

Stochastik I

Sommersemester 2003

Prof.Dr. Rudolf Grübel

Dieses Skript wurde erstellt von Alexander Seifert.
Anregungen und Anmerkungen schickt bitte an alexseifert@gmx.net.
Meine Homepage: www.blu7.com.

Stand des Inhalts: 31.03.2005

Version 2006.04.18.a

Inhaltsverzeichnis

| | |
|---|-----------|
| Kapitel 1: Grundbegriffe, Kolmogorov-Axiome..... | 4 |
| Kapitel 2: Bedingte Wahrscheinlichkeiten und Unabhängigkeit..... | 8 |
| Kapitel 3: Laplace-Experimente | 12 |
| 3.1 Allgemeines | 12 |
| 3.2 Etwas Kombinatorik, oder: Die Kunst des Zählens | 13 |
| 3.3 Einige typische Probleme..... | 15 |
| 3.3.1 Das Geburtstagsproblem | 15 |
| 3.3.2 Ein Bridge-Problem | 15 |
| 3.3.3 Der zerstreute Postbote..... | 16 |
| 3.3.4 Das Paradox von de Méré *..... | 17 |
| 3.4 Ein Blick in die Ferne | 17 |
| 3.4.1 Die Nadel von Buffon | 17 |
| 3.4.2 Das Paradox von Bertrand | 18 |
| 3.4.3 You can't always get what you want | 18 |
| Kapitel 4: Diskrete Wahrscheinlichkeitsräume und Zufallsgrößen | 19 |
| 4.1 Allgemeines | 19 |
| 4.2 Einige wichtige diskrete Verteilungen..... | 20 |
| 4.3 Erwartungswert und Varianz von Zufallsvariablen..... | 22 |
| 4.4 Bedingte Verteilungen und Unabhängigkeit | 24 |
| 4.5 Reellwertige diskrete Zufallsgrößen..... | 27 |
| 4.6 Wahrscheinlichkeitserzeugende Funktionen | 31 |
| 4.7 Das schwache Gesetz der großen Zahlen | 33 |
| Kapitel 5: Allgemeine Wahrscheinlichkeitsräume..... | 35 |
| 5.1 Mengensysteme..... | 35 |
| 5.2 Zufallsgrößen und Verteilungen | 38 |
| 5.3 Reellwertige Zufallsgrößen..... | 39 |
| 5.4 Verteilungsfunktionen | 41 |
| 5.5 Einige wichtige Verteilungen mit Riemann-Dichten..... | 43 |
| 5.6 Erwartungswerte..... | 45 |
| 5.7 Unabhängigkeit | 45 |
| Kapitel 6: Verteilungskonvergenz und Normalapproximation | 49 |
| 6.1 Verteilungskonvergenz..... | 49 |

| | |
|---|-----------|
| 6.2 Normalapproximation bei Poisson-Verteilungen..... | 50 |
| 6.3 Normalapproximation bei Binomialverteilungen | 52 |
| 6.4 Normalapproximation bei Gamma-Verteilungen | 53 |
| Kapitel 7: Grundbegriffe der mathematischen Statistik..... | 55 |
| 7.1 Allgemeines | 55 |
| 7.2 Schätztheorie..... | 55 |
| 7.3 Tests | 57 |
| 7.4 Konfidenzbereiche..... | 62 |

Anmerkung des Erstellers (dieses Skripts): Dieses Skript basiert auf dem vom Prof.Dr. Grübel im Sommersemester 2003 erstellten Skript. Zudem habe ich das überaus interessante Skript vom Sommersemester 1996, das von Sven Meyer erstellt wurde, zum Vergleich genommen und alle aus meiner Sicht interessanten und wichtigen Dinge (Beispiele, Bemerkungen, Zeichnungen, etc.) übernommen, die in dem neueren fehlen. Die mit * gekennzeichneten Abschnitte sind aus dem genannten Skript übernommen. Die mit ** gekennzeichneten Stellen sind dem aktuellsten Skript vom Sommersemester 2006 entnommen.

Kapitel 1

Grundbegriffe, Kolmogorov-Axiome

Stochastik (ein moderner Sammelbegriff für die Gebiete Wahrscheinlichkeitstheorie und mathematische Statistik) beschäftigt sich mit Zufallsexperimenten, bei denen das Ergebnis nicht durch die Randbedingungen des Experiments festgelegt ist. Der **Ergebnisraum** Ω ist eine Menge, die die möglichen Ergebnisse (Resultate) des Experiments enthält, **Ereignisse** werden durch Teilmengen von Ω beschrieben. Aussagen über das Ergebnis werden dabei in Teilmengen des Ergebnisraumes übersetzt: eine Aussage wird zu der Menge aller $\omega \in \Omega$, für die diese Aussage richtig ist.

Beispiel 1.1

- (i) Würfelwurf: $\Omega = \{1, 2, 3, 4, 5, 6\}$, das Ereignis „eine gerade Zahl erscheint“ wird repräsentiert durch (ist) $A = \{2, 4, 6\}$.
- (ii) Eine Münze wird n -mal geworfen. Schreibt man 0 für Kopf und 1 für Zahl, so ist
- $$\Omega = \{(i_1, \dots, i_n) \mid i_j \in \{0, 1\}\} = \{0, 1\}^n$$
- ein geeigneter Ergebnisraum. Hierbei bedeutet $\omega = (i_1, \dots, i_n)$, dass für $j = 1, \dots, n$ im j -ten Wurf i_j erscheint. Das Ereignis „in allen n Würfeln erscheint dieselbe Seite“ wird beschrieben durch
- $$\{(0, 0, \dots, 0), (1, 1, \dots, 1)\}.$$
- (iii) Eine Probe radioaktiven Materials emittiert Partikel eines bestimmten Typs. Zählt man die Emissionen in einem bestimmten Zeitraum, so wird man auf $\Omega = \{0, 1, 2, \dots\} =: \mathbb{N}_0$ geführt. $\{10, 11, 12, \dots\}$ beschreibt das Ereignis „mindestens 10 Partikel werden emittiert“. Wartet man auf den ersten Zerfall, so ist $\Omega = [0, \infty)$ ein geeigneter Ergebnisraum.
- (iv) Rotierender Zeiger (Roulette etc.) $\theta = 2\pi x$, $0 \leq x < 1$, sei der Winkel beim Stillstand, $\Omega = [0, 1)$. Das Ereignis „Zeiger stoppt im rechten oberen Quadranten“ wird durch $A = (0, \frac{1}{4})$ bzw. $A = [0, \frac{1}{4}]$ beim Einschluss der Ränder beschrieben.

Ein Ereignis A mit exakt einem Element, also $A = \{\omega\}$ mit einem $\omega \in \Omega$, nennt man ein **Elementarereignis**. Ergebnisse sind also Elemente von Ω , Ereignisse Teilmengen von Ω . Kombinationen von Ereignissen können durch mengentheoretische Operationen beschrieben werden:

$$\begin{aligned} A \cap B: & \quad A \text{ und } B \text{ treten beide ein,} \\ A \cup B: & \quad A \text{ oder } B \text{ (oder beide) tritt (treten) ein,} \\ A^c: & \quad A \text{ tritt nicht ein.} \end{aligned}$$

Beim Würfelwurf wird beispielsweise das Ereignis „es erscheint keine gerade Zahl“ beschrieben durch

$$\{2, 4, 6\}^c = \{1, 3, 5\}.$$

Beispiel 1.2 (Kombinationen von mehr als zwei Ereignissen)

- (i) „Genau eines der Ereignisse A, B, C tritt ein“ wird beschrieben durch

$$A \cap B^c \cap C^c + A^c \cap B \cap C^c + A^c \cap B^c \cap C.$$

Hierbei steht $A + B$ für $A \cup B$ bei disjunkten Mengen A, B .

- (ii) Es sei A_1, A_2, \dots eine Folge von Ereignissen. Dann wird das Ereignis „unendlich viele der A_i treten ein“ repräsentiert durch den **Limes Superior** der Mengenfølge,

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

Klar: $\bigcup_{m=n}^{\infty} A_m$ steht für „mindestens eines der Ereignisse mit Index $\geq n$ tritt ein“ und es gilt:

$$\omega \in \limsup_{n \rightarrow \infty} A_n \iff \forall n \in \mathbb{N} : \exists m \geq n : \omega \in A_m \iff \#\{n \in \mathbb{N} \mid \omega \in A_n\} = \infty.$$

Die Menge der Ereignisse (eine Menge von Mengen!) in einem Zufallsexperiment bildet ein Mengensystem \mathcal{A} über Ω , also eine Teilmenge der Potenzmenge $\mathfrak{P}(\Omega)$ von Ω . Bei endlichem oder abzählbar unendlichem Ergebnisraum (Beispiel 1.1 (i), (ii), erster Teil von (iii)) können wir problemlos $\mathcal{A} = \mathfrak{P}(\Omega)$ voraussetzen (jede Zusammenfassung von Ergebnissen ist ein Ereignis), bei überabzählbarem Ω geht dies in vielen wichtigen Fällen (Beispiel 1.1 (iv)) nicht. Wir werden dies später präzisieren, es besteht eine entfernte Verwandtschaft zur Russellschen Antinomie (nicht jede Zusammenfassung von Objekten ist eine Menge). Die obigen Beispiele führen auf gewisse Mindestvoraussetzungen an das System \mathcal{A} und damit zur folgenden Definition.

