Buch von Casella und Berger auf Seite 445.

In den meisten Anwendungsfällen wird die Varianz, wie in Beispiel 9.12 angenommen, nicht bekannt sein. Um auch hier Konfidenzintervalle zu bestimmen, wird zunächst die Verteilung von $(\bar{X} - \mu)/(S/\sqrt{n})$ eingeführt.

Satz 9.14 (Student's *t*-Verteilung). Seien $X_1, \ldots, X_n$ i.i.d. $\sim \mathrm{N}(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ und $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Es gilt

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

wobei t_{n-1} die t-Verteilung mit $n-1$ Freiheitsgraden bezeichnet. Sie besitzt die Dichte

$$f_{t_m}(x) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}\frac{1}{\sqrt{m\pi}}\frac{1}{(1+\frac{x^2}{m})^{(m+1)/2}}, \quad x \in \mathbb{R},\ m \in \mathbb{N}.$$

Es gilt sogar, daß $\bar{X}$ und S^2 stochastisch unabhängig sind. Der Beweis dieser Aussage und des Satzes findet sich in Casella, Berger auf Seite 226.
Man beachte folgende Extremfälle für den Parameter m der t-Verteilung.

- Für $m = 1$ ergibt sich die Dichte der *Cauchy-Verteilung*.
- Für $Y_m \sim t_m$, $m \in \mathbb{N}$, gilt $Y_m \overset{\text{as}}{\approx} \mathrm{N}(0,1)$ mit $m \to \infty$. t_m-Verteilungen konvergieren also mit $m \to \infty$ gegen die Standardnormalverteilung.

Für $m = 1, 3, 10$ sind die Dichten der jeweiligen t_m-Verteilung und die Dichte der $\mathrm{N}(0,1)$-Verteilung in Abbildung 9.2 dargestellt.

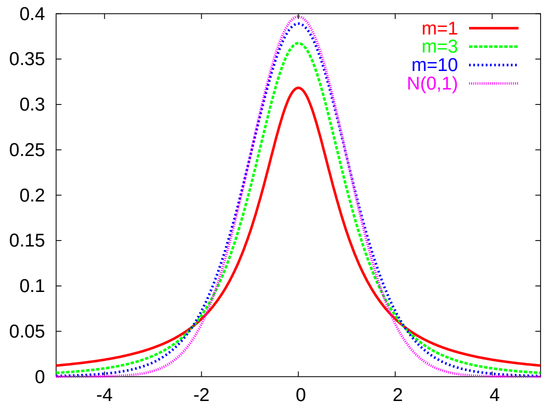


Abbildung 9.2: t_m-Verteilungsdichten für $m = 1, 3, 10$ und die Dichte der $\mathrm{N}(0,1)$-Verteilung.

Die t-Verteilung wurde von W.S. Gosset unter dem Pseudonym Student in den frühen 90er Jahren publiziert. Ihr Name "*Student's t*" erklärt sich hieraus.

Beispiel 9.15. Seien $X_1, \ldots, X_n$ i.i.d. $\sim \mathrm{N}(\mu, \sigma^2)$. Bestimmt wird ein Konfidenzintervall für μ bei unbekanntem σ^2. Die Idee ist, in Beispiel 9.12 σ^2 durch den erwartungstreuen Schätzer S^2 zu ersetzen und dann die Fraktile der t-Verteilung zu benutzen.
Bestimmt wird $v \in \mathbb{R}$ so, daß

$$P\Big(\Big|\frac{\bar{X} - \mu}{S/\sqrt{n}}\Big| \le v\Big) \le 1 - \alpha.$$

Analog zu Beispiel 9.12 wird $v = v_{1-\alpha/2}$ als das $(1 - \alpha/2)$-Fraktil der t_{n-1}-Verteilung gewählt. Auflösen obiger Gleichung nach μ liefert

$$P\Big(\big|\bar{X} - \mu\big| \le v_{1-\alpha/2}\,\frac{S}{\sqrt{n}}\Big) = P\Big(\bar{X} - v_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + v_{1-\alpha/2}\,\frac{S}{\sqrt{n}}\Big).$$

Also ist

$$\Big[\bar{X} - v_{1-\alpha/2}\,\frac{S}{\sqrt{n}}\,,\ \bar{X} + v_{1-\alpha/2}\,\frac{S}{\sqrt{n}}\Big]$$

ein $1 - \alpha$-Konfidenzintervall für μ bei unbekannter Varianz σ^2.

Index