Definition 1.3

$\mathcal{A} \subset \mathfrak{P}(\Omega)$ heißt eine σ -Algebra über Ω , wenn gilt:

- (i) $\Omega \in \mathcal{A}$,
- (ii) Aus $A \in \mathcal{A}$ folgt $A^c \in \mathcal{A}$,
- (iii) Aus $A_1, A_2, \dots \in \mathcal{A}$ folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

In Worten: Ein Mengensystem über Ω ist eine σ -Algebra, wenn es die Grundmenge enthält und stabil ist gegenüber den Operationen „Komplement“ und „abzählbare Vereinigung“.

Was ist nun „Wahrscheinlichkeit“? Streng genommen ist dies keine mathematische Frage (analog zu: Was ist eine Gerade? Was ist eine Menge?) Als mathematischer Gegenstand ist Wahrscheinlichkeit eine Funktion, die Ereignissen Zahlen zwischen 0 und 1 zuordnet und dabei gewissen Axiomen genügt. Diese Axiome (Forderungen) werden durch den umgangssprachlichen Wahrscheinlichkeitsbegriff motiviert. Zur Erläuterung betrachten wir die Aussage „das Ereignis A hat die Wahrscheinlichkeit p “ (z.B.: „beim Wurf eines fairen Würfels erscheint mit Wahrscheinlichkeit $\frac{1}{2}$ eine gerade Zahl“). Es gibt zwei hauptsächliche Interpretationen:

(F) Die „Häufigkeitsauffassung“, deren Anhänger auch Frequentisten genannt werden. Es sei $N_n(A)$ die Häufigkeit des Auftretens von A bei n Wiederholungen des Zufallsexperiments, $\frac{1}{n}N_n(A)$ ist die **relative Häufigkeit** von A . Bei großem n würde man erwarten, dass die relative Häufigkeit von A in der Nähe von p liegt (ungefähr die Hälfte der Würfelwürfe sollte eine gerade Zahl liefern).

(S) Die „Glaubens- oder Plausibilitätsauffassung“, deren Anhänger man gelegentlich als Subjektivisten bezeichnet. Der Wert p gibt auf einer Skala von 0 bis 1 die „Stärke meines Glaubens“ an das Eintreten von A wieder. Dies kann über Wetten formalisiert werden und ist im Gegensatz zu (F) auch bei nicht wiederholbaren Experimenten anwendbar (aber eben subjektiv).

Diese Auffassungen sind natürlich nicht disjunkt. Für relative Häufigkeiten gelten die Regeln

$$\frac{1}{n}N_n(\Omega) = 1, \quad \frac{1}{n}N_n(A) \geq 0 \text{ für alle } A \in \mathcal{A},$$

sowie für alle paarweise disjunkten $A_1, \dots, A_k \in \mathcal{A}$

$$\frac{1}{n}N_n(A_1 + \dots + A_k) = \frac{1}{n}N_n(A_1) + \dots + \frac{1}{n}N_n(A_k).$$

Insgesamt motiviert dies das folgende mathematische Modell für Zufallsexperimente:

Definition 1.4 (Die Kolmogorov-Axiome)

Ein **Wahrscheinlichkeitsraum** ist ein Tripel (Ω, \mathcal{A}, P) , bestehend aus einer nichtleeren Menge Ω (dem **Ergebnisraum**), einer σ -Algebra \mathcal{A} (dem **Ereignissystem**) über Ω und einer Abbildung $P: \mathcal{A} \rightarrow \mathbb{R}$ mit den Eigenschaften

- (i) $P(\Omega) = 1$,
- (ii) $P(A) \geq 0$ für alle $A \in \mathcal{A}$,
- (iii) $P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ für alle paarweise disjunkten $A_1, A_2, \dots \in \mathcal{A}$.

Eine Abbildung mit diesen Eigenschaften heißt **Wahrscheinlichkeitsmaß**, Eigenschaft (iii) nennt man die σ -**Additivität**.

Beispiel 1.5

- (i) Ist Ω eine endliche und nichtleere Menge, so wird durch

$$P(A) := \frac{\#A}{\#\Omega} \text{ für alle } A \subset \Omega$$

ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{P}(\Omega))$ definiert. Man nennt (Ω, \mathcal{A}, P) mit $\mathcal{A} = \mathfrak{P}(\Omega)$ das **Laplace-Experiment** über Ω . Solche Modelle werden häufig durch Symmetrieüberlegungen nahe gelegt, siehe

Beispiel 1.1 (i) und (ii). Beim Wurf eines fairen (d.h. symmetrischen) Würfels ergibt sich damit als Wahrscheinlichkeit dafür, dass eine gerade Zahl geworfen wird,

$$P(A) = \frac{\#\{2, 4, 6\}}{\#\{1, 2, 3, 4, 5, 6\}} = \frac{1}{2}$$

(Anzahl der günstigen Fälle dividiert durch die Anzahl der möglichen Fälle – eine vielleicht schon aus dem Schulunterricht bekannte Regel).

- (ii) Ein deterministisches Experiment, bei dem nur ein einziges Ergebnis ω_0 möglich ist, kann als degeneriertes Zufallsexperiment $(\Omega, \mathcal{A}, \delta_{\omega_0})$ betrachtet werden. Hierbei ist Ω irgendeine Menge, die ω_0 enthält, \mathcal{A} eine σ -Algebra über Ω und δ_{ω_0} das **Dirac-Maß** oder auch **Einpunktmaß** in ω_0 :

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

Man macht sich leicht klar, dass δ_{ω_0} in der Tat ein Wahrscheinlichkeitsmaß ist.

Im folgenden Satz sind einige erste Folgerungen aus den Axiomen zusammengefasst.

Satz 1.6

Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt:

- (i) $P(\emptyset) = 0$, $P(A) \leq 1$ für alle $A \in \mathcal{A}$,
- (ii) $P(A^c) = 1 - P(A)$ für alle $A \in \mathcal{A}$,
- (iii) Endliche Additivität:
 $P(A_1 + \dots + A_k) = P(A_1) + \dots + P(A_k)$ für alle paarweise disjunkten $A_1, \dots, A_k \in \mathcal{A}$,
- (iv) Monotonie: Aus $A \subset B$ folgt $P(A) \leq P(B)$ für alle $A, B \in \mathcal{A}$,
- (v) Boolesche Ungleichung:
 $P(A_1 \cup \dots \cup A_k) \leq P(A_1) + \dots + P(A_k)$ für alle (nicht notwendigerweise disjunkten) $A_1, \dots, A_k \in \mathcal{A}$,
- (vi) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für alle $A, B \in \mathcal{A}$,
- (vii) Formel von Poincaré (auch: Einschluss-Ausschluss-Formel oder Siebformel)

$$P(A_1 \cup \dots \cup A_k) = \sum_{\emptyset \neq H \subset \{1, \dots, k\}} (-1)^{\#H-1} P\left(\bigcap_{i \in H} A_i\right).$$

Beweis

Der Nachweis, dass die beteiligten Mengenkonstruktionen nicht aus der σ -Algebra herausführen, ist Gegenstand einer Übungsaufgabe. Beispielsweise gilt $\emptyset \in \mathcal{A}$ wegen $\Omega \in \mathcal{A}$ und $\emptyset = \Omega^c$.

- (i) Verwendet man die σ -Additivität von P mit $A_1 = A_2 = \dots = \emptyset$, so folgt $P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots$, also $P(\emptyset) = 0$. Die Aussage $P(A) \leq 1$ folgt aus $P(\Omega) = 1$ und der Monotonie (iv).
- (ii) $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$, verwende nun die endliche Additivität.
- (iii) Setze $A_{k+1} = A_{k+2} = \dots = \emptyset$, verwende die σ -Additivität und $P(\emptyset) = 0$.
- (iv) Es gilt $B = A + B \cap A^c$, also $P(B) = P(A) + P(B \cap A^c) \geq P(A)$, da $P(B \cap A^c) \geq 0$.
- (v) Im Falle $k = 2$ folgt die Aussage aus (vi) und $P(A \cap B) \geq 0$. Angenommen, die Aussage ist für ein $k \geq 2$ richtig. Dann folgt

$$P((A_1 \cup \dots \cup A_k) \cup A_{k+1}) \leq P(A_1 \cup \dots \cup A_k) + P(A_{k+1}),$$

denn für zwei Ereignisse gilt die Formel, also

$$P((A_1 \cup \dots \cup A_k) \cup A_{k+1}) \leq (P(A_1) + \dots + P(A_k)) + P(A_{k+1}),$$

d.h. die Aussage gilt dann auch für $k + 1$. Vollständige Induktion liefert nun die gewünschte Aussage.

- (vi) $A = A \cap B + A \cap B^c$, also ergibt der bereits bewiesene Teil (iii) $P(A \cap B^c) = P(A) - P(A \cap B)$. Weiter gilt $A \cup B = B + A \cap B^c$, also

$$P(A \cup B) = P(B) + P(A \cap B^c) = P(B) + P(A) - P(A \cap B).$$

- (vii) Im Falle $k = 2$ erhält man (vi). Induktionsschritt: Übungsaufgabe. □

Als einfache Anwendung betrachten wir die Frage, mit welcher Wahrscheinlichkeit beim n -fachen Wurf einer (fairen) Münze Kopf und Zahl beide je mindestens einmal erscheinen. Bezeichnet A dieses Ereignis, so gilt

$$A^c = \{(0, 0, \dots, 0), (1, 1, \dots, 1)\}.$$

Man erhält also

$$P(A) = 1 - P(A^c) = 1 - \frac{\#A^c}{\#\Omega} = 1 - \frac{2}{2^n} = 1 - \frac{1}{2^{n-1}}.$$

Warum wird in den Kolmogorov-Axiomen die σ -Additivität anstelle beispielsweise der (schwächeren) endlichen Additivität gefordert? Man sieht leicht, dass letztere bereits aus

$$P(A \cup B) = P(A) + P(B) \text{ für alle disjunkten } A, B \in \mathcal{A}$$

folgt. Das folgende Resultat zeigt, dass man σ -Additivität als Stetigkeitseigenschaft interpretieren kann. Wir nennen eine Folge $(A_n)_{n \in \mathbb{N}}$ von Teilmengen von Ω **isoton**, wenn $A_n \subset A_{n+1}$ für alle $n \in \mathbb{N}$ gilt, **antiton** im Falle $A_n \supset A_{n+1}$ für alle $n \in \mathbb{N}$. Wir schreiben beispielsweise $A_n \downarrow A$, wenn $(A_n)_{n \in \mathbb{N}}$ eine antitone Mengenfølge ist mit der Eigenschaft $\bigcap_{n=1}^{\infty} A_n = A$.

Satz 1.7

Es seien $\Omega \neq \emptyset$, \mathcal{A} eine σ -Algebra auf Ω und $P : \mathcal{A} \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften

- (a) $P(\Omega) = 1$,
- (b) $P(A) \geq 0$ für alle $A \in \mathcal{A}$,
- (c) $P(A \cup B) = P(A) + P(B)$ für alle $A, B \in \mathcal{A}$ mit $A \cap B = \emptyset$.

Dann sind äquivalent:

- (i) P ist ein Wahrscheinlichkeitsmaß (also σ -additiv),
- (ii) P ist stetig von unten, d.h. für jede isotone Folge A_1, A_2, \dots von Ereignissen gilt

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right),$$

- (iii) P ist stetig von oben, d.h. für jede antitone Folge A_1, A_2, \dots von Ereignissen gilt

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right),$$

- (iv) P ist stetig in \emptyset , d.h. für jede Folge $(A_n)_{n \in \mathbb{N}}$ von Ereignissen mit der Eigenschaft $A_n \downarrow \emptyset$ gilt

$$\lim_{n \rightarrow \infty} P(A_n) = 0.$$

Beweis

(i) \Rightarrow (ii): Es sei $B_1 := A_1$, $B_n := A_n \cap A_{n-1}^c$ für alle $n > 1$. Klar: $B_n \in \mathcal{A}$ für alle $n \in \mathbb{N}$, $(B_n)_{n \in \mathbb{N}}$ paarweise disjunkt, $A_n = B_1 + \dots + B_n$ für alle $n \in \mathbb{N}$, $\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} B_n$. Die σ -Additivität von P liefert

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{m=1}^n P(B_m) = \lim_{n \rightarrow \infty} P\left(\sum_{m=1}^n B_m\right) = \lim_{n \rightarrow \infty} P(A_n).$$

(ii) \Rightarrow (iii): Über Komplementbildung: Ist $A_n \downarrow$, so ist $A_n^c \uparrow$ und man erhält

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = 1 - \lim_{n \rightarrow \infty} P(A_n^c) = 1 - \lim_{n \rightarrow \infty} (1 - P(A_n)) = \lim_{n \rightarrow \infty} P(A_n).$$

(iii) \Rightarrow (iv): Trivial.

(iv) \Rightarrow (i): Sind A_1, A_2, \dots disjunkt, so gilt $B_n \downarrow \emptyset$ für $B_n := \sum_{k=n+1}^{\infty} A_k$, also folgt unter Verwendung der endlichen Additivität

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\sum_{k=1}^n A_k + B_n\right) = \sum_{k=1}^n P(A_k) + P(B_n).$$

Wegen $P(B_n) \rightarrow 0$ konvergiert die Reihe und ist gleich $P\left(\bigcup_{k=1}^{\infty} A_k\right)$. □

Als Ersatz für die σ -Additivität ist die endliche Additivität zu schwach für eine befriedigende mathematische Theorie. Fordert man dagegen die Additivität für beliebige, also auch überabzählbare Mengenfamilien (eine Eigenschaft, die für relative Häufigkeiten gilt), so bleibt nicht genug übrig: beim rotierenden Zeiger (Beispiel 1.1 (iv)) erwartet man $P(\{\omega\}) = 0$ für alle $\omega \in \Omega$, „Hyperadditivität“ würde dann auf $P \equiv 0$ führen.

Kapitel 2

Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Es seien A und B Ereignisse in einem Zufallsexperiment, das durch einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) beschrieben wird. Was ist die Wahrscheinlichkeit von B unter der Bedingung, dass A eintritt? Bei n Wiederholungen tritt A $N_n(A)$ -mal ein, unter diesen ist $N_n(A \cap B)$ die (absolute) Häufigkeit von B . Für die relative Häufigkeit von B unter den Experimenten, die A liefern, gilt

$$\frac{N_n(A \cap B)}{N_n(A)} = \frac{\frac{1}{n} N_n(A \cap B)}{\frac{1}{n} N_n(A)}.$$

Durch den frequentistischen Wahrscheinlichkeitsbegriff wird somit die folgende Definition motiviert.

Definition 2.1

Es sei A ein Ereignis mit $P(A) > 0$. Die **bedingte Wahrscheinlichkeit** eines Ereignisses B unter A wird definiert durch

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

Man sieht leicht, dass dann $B \mapsto P(B | A)$ ein Wahrscheinlichkeitsmaß ist, d.h. $(\Omega, \mathcal{A}, P(\cdot | A))$ ist ein Wahrscheinlichkeitsraum. Er repräsentiert das gegenüber (Ω, \mathcal{A}, P) dahingehend veränderte Experiment, dass das Eintreten von A bekannt ist.

Satz 2.2

- (i) Die Multiplikationsregel: Es seien A_1, \dots, A_n Ereignisse mit $P(A_1 \cap \dots \cap A_n) > 0$. Dann gilt
- $$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$
- (ii) Das Gesetz von der totalen Wahrscheinlichkeit: Es sei A_1, \dots, A_n eine Ereignispartition von Ω , d.h.

$$A_1, \dots, A_n \in \mathcal{A}, \quad \bigcup_{i=1}^n A_i = \Omega, \quad A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Dann gilt für alle $B \in \mathcal{A}$

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

(wir lassen hierbei $P(A_i) = 0$ zu und setzen dann $P(B | A_i)P(A_i) = 0$).

- (iii) Die Formel von Bayes: Es seien A_1, \dots, A_n, B wie in (ii) und es gelte $P(B) > 0$. Dann folgt

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{k=1}^n P(B | A_k)P(A_k)}.$$

Beweis

Verwende $B = \sum_{i=1}^n B \cap A_i$ und die Additivität von P bei (ii). Alles andere folgt unmittelbar aus den Definitionen. \square

Beispiel 2.3

Ein bestimmter medizinischer Test ist zu 95% effektiv beim Erkennen einer bestimmten Krankheit, liefert allerdings bei 1% der gesunden Personen einen „falschen Alarm“. Angenommen, 0,5% der Bevölkerung leiden unter dieser Krankheit – mit welcher Wahrscheinlichkeit hat jemand die Krankheit, wenn der Test dies behauptet? Wir schreiben A für das Ereignis, dass die getestete Person die Krankheit hat, B für das Ereignis, dass der Test das Vorliegen der Krankheit anzeigt und übersetzen die obigen Annahmen in

$$P(A) = 0,005, \quad P(B | A) = 0,95, \quad P(B | A^c) = 0,01.$$

Mit der Bayes-Formel ergibt sich dann

$$P(A | B) = \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | A^c)P(A^c)} = \frac{0,95 \cdot 0,005}{0,95 \cdot 0,005 + 0,01 \cdot 0,995} \approx 0,323,$$

ein zumindest auf den ersten Blick überraschend hoher Wert. Man beachte, dass der Übersetzung von Prozentzahlen in Wahrscheinlichkeiten bestimmte Annahmen über die Auswahl der Testpersonen etc. zugrunde liegen.

Beispiel 2.3 zeigt auch, dass es nicht immer nötig bzw. sinnvoll ist, einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) explizit anzugeben.

Einer der zentralen Begriffe der Stochastik ist der der (stochastischen) Unabhängigkeit. Die mathematische Definition soll das intuitive Konzept wiedergeben: B wird von A nicht beeinflusst, wenn sich die Wahrscheinlichkeit von B nicht durch die Information ändert, dass A eingetreten ist. Dies führt auf die Forderung $P(B | A) = P(B)$. Langweilige Fallunterscheidungen (ist $P(A)$ größer als 0?) werden vermieden durch ...

Definition 2.4

Zwei Ereignisse A und B heißen **stochastisch unabhängig**, wenn $P(A \cap B) = P(A)P(B)$ gilt.

Bei mehr als zwei Ereignissen ist Vorsicht angesagt:

Definition 2.5

Eine Familie $\{A_i | i \in I\}$ von Ereignissen heißt **paarweise unabhängig**, wenn gilt:

$$P(A_i \cap A_j) = P(A_i)P(A_j) \text{ für alle } i, j \in I \text{ mit } i \neq j.$$

Sie heißt **unabhängig**, wenn gilt:

$$P\left(\bigcap_{i \in H} A_i\right) = \prod_{i \in H} P(A_i) \text{ für jede endliche Teilmenge } H \text{ von } I.$$

Beispiel 2.6

Eine faire Münze wird zweimal geworfen (siehe Beispiel 1.1 (ii)). Wir haben also ein Laplace-Experiment über $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$. Es sei

- $A_1 := \{(0,0), (0,1)\}$ „Kopf“ im ersten Wurf,
- $A_2 := \{(0,0), (1,0)\}$ „Kopf“ im zweiten Wurf,
- $A_3 := \{(0,1), (1,0)\}$ Resultate verschieden.

Man sieht leicht (die Durchschnitte sind jeweils einelementig)

$$P(A_1 \cap A_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2)$$

und erhält analog

$$P(A_1 \cap A_3) = P(A_1)P(A_3), \quad P(A_2 \cap A_3) = P(A_2)P(A_3).$$

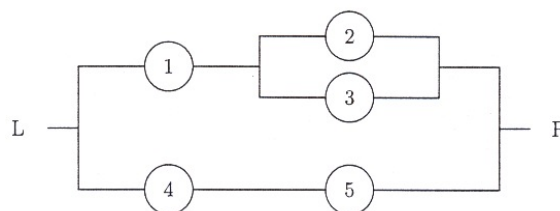
Die Familie $\{A_1, A_2, A_3\}$ ist also paarweise unabhängig. Es gilt jedoch

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq P(A_1)P(A_2)P(A_3),$$

die Familie ist also nicht unabhängig. Moral: paarweise Unabhängigkeit impliziert nicht die (volle) Unabhängigkeit.

Beispiel 2.7

Eine typische Fragestellung der Angewandten Wahrscheinlichkeitsrechnung bezieht sich auf das Funktionieren von Netzwerken. Wir betrachten einen einfachen Fall, in dem ein System aus fünf wie folgt angeordneten Komponenten besteht:



Wir nehmen an, dass die Komponenten unabhängig voneinander und zwar jeweils mit Wahrscheinlichkeit p funktionieren. Das Gesamtsystem funktioniert, wenn es einen Pfad funktionierender Komponenten vom Eingang zum Ausgang gibt. Mit welcher Wahrscheinlichkeit funktioniert das Gesamtsystem?

Es sei A_i das Ereignis, dass Komponente i funktioniert, B das interessierende Ereignis. Dann gilt $B = B_1 \cup B_2$ mit

$$B_1 := A_4 \cap A_5 \quad \text{unterer Pfad passierbar,}$$

$$B_2 := A_1 \cap (A_2 \cup A_3) \quad \text{oberer Pfad passierbar.}$$

Mit Hilfe der Unabhängigkeit und der Formel $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ aus Satz 1.6 erhalten wir

$$P(B_1) = P(A_4)P(A_5) = p^2,$$

$$P(B_2) = P((A_1 \cap A_2) \cup (A_1 \cap A_3)) = P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3) = 2p^2 - p^3,$$

$$P(B_1 \cap B_2) = P(A_4 \cap A_5 \cap A_1 \cap A_2) + P(A_4 \cap A_5 \cap A_1 \cap A_3) - P(A_4 \cap A_5 \cap A_1 \cap A_2 \cap A_3) = 2p^4 - p^5$$

(man könnte auch „ B_1, B_2 unabhängig“ verwenden – allerdings erfordert dies eine abstrakte Zusatzüberlegung), also insgesamt

$$P(B) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = p^2 + 2p^2 - p^3 - (2p^4 - p^5) = p^2(3 - p - 2p^2 + p^3).$$

Man beachte, dass paarweise Unabhängigkeit hier nicht gereicht hätte.

Beispiel 2.8 (Simpson’s paradox)

Das Rechnen mit bedingten Wahrscheinlichkeiten kann gelegentlich in als paradox empfundenen Situationen eine einfache Lösung oder Erklärung liefern. Siehe auch das in den Übungen besprochene „Ziegenproblem“. Ein klassisches Beispiel für das, worum es uns hier geht, liefern die Zulassungszahlen einer amerikanischen Universität aus dem Jahr 1973: Von 1576 männlichen Bewerbern wurden etwa 58% angenommen, von 526 weiblichen Bewerbern nur etwa 46% (aus Zeitgründen betrachten wir nur einen Teil der Daten). Dies wurde damals als Beleg für die Diskriminierung von Frauen angesehen. Die Aufschlüsselung nach Fächern sah wie folgt aus:

| Fach | Männer | | Frauen | |
|-------|------------|------------|------------|------------|
| | # Bewerber | zugelassen | # Bewerber | zugelassen |
| 1 | 825 | 511 (62%) | 108 | 82 (89%) |
| 2 | 560 | 352 (63%) | 25 | 17 (68%) |
| 3 | 191 | 53 (28%) | 393 | 134 (34%) |
| Summe | 1576 | 916 (58%) | 526 | 240 (46%) |

Berücksichtigt man also den Faktor „Fach“, so ergibt sich ein ganz anderes Bild – offensichtlich bewerben sich Frauen eher in Fächern mit einer höheren Ablehnungsquote.

Was hat dies mit bedingten Wahrscheinlichkeiten zu tun? Wie im Beispiel 2.3 werden Häufigkeiten und Wahrscheinlichkeiten dadurch in Zusammenhang gebracht, dass man die zufällige Auswahl einer Person aus der Grundpopulation der 1576 + 526 Bewerber, also ein Laplace-Experiment über $\{1, 2, \dots, 2102\}$ betrachtet. Es seien

- S_k : die ausgewählte Person hat sich für Studiengang k beworben,
- Z : die ausgewählte Person wird zugelassen,
- F, M : die ausgewählte Person ist eine Frau bzw. ein Mann.

Es gilt dann beispielsweise $P(S_1 | M) = \frac{825}{1576}$. Die oben eingeführten Rechenregeln liefern

$$P(Z | F) = \sum_{k=1}^3 P(Z | F \cap S_k)P(S_k | F), \quad P(Z | M) = \sum_{k=1}^3 P(Z | M \cap S_k)P(S_k | M).$$

Man landet also bei dem (ziemlich trivialen) Sachverhalt, dass durchaus

$$P(Z | F \cap S_k) > P(Z | M \cap S_k) \quad \text{für } k = 1, 2, 3$$

und trotzdem $P(Z | F) < P(Z | M)$ gelten kann, da ja die Gewichte verschieden sein können.

Beispiel 2.8*

Bei einer Game-Show befindet sich hinter einer von drei Türen ein Auto. Kandidat wählt eine Tür, Quizmaster öffnet eine der beiden übrigen Türen bzw. die übrige Tür, hinter der kein Auto steht. Sollte der Kandidat wechseln?

Wir nummerieren die Türen mit 1,2,3, wobei Tür 1 die ursprünglich vom Kandidaten gewählte bezeichnet. Seien:

A_i : Auto steht hinter Tür i ,

B_i : Quizmaster wählt Tür i ,

C : Auto steht weder hinter der ursprünglich, noch hinter der vom Quizmaster gewählten Tür.

Klar: $C = A_2 \cap B_3 + A_3 \cap B_2$. Es gilt:

$$P(C) = P(A_2 \cap B_3) + P(A_3 \cap B_2) = P(B_3 | A_2)P(A_2) + P(B_2 | A_3)P(A_3).$$

Aus der Problemstellung geht hervor:

$$P(B_3 | A_2) = P(B_2 | A_3) = 1, \quad P(A_i) = \frac{1}{3} \text{ für } i = 1, 2, 3, \text{ also } P(C) = \frac{2}{3}.$$

Kapitel 3

Laplace-Experimente

3.1 Allgemeines

Ist Ω eine endliche nichtleere Menge, so beschreibt der Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $\mathcal{A} = \mathfrak{P}(\Omega)$,

$$P(A) = \frac{\#A}{\#\Omega} \text{ für alle } A \subset \Omega$$

das Laplace-Experiment über Ω : Es gibt nur endlich viele mögliche Ergebnisse und alle Elementarereignisse haben dieselbe Wahrscheinlichkeit. Zufallsexperimente dieser Art tauchen auf:

- beim Werfen eines symmetrischen Gegenstands (Münze, Würfel, etc.). Symmetrisch heißt dabei, dass alle Seiten mit derselben Wahrscheinlichkeit oben landen.
- beim Mischen von Karten oder allgemeiner beim Herstellen einer zufälligen Reihenfolge. Gut gemischt bzw. zufällige Reihenfolge heißt dabei, dass alle möglichen Anordnungen dieselbe Wahrscheinlichkeit haben.
- beim Entnehmen einer zufälligen Stichprobe aus einer Grundgesamtheit. Zufällige Entnahme einer Stichprobe vom Umfang k aus einer Grundgesamtheit M von n Gegenständen/Personen o.Ä. heißt dabei, dass alle Teilmengen vom Umfang k von M mit derselben Wahrscheinlichkeit gezogen werden.

Wir betrachten nun Koppelungen von solchen Experimenten: Sind $(\Omega_i, \mathcal{A}_i, P_i)$, $1 \leq i \leq n$, Laplace-Experimente, so können wir das Ausführen aller dieser Einzelexperimente als ein neues Zufallsexperiment betrachten. Ein geeigneter Ergebnisraum ist

$$\Omega := \prod_{i=1}^n \Omega_i = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_i \in \Omega_i \text{ für } i = 1, \dots, n\}.$$

Dies ist wieder eine endliche Menge. Sei $\mathcal{A} = \mathfrak{P}(\Omega)$. Wie sieht ein geeignetes Wahrscheinlichkeitsmaß hierauf aus? Wenn die Experimente unabhängig voneinander ausgeführt werden, dann sollten für alle $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ die Ereignisse A_1, \dots, A_n ,

$$A_i: \text{ im } i\text{-ten Telexperiment erscheint } \omega_i,$$

im Sinne von Abschnitt 2 unabhängig sein und natürlich sollte $P(A_i) = P_i(\{\omega_i\})$ gelten. Wegen

$$A_i = \Omega_1 \times \dots \times \Omega_{i-1} \times \{\omega_i\} \times \Omega_{i+1} \times \dots \times \Omega_n$$

führt dies auf

$$P(\{\omega\}) = P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i) = \prod_{i=1}^n P_i(\{\omega_i\}) = \prod_{i=1}^n \frac{1}{\#\Omega_i} = \frac{1}{\#\Omega}$$

(die letzte Gleichung werden wir im folgenden Absatz näher begründen). Da dies nicht von ω abhängt, erhalten wir

$$P(A) = \frac{\#A}{\#\Omega} \text{ für alle } A \subset \Omega,$$

d.h. man erhält wieder ein Laplace-Experiment. Bei dieser Konstruktion sind Ereignisse, die sich auf verschiedene Telexperimente beziehen, stochastisch unabhängig im Sinne von Definition 2.5.

Beispiel 3.1

Zwei Münzen werden (gleichzeitig) geworfen. Als Modell würde ein Laplace-Experiment über $\Omega = \{\omega_0, \omega_1, \omega_2\}$ mit

- ω_0 : beide Münzen zeigen „Kopf“,
- ω_1 : beide Münzen zeigen „Zahl“,
- ω_2 : Ergebnisse verschieden

auf $P(\omega_2) = \frac{1}{3}$ führen – was der Erfahrung widerspricht. Das „korrekte“ Modell ist (wie bereits in Zusammenhang mit Beispiel 1.1 (ii) verwendet) das Laplace-Experiment über $\{0,1\}^2$ und dies führt auf

$$P(\text{Ergebnisse verschieden}) = \frac{1}{2}.$$

Wichtig hierbei ist Folgendes: Die Bestimmung des korrekten Modells ist kein rein mathematisches Problem, dieser Vorgang involviert die Außenwelt. Wir gehen von unterscheidbaren Münzen aus. Die Elementarteilchenphysik liefert Beispiele, die nicht in diesen Rahmen passen.

3.2 Etwas Kombinatorik, oder: Die Kunst des Zählens

Bei Laplace-Experimenten ist häufig die Anzahl der Elemente einer Menge A ($\#A$) zu bestimmen. In diesem Absatz besprechen wir einige wichtige Formeln für häufig vorkommende Mengen A .

Wir schreiben (wieder) $A \times B = \{(a,b) \mid a \in A, b \in B\}$ für das kartesische Produkt der Mengen A und B ,

$$A^k = \underbrace{A \times \dots \times A}_{k\text{-mal}} = \{(x_1, \dots, x_k) \mid x_i \in A \text{ für } i = 1, \dots, k\}.$$

Zwei elementare Grundregeln:

Regel 1: Gibt es eine bijektive Abbildung von A nach B , so gilt: $\#A = \#B$.

Regel 2: Sind A und B disjunkt, so gilt $\#(A \cup B) = \#A + \#B$.

Hat beispielsweise $C \subset A \times B$ die Eigenschaft

$$\#B_x = n \text{ für alle } x \in A \text{ mit } B_x := \{y \in B \mid (x,y) \in C\},$$

so gilt $\#C = n\#A$. Um dies einzusehen, schreibt man die Menge der Paare als disjunkte Vereinigung der Mengen $\{x\} \times B_x$, $x \in A$, verwendet bei den einzelnen Mengen Regel 1 (mit $y \mapsto (x,y)$) und anschließend die von zwei auf endlich viele Mengen verallgemeinerte Variante von Regel 2. Als Spezialfall (B_x hängt nicht von x ab) erhält man die Formel $\#(A \times B) = \#A \cdot \#B$.

Wir schreiben abkürzend M_n für $\{1, \dots, n\}$ (im Folgenden kann anstelle von M_n eine beliebige Menge mit n Elementen stehen). Die obigen Regeln liefern, zusammen mit der anschließenden Diskussion,

$$\#M_n^k = \#\{(i_1, \dots, i_k) \mid 1 \leq i_j \leq n \text{ für } j = 1, \dots, k\} = n^k.$$

Die Elemente von M_n^k werden gelegentlich **k -Permutationen von M_n mit Wiederholung** genannt. Wir geben zwei typische Anwendungen, bei denen Mengen dieses Typs auftauchen:

- (i) Einer Menge von n Elementen kann man n^k Stichproben vom Umfang k mit Zurücklegen bei Berücksichtigung der Reihenfolge des Ziehens entnehmen. Das Element (i_1, \dots, i_k) von M_n^k steht dabei für die Stichprobe, bei der im l -ten Zug das Element i_l erscheint.
- (ii) Es gibt n^k Möglichkeiten, k verschiedene Objekte auf n mögliche Plätze zu verteilen, wieder bei Berücksichtigung der Reihenfolge (Rollenwechsel der Objekte ist nicht möglich) und mit möglicher Mehrfachbelegung. Hierbei steht $(i_1, \dots, i_k) \in M_n^k$ für die Austeilung, bei der im l -ten Schritt das Objekt mit der Nummer l auf den Platz i_l gelegt wurde.

Ein formaler, aber vielleicht weniger anschaulicher Zugang verwendet die Bezeichnung A^B für die Menge der Funktionen $f: B \rightarrow A$ und führt auf

$$\#(A^B) = (\#A)^{\#B} \text{ für endliche Mengen } A, B.$$

Was passiert, wenn wir nur injektive Funktionen zulassen?

Satz 3.2

Für $1 \leq k \leq n$ gilt

$$\#\{(i_1, \dots, i_k) \in M_n^k \mid i_l \neq i_j \text{ für } l \neq j\} = \frac{n!}{(n-k)!}.$$

Beweis

Es gibt n Möglichkeiten für i_1 , bei gegebenem i_1 bleiben $n-1$ Möglichkeiten für i_2 , bei gegebenen i_1, i_2 bleiben $n-2$ Möglichkeiten für i_3 etc., die gesuchte Anzahl ist also gemäß der oben skizzierten Anwendung der Elementarregeln gleich $n(n-1) \cdot \dots \cdot (n-k+1)$. \square

Als wichtigen Spezialfall dieses Satzes erhält man bei $k = n$, dass es genau $n!$ Permutationen einer Menge mit n Elementen gibt. Die Elemente der Menge aus Satz 3.2 werden auch **k -Permutationen von M_n ohne Wiederholung** genannt. Wir haben wieder zwei hauptsächliche Interpretationen:

- (i) Einer Menge von n Elementen kann man $\frac{n!}{(n-k)!}$ verschiedene Stichproben vom Umfang k ohne Zurücklegen bei Berücksichtigung der Reihenfolge entnehmen.
- (ii) Es gibt $\frac{n!}{(n-k)!}$ verschiedene Möglichkeiten, k Objekte auf n Plätze so zu verteilen, dass keine Mehrfachbesetzungen vorkommen.

Satz 3.3

Für $1 \leq k \leq n$ gilt

$$\#\{(i_1, \dots, i_k) \in M_n^k \mid i_1 < i_2 < \dots < i_k\} = \binom{n}{k}.$$

Beweis

Zu jedem Element dieser Menge gehören genau $k!$ Elemente der Menge aus Satz 3.2, nämlich alle die k -Tupel, die durch Permutationen der Koordinaten aus dem geordneten Tupel hervorgehen. \square

Man nennt die Elemente der Menge aus Satz 3.3 auch **k -Kombinationen von M_n ohne Wiederholung**. Als wichtigen Spezialfall erhalten wir die Aussage, dass eine Menge mit n Elementen $\binom{n}{k}$ Teilmengen vom Umfang k hat – was wiederum zusammen mit der bekannten Formel für die Mächtigkeit der Potenzmenge einer Menge einen Beweis für $\sum_{k=0}^n \binom{n}{k} = 2^n$ liefert. (Wir sehen, dass man Identitäten für Binomialkoeffizienten mit kombinatorischen Überlegungen beweisen kann.)

Wie in den vorangegangenen Fällen haben wir auch hier zwei Standardanwendungen:

- (i) Es gibt $\binom{n}{k}$ Möglichkeiten, aus n verschiedenen Objekten k verschiedene herauszugreifen (Stichproben ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge des Ziehens).
- (ii) Es gibt $\binom{n}{k}$ verschiedene Möglichkeiten, k Objekte ohne Mehrfachbesetzung auf n Plätze zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Satz 3.4

Für alle $k \in \mathbb{N}$ gilt

$$\#\{(i_1, \dots, i_k) \in M_n^k \mid i_1 \leq i_2 \leq \dots \leq i_k\} = \binom{n+k-1}{k}.$$

Beweis

Wir definieren eine bijektive Abbildung ϕ von

$$\{(i_1, \dots, i_k) \in M_n^k \mid i_1 \leq \dots \leq i_k\}$$

nach

$$\{(i_1, \dots, i_k) \in M_{n+k-1}^k \mid i_1 < \dots < i_k\}$$

durch

$$\phi((i_1, \dots, i_k)) = (i_1, i_2 + 1, i_3 + 2, \dots, i_k + k - 1)$$

und verwenden Regel 1 und Satz 3.3. \square

Auch für die Elemente der Menge aus Satz 3.4 gibt es einen Namen, **k -Kombinationen von M_n mit Wiederholung**, sowie zwei klassische Interpretationen:

- (i) Einer Menge von n Elementen kann man $\binom{n+k-1}{k}$ verschiedene Stichproben vom Umfang k entnehmen, wenn zurückgelegt wird und die Ziehungsreihenfolge unbeachtet bleibt.
- (ii) Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, k Objekte mit möglicher Mehrfachbesetzung auf n Plätze zu verteilen, wenn die Verteilungsreihenfolge nicht berücksichtigt wird.

Aus der zweiten Interpretation ergibt sich als Anwendung, dass man eine natürliche Zahl k auf $\binom{n+k-1}{k}$ Weisen als Summe von n nichtnegativen ganzen Zahlen schreiben kann:

$$\#\{(i_1, \dots, i_n) \in \mathbb{N}_0^n \mid i_1 + \dots + i_n = k\} = \binom{n+k-1}{k}.$$

Hierbei ist i_l die Anzahl der Objekte auf Platz l , ein leeres Fach beispielsweise entspricht einem Summanden 0.

Gibt es auch bei Kombinationen eine Definition über Funktionen? Bei den Permutationen sieht man den Zusammenhang zu Funktionen, wenn man (i_1, \dots, i_k) als Tabelle auffasst: i_j ist dann die Nummer der Funktionswertes, der zu dem Argument mit Nummer j gehört. Bei den Kombinationen haben wir nur isotone Tupel zugelassen. Definiert man nun eine Äquivalenzrelation \sim auf A^B durch

$$f \sim g \iff \text{Es gibt } \pi : B \rightarrow B \text{ mit } \pi \text{ bijektiv und } f \circ \pi = g,$$

so entsprechen die Kombinationen mit Wiederholung den Äquivalenzklassen in A^B , die ohne Wiederholung den Äquivalenzklassen im Teilraum der injektiven Funktionen. Dies folgt aus zwei einfachen Überlegungen: Zum einen ist Injektivität in dem Sinn mit \sim verträglich, dass entweder alle Elemente einer Äquivalenzklasse injektiv sind oder keines, zum anderen gibt es bei einer festgelegten Nummerierung der Elemente von A und B stets einen kanonischen Vertreter, nämlich das isotone Element. Satz 3.3 und Satz 3.4 können also auch wie folgt geschrieben werden:

$$\#\left(\{f \in A^B \mid f \text{ injektiv}\} / \sim\right) = \binom{\#B}{\#A}, \quad \#(A^B / \sim) = \binom{\#B + \#A - 1}{\#A}.$$

3.3 Einige typische Probleme

3.3.1 Das Geburtstagsproblem

In einem Raum befinden sich n Personen. Mit welcher Wahrscheinlichkeit haben mindestens zwei dieser Personen am gleichen Tag Geburtstag? Wir machen einige vereinfachende Annahmen: Der 29. Februar wird vernachlässigt, ebenso die Möglichkeit von Zwillingen etc., auch saisonale Schwankungen der Geburtenrate werden nicht berücksichtigt. Dann ist ein Laplace-Experiment über

$$\Omega := \{(i_1, \dots, i_n) \mid 1 \leq i_1, \dots, i_n \leq 365\} = \{1, \dots, 365\}^n$$

plausibel, wobei $i_j = k$ bedeutet, dass Person j am k -ten Tag des Jahres Geburtstag hat. Es geht um

$$A := \{(i_1, \dots, i_n) \in \Omega \mid i_l = i_j \text{ für ein Paar } (l, j) \text{ mit } l \neq j\}.$$

Man hat

$$A^c = \{(i_1, \dots, i_n) \in \Omega \mid i_l \neq i_j \text{ für } l \neq j\}$$

und erhält mit den Formeln aus Abschnitt 3.2

$$P(A) = 1 - \frac{\#A^c}{\#\Omega} = 1 - \frac{365!}{365^n (365 - n)!}.$$

Dies ist eine (in n) steigende Folge, denn beim Übergang von n zu $n + 1$ wird im Nenner ein Faktor $(365 - n)$ durch 365 ersetzt. Ab $n = 23$ gilt $P(A) \geq 0,5$, bei $n = 50$ hat man bereits $P(A) \approx 0,97$.

3.3.2 Ein Bridge-Problem

Beim Kartenspiel Bridge werden 52 Karten an die vier Spieler (Nord, Süd, Ost und West) verteilt. Wir wollen die Wahrscheinlichkeit der Ereignisse

A : einer der Spieler erhält vier Asse,

B : jeder der Spieler erhält ein As

bestimmen. Das Mischen der Karten liefert eine zufällige Reihenfolge,

$$\Omega' = \{(\omega_1, \dots, \omega_{52}) \in \{1, \dots, 52\}^{52} \mid \omega_i \neq \omega_j \text{ für } i \neq j\},$$

Ω' ist also die Menge der Permutationen von $\{1, \dots, 52\}$. Hierbei werden die Karten mit $1, \dots, 52$ durchnummeriert. $\omega_k = j$ bedeutet, dass die k -te Karte im Stapel die Nummer j hat. Alle Elementarereignisse haben dieselbe Wahrscheinlichkeit $\frac{1}{52!}$ (wir können diese Annahme als Definition von „Karten gut gemischt“ betrachten). Die Ereignisse A und B hängen nicht von der Reihenfolge ab, mit der die Karten bei den Spielern ankommen. Man kann also auch mit

$$\Omega := \{(D_1, D_2, D_3, D_4) \mid D_i \subset \{1, \dots, 52\}, \#D_i = 13 \text{ für } i = 1, \dots, 4, D_i \cap D_j = \emptyset \text{ für } i \neq j\}$$

arbeiten. Hierbei ist D_i die Menge der Karten für Spieler i . Die Austeilreihenfolge definiert eine Abbildung von Ω' in Ω , die jeweils $(13!)^4$ verschiedene Elemente von Ω' auf genau ein Element von Ω abbildet (alle $13!$ Permutationen der an Spieler 1 ausgegebenen Karten liefern dieselbe Menge D_1 etc.). Betrachten wir also als Resultat des Zufallsexperiments das Vierer-Tupel der „Hände“, so liegt noch stets ein Laplace-Experiment vor, denn es werden jeweils gleich viele Elemente (bei Beispiel 3.1 war dies anders!) von Ω' zu einem Element von Ω zusammengefasst. Hieraus ergibt sich auch

$$\#\Omega = \frac{\#\Omega'}{(13!)^4} = \frac{52!}{13!13!13!13!}.$$

Man kann dies auch wie folgt einsehen: D_1 ist eine Teilmenge vom Umfang 13 von einer Menge mit 52 Elementen, es gibt also $\binom{52}{13}$ Möglichkeiten für D_1 . D_2 ist eine Teilmenge vom Umfang 13 der Menge $\{1, \dots, 52\} - D_1$, die $52 - 13 = 39$ Elemente hat. Ist also D_1 festgelegt, so bleiben $\binom{39}{13}$ Möglichkeiten für D_2 .

Für D_3 bleiben $\binom{26}{13}$ Möglichkeiten und der vierte Spieler erhält automatisch die übrigen Karten: Anwendung der Regeln aus Abschnitt 3.2 führt also auf

$$\#\Omega = \binom{52}{13} \binom{39}{13} \binom{26}{13} \cdot 1 = \frac{52!}{13!13!13!13!}.$$

Es sei nun A_i das Ereignis, dass Spieler i alle vier Asse erhält (wir können annehmen, dass diese mit $1, \dots, 4$ durchnummeriert sind). Dann gilt

$$A_i = \{(D_1, D_2, D_3, D_4) \in \Omega \mid D_i \supset \{1, \dots, 4\}\}.$$

Für $D_1 \cap \{1, \dots, 4\}^c$ bleiben $\binom{48}{9}$ Möglichkeiten (9 Karten aus der Menge der „Nicht-Asse“). Die Anzahl der Möglichkeiten für D_2, D_3 und D_4 bleibt unverändert, also gilt

$$P(A_i) = \frac{1}{\#\Omega} \binom{48}{9} \binom{39}{13} \binom{26}{13} = \frac{13 \cdot 12 \cdot 11 \cdot 10}{52 \cdot 51 \cdot 50 \cdot 49}.$$

Dieselben Argumente funktionieren bei A_2, A_3, A_4 und führen auf dasselbe Ergebnis. Offensichtlich sind A_1, \dots, A_4 disjunkt und haben Vereinigung A , also ergibt sich

$$P(A) = P(A_1) + \dots + P(A_4) = 4P(A_1) \approx 0,01056,$$

in ungefähr einem von 100 Spielen wird ein Spieler alle Asse erhalten.

Bei der Behandlung von B kann man ganz analog verfahren. Wir kürzen die Argumentation wie folgt ab: Es gibt $4!$ Möglichkeiten, die vier Asse so an die vier Spieler zu verteilen, dass jeder genau ein As erhält (4 Möglichkeiten für das Kreuz-As, 3 für das Pik-As etc.). Sind die Asse verteilt, so bleiben

$$\binom{48}{12} \binom{36}{12} \binom{24}{12} = \frac{48!}{12!12!12!12!}$$

Möglichkeiten für die übrigen Karten. Dies ergibt

$$P(B) = \frac{\#B}{\#\Omega} = \frac{4!13^4}{52 \cdot 51 \cdot 50 \cdot 49} \approx 0,1055,$$

in ungefähr einem von 10 Spielen sind also die Asse gleichmäßig verteilt.

3.3.3 Der zerstreute Postbote

Ein Postbote verteilt n Briefe zufällig auf n Briefkästen, einen pro Kasten. Wir nehmen an, dass zu jeder der n Adressen genau einer der n Briefe gehört. Mit welcher Wahrscheinlichkeit erhält keine Person den für sie bestimmten Brief?

Wir nummerieren Briefe und Briefkästen so, dass Brief i in Kasten i gehört, $1 \leq i \leq n$. „Zufällig“ soll heißen, dass ein Laplace-Experiment über

$$\Omega_n := \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \{1, \dots, n\}, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

vorliegt. Sei zunächst

$$A_n := \{\omega \in \Omega_n \mid \omega_i \neq i \text{ für alle } i = 1, \dots, n\}$$

die Menge der **fixpunktfreien** Permutationen sowie

$$B_{ni} := \{\omega \in \Omega_n \mid \omega_i = i\}, \quad 1 \leq i \leq n.$$

Offensichtlich gilt $A_n^c = \bigcup_{i=1}^n B_{ni}$, also folgt mit der Siebformel (Satz 1.6 (vii))

$$P_n(A_n) = 1 - P_n\left(\bigcup_{i=1}^n B_{ni}\right) = 1 - \sum_{\emptyset \neq H \subset \{1, \dots, n\}} (-1)^{\#H-1} P_n\left(\bigcap_{i \in H} B_{ni}\right).$$

Wir haben

$$\bigcap_{i \in H} B_{ni} = \{\omega \in \Omega_n \mid \omega_i = i \text{ für alle } i \in H\}.$$

Für ein ω aus diesem Durchschnitt sind $\#H$ Positionen festgelegt. Die übrigen $n - \#H$ Positionen können beliebig permutiert werden, also gilt

$$\#\bigcap_{i \in H} B_{ni} = (n - \#H)!.$$

Schließlich ist die Anzahl aller H mit k Elementen gleich $\binom{n}{k}$, also erhalten wir insgesamt

$$P_n(A_n) = 1 - \sum_{\emptyset \neq H \subset \{1, \dots, n\}} \frac{(-1)^{\#H-1} (n - \#H)!}{n!} = 1 - \sum_{k=1}^n \binom{n}{k} (-1)^{k-1} \frac{(n-k)!}{n!} = \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Aus der Analysis ist $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$ bekannt. Für große n ist also die Wahrscheinlichkeit dafür, dass kein Brief beim richtigen Empfänger landet, ungefähr $e^{-1} \approx 0,3679$. Wir haben hier ein erstes Grenzwertresultat. Da es im vorliegenden Fall um eine alternierende Reihe geht, können wir darüber hinaus sogar eine Fehlerabschätzung angeben:

$$|P_n(A_n) - e^{-1}| \leq \frac{1}{(n+1)!}.$$

Gleichzeitig haben wir eine Aussage bewiesen, die nicht auf Wahrscheinlichkeiten Bezug nimmt: Die Anzahl der fixpunktfreien Permutationen einer Menge von n Elementen ist

$$n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

3.3.4 Das Paradox von de Méré *

(i) Ein Würfel wird 4 Mal geworfen. Sei A das Ereignis, dass mindestens eine 6 erscheint.

(ii) Zwei Würfel werden 24 Mal geworfen. Sei B das Ereignis, dass mindestens eine Doppel-6 erscheint.

Da in (i) das Einzelereignis die Wahrscheinlichkeit $\frac{1}{6}$ hat und in (ii) das Sechsfache an Wiederholungen ausgeführt werden, kann man vermuten, dass A und B dieselbe Wahrscheinlichkeit haben.

(i) Sei A_i das Ereignis, dass im i -ten Wurf eine 6 erscheint.

A_1, \dots, A_4 sind unabhängig und haben die Wahrscheinlichkeit $\frac{1}{6}$, also:

$$\begin{aligned} P(A) &= 1 - P(A^c) = 1 - P(A_1^c \cap \dots \cap A_4^c) = 1 - P(A_1^c) \cdot \dots \cdot P(A_4^c) \\ &= 1 - (1 - P(A_1)) \cdot \dots \cdot (1 - P(A_4)) = 1 - \left(1 - \frac{1}{6}\right)^4 \approx 0.5177\dots \end{aligned}$$

(ii) Sei B_i das Ereignis, dass im i -ten Wurf das Paar (6,6) erscheint. B_1, \dots, B_{24} sind unabhängig und haben je die Wahrscheinlichkeit $\frac{1}{36}$, also:

$$P(B) = 1 - P(B^c) = 1 - P(B_1^c \cap \dots \cap B_{24}^c) = 1 - \left(1 - \frac{1}{36}\right)^{24} \approx 0.4914\dots$$

3.4 Ein Blick in die Ferne

Kann man auch bei unendlichem Ergebnisraum von gleich wahrscheinlichen Resultaten sprechen? Bei abzählbar unendlichem Ω wie beispielsweise \mathbb{N} erhält man sofort einen Widerspruch zur σ -Additivität – es gibt in einem Rahmen kein Modell für eine zufällige natürliche Zahl, bei dem alle Elementarereignisse $\{n\}$, $n \in \mathbb{N}$, dieselbe Wahrscheinlichkeit haben. In diesem Unterabschnitt betrachten wir kurz die Situation bei überabzählbarem Ergebnisraum. Bereits der in Beispiel 1.1 (iv) aufgetauchte rotierende Zeiger führt auf diese Fragestellung.

3.4.1 Die Nadel von Buffon

Eine große Fläche wird mit parallelen Linien im Abstand D bedeckt. Eine Nadel der Länge L wird „in zufälliger Weise“ auf diese Fläche geworfen. Mit welcher Wahrscheinlichkeit schneidet die Nadel eine der Linien? Wir setzen einfachheitshalber $L \leq D$ voraus. Das Wurfresultat kann durch ein Paar (x, θ) beschrieben werden, wobei x der Abstand des Nadelzentrums zur nächsten Linie und θ den Winkel zwischen Nadel- und Linienrichtung angibt. Entscheidend ist nun eine Invarianzüberlegung: Drehungen und Verschiebungen sollten keine Rolle spielen, also sollten alle Elemente von

$$\Omega := \{(x, \theta) \mid 0 \leq x \leq \frac{D}{2}, 0 \leq \theta < \pi\}$$

„dieselbe Wahrscheinlichkeit“ haben. Schaut man sich die Formel an, auf die diese Forderung bei endlichem Ergebnisraum führt, so liegt es nahe,

$$P(A) = \frac{\text{Fläche von } A}{\text{Fläche von } \Omega}$$

zu fordern. Bei gegebenem θ schneidet die Nadel genau dann eine der Linien, wenn $x \leq \frac{L}{2} \sin(\theta)$ gilt, das interessierende Ereignis wird also beschrieben durch

$$A = \left\{ (x, \theta) \in \Omega \mid x \leq \frac{L}{2} \sin(\theta) \right\}$$

und man erhält

$$P(A) = \left(\frac{\pi D}{2}\right)^{-1} \int_0^\pi \frac{L}{2} \sin(\theta) d\theta = \frac{2L}{\pi D}.$$

Schätzt man $P(A)$ durch die beobachtete relative Häufigkeit der Linienüberquerungen beim Wurf einer großen Anzahl von Nadeln, so lässt sich auf diese Weise ein (zufälliger) Näherungswert für π bestimmen. Diese Beobachtung hat allerdings bestenfalls didaktischen Wert als Einstieg in die Monte-Carlo-Methode, da selbst die aus der Numerik als praktisch unbrauchbar bekannte Leibniz-Reihe bessere Resultate liefert.

3.4.2 Das Paradox von Bertrand

Mit welcher Wahrscheinlichkeit ist die von einer zufälligen Geraden im Einheitskreis gebildete Sekante länger als $\sqrt{3}$, die Seite eines einbeschriebenen gleichseitigen Dreiecks? Wird von einem zufällig auf dem Kreisrand gewählten Punkt x in einem zufällig aus $[0, 2\pi)$ gewählten Winkel ϕ zur Tangente an den Einheitskreis in x eine Linie gezogen, so ist das interessierende Ereignis äquivalent zu $\phi \in (\frac{\pi}{3}, \frac{2\pi}{3}]$, man erhält also die Antwort $\frac{1}{3}$. Wird stattdessen ein zufälliger Kreisdurchmesser gewählt und auf diesem ein zufälliger Punkt x , so ist die Sekante, die man als Senkrechte zu dem Durchmesser in x erhält, genau dann länger als $\sqrt{3}$, wenn $x \in (-\frac{1}{2}, \frac{1}{2})$ gilt. Mit dieser Argumentation erhält man also die Antwort $\frac{1}{2}$. Welches die richtige Antwort ist (in den Übungen wird noch eine dritte Möglichkeit erarbeitet), hängt davon ab, wie das Zufallsexperiment ausgeführt wird. Invarianzüberlegungen führen auf die Antwort $\frac{1}{2}$. Man sieht, dass man bei überabzählbarem Ergebnisraum mit dem Konzept „gleich wahrscheinlich“ vorsichtig umgehen muss.

3.4.3 You can't always get what you want

In den obigen Beispielen mit überabzählbarem Ergebnisraum haben wir uns nicht um den konkreten Definitionsbereich der Wahrscheinlichkeitsmaße gekümmert – aus gutem Grund, wie wir jetzt sehen werden. Bereits im allereinfachsten Beispiel des rotierenden Zeigers benötigen wir eine Gleichverteilung auf $[0, 1)$, also einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $\Omega = [0, 1)$ und

$$P(x + A) = P(A) \text{ für alle } x \in [0, 1), A \in \mathcal{A}, \tag{3.1}$$

wobei die Addition modulo 1 zu verstehen ist und $x + A := \{x + y \mid y \in A\}$.

Satz 3.5

Ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}([0, 1))$ mit der Eigenschaft (3.1) existiert nicht.

Beweis

(Unter Verwendung des Auswahlaxioms): Auf $[0, 1)$ wird durch

$$x \sim y \iff x - y \in \mathbb{Q}$$

eine Äquivalenzrelation definiert. Das Auswahlaxiom erlaubt es, aus jeder der zugehörigen Äquivalenzklassen ein Element auszuwählen. Sei A die so erhaltene Menge. Da die Äquivalenzklassen disjunkt sind, enthält A von jeder Äquivalenzklasse genau ein Element. Wir behaupten nun:

- (i) $(A + x) \cap (A + y) = \emptyset$ für alle $x, y \in \mathbb{Q} \cap [0, 1)$, $x \neq y$,
- (ii) $\bigcup_{x \in \mathbb{Q} \cap [0, 1)} (x + A) = [0, 1)$.

Zu (i): Angenommen, man hat $a + x = b + y$ mit $x, y \in \mathbb{Q} \cap [0, 1)$, $x \neq y$, und $a, b \in A$. Wegen $x - y \notin \mathbb{Z}$ bedeutet dies $a \neq b$, wegen $a - b \in \mathbb{Q}$ würde A also im Widerspruch zur Konstruktion zwei Elemente aus einer Äquivalenzklasse enthalten.

Zu (ii): Die Richtung „ \subset “ ist klar, da die Addition modulo 1 geschieht. Ist andererseits $z \in [0, 1)$, dann existiert ein $a \in A$ mit $a \sim z$, d.h. $x := a - z \in \mathbb{Q}$ (mit dem „üblichen“ Minus). Ersetzt man ggf. x durch $x + 1$, so erhält man die gewünschte Darstellung von z .

Ist nun P ein Wahrscheinlichkeitsmaß auf $\mathfrak{P}([0, 1))$ mit der Eigenschaft (3.1), so muss P auch der Menge A einen Wert zuordnen. Mit (3.1), (ii) und der σ -Additivität von P (deren Anwendbarkeit (i) benötigt) würde dann

$$\sum_{x \in \mathbb{Q} \cap [0, 1)} P(x + A) = \sum_{x \in \mathbb{Q} \cap [0, 1)} P(A) = 1$$

folgen – dies ist unmöglich. □

Die Potenzmenge ist also zu groß, wir werden uns mit einer kleineren σ -Algebra zufrieden geben müssen. Wir werden dies im übernächsten Abschnitt weiterverfolgen, betrachten aber im folgenden Abschnitt zunächst wieder Wahrscheinlichkeitsräume mit endlichem oder abzählbar unendlichem Ergebnisraum.

Kapitel 4

Diskrete Wahrscheinlichkeitsräume und Zufallsgrößen

4.1 Allgemeines

Wir nennen (Ω, \mathcal{A}, P) einen **diskreten Wahrscheinlichkeitsraum**, wenn Ω eine endliche oder abzählbar unendliche Menge ist und $\mathcal{A} = \mathfrak{P}(\Omega)$ gilt. Aufgrund der σ -Additivität ist P dann durch die zugehörige **Wahrscheinlichkeitsmassenfunktion** (kurz: **Massenfunktion**) p ,

$$p : \Omega \rightarrow \mathbb{R}, \quad p(\omega) := P(\{\omega\}),$$

eindeutig festgelegt:

$$P(A) = \sum_{\omega \in A} p(\omega) \text{ für alle } A \in \mathcal{A}.$$

Dies verallgemeinert die im letzten Abschnitt behandelten Laplace-Experimente, bei denen Ω endlich und p eine konstante Funktion ist. Wie im Falle der Laplace-Experimente lassen sich Produkte von (endlich vielen) diskreten Wahrscheinlichkeitsräumen bilden, diese ergeben wieder diskrete Wahrscheinlichkeitsräume.

Im Beispiel 3.1 erhielten wir Ergebnisse (genauer: Elementarereignisse) mit unterschiedlicher Wahrscheinlichkeit durch Zusammenfassen von unterschiedlich vielen Ergebnissen bei einem Laplace-Experiment. Dies führt auf Abbildungen auf Wahrscheinlichkeitsräumen und damit zu den folgenden wichtigen Begriffsbildungen.

Definition 4.1

Es seien (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum und S eine nichtleere Menge. Dann heißt eine Abbildung $X : \Omega \rightarrow S$ eine S -wertige **diskrete Zufallsgröße**. Im Falle $S = \mathbb{R}$ sprechen wir von **Zufallsvariablen**, bei $S = \mathbb{R}^d$ mit $d > 1$ von **Zufallsvektoren**.

Mit ω ist auch $X(\omega)$ zufällig, triviale Extremfälle ausgenommen. Es wird bei der Behandlung von Zufallsgrößen also nicht darum gehen (können), welchen Wert X annimmt, sondern darum, mit welcher Wahrscheinlichkeit X in einer Teilmenge A von S liegt. Im Folgenden sei $X^{-1}(A) := \{\omega \in \Omega \mid X(\omega) \in A\}$.

Satz 4.2

Es seien (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow S$ eine diskrete Zufallsgröße. Dann wird durch

$$P^X : \mathfrak{P}(S) \rightarrow \mathbb{R}, \quad P^X(A) := P(X^{-1}(A)) \text{ für alle } A \subset S,$$

ein Wahrscheinlichkeitsmaß auf $(S, \mathfrak{P}(S))$ definiert, die **Verteilung** von X .

Beweis

(i) $P^X(S) = P(\{\omega \in \Omega \mid X(\omega) \in S\}) = P(\Omega) = 1.$

(ii) Sind $A_1, A_2, \dots \subset S$ paarweise disjunkt, so sind auch die Mengen $X^{-1}(A_1), X^{-1}(A_2), \dots$ paarweise disjunkt und mit der σ -Additivität von P folgt

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

Dies zeigt, dass P^X σ -additiv ist. □

Als alternative Schreibweise für die Verteilung einer Zufallsgröße verwenden wir auch $\mathcal{L}(X)$ (das \mathcal{L} steht für das englische Wort „Law“) und schreiben häufig $P(X \in A)$ für $P(X^{-1}(A))$.

Beispiel 4.3

Wie oft erscheint „Kopf“ beim fünfmaligen Wurf einer fairen Münze? Das Ausgangsexperiment ist ein Laplace-Experiment über $\Omega = \{0,1\}^5$ (1: Kopf, 0: Wappen). Die Anzahl der „Kopf“-Würfe ist

$$X(\omega) := \omega_1 + \omega_2 + \dots + \omega_5, \quad \omega = (\omega_1, \dots, \omega_5) \in \Omega.$$

Als Bildbereich kommt beispielsweise $S = \{0, \dots, 5\}$ in Frage. Als Wahrscheinlichkeitsmaß auf einer endlichen Menge wird $\mathcal{L}(X)$ wieder durch die zugehörige Massenfunktion beschrieben, wir benötigen also die Werte

$$P(X = k) = P(\{\omega \in \Omega \mid X(\omega) = k\}) = P(X^{-1}(\{k\}))$$

für $k = 0, \dots, 5$. Man erhält

$$P(\{\omega \in \Omega \mid X(\omega) = k\}) = \frac{\#\{\omega \in \Omega \mid X(\omega) = k\}}{\#\Omega} = \frac{\#\{(\omega_1, \dots, \omega_5) \in \{0,1\}^5 \mid \sum_{i=1}^5 \omega_i = k\}}{2^5} = \frac{\binom{5}{k}}{32}$$

für $k = 0, \dots, 5$, denn es gibt $\binom{5}{k}$ Möglichkeiten, die k 1-Werte auf die fünf möglichen Positionen zu verteilen.

Man beachte, dass $\mathcal{L}(X)$ die im Zusammenhang mit X interessierenden Wahrscheinlichkeiten festlegt, keineswegs aber die Zufallsgröße selbst. Bezeichnet beispielsweise Y die Anzahl der „Wappen“-Würfe in der Situation von Beispiel 4.3, so erhält man $\mathcal{L}(Y) = \mathcal{L}(X)$, obwohl offensichtlich X und Y niemals denselben Wert annehmen.

4.2 Einige wichtige diskrete Verteilungen**4.2.1**

Eine diskrete Zufallsvariable X heißt **binomialverteilt** mit Parametern n und p , kurz: $\mathcal{L}(X) = \text{Bin}(n, p)$ oder $X \sim \text{Bin}(n, p)$, wobei $n \in \mathbb{N}$ und $p \in [0, 1]$, wenn

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

gilt. Dies impliziert wegen

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1 \quad (\text{binomische Formel}),$$

dass die Wahrscheinlichkeit für X -Werte außerhalb von $\{0, \dots, n\}$ gleich 0 ist, also $P(X \in \{0, \dots, n\}) = 1$ gilt.

Die Zufallsvariable X aus Beispiel 4.3 ist $\text{Bin}(5, \frac{1}{2})$ -verteilt. In Verallgemeinerung der in diesem Beispiel betrachteten Situation tauchen Binomialverteilungen stets bei Erfolgsanzahlen bei unabhängigen Wiederholungen auf, wenn man „Erfolg“ als das Eintreten eines bestimmten Ereignisses A in einem Einzelexperiment (beispielsweise „Kopf“ beim Münzwurf) interpretiert. Hierbei ist n die Anzahl der Versuchswiederholungen und p die Erfolgswahrscheinlichkeit, d.h. die Wahrscheinlichkeit für das Eintreten von A in einem Einzelexperiment. Zur Begründung bemerken wir, dass jedes Ereignis im Produktexperiment von der Form $B_1 \times \dots \times B_n$ mit k Faktoren gleich A , den übrigen gleich A^c , das auf $X = k$ führt, wegen der vorausgesetzten Unabhängigkeit die Wahrscheinlichkeit $p^k (1-p)^{n-k}$ hat. Es gibt $\binom{n}{k}$ Möglichkeiten, die k A -Faktoren auf die n möglichen Positionen zu verteilen.

Im Falle $n = 1$ spricht man auch von **Bernoulli-Verteilungen**. X nimmt dann mit Wahrscheinlichkeit 1 nur die Werte 0 und 1 an.

4.2.2

Die Zufallsvariable X heißt **Poisson-verteilt** mit Parameter $\lambda > 0$, wenn

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{für alle } k \in \mathbb{N}_0$$

gilt. Diese Verteilung spielt eine wichtige Rolle als Grenzverteilung, sie approximiert beispielsweise Binomialverteilungen $\text{Bin}(n, p)$ bei großem n und kleinem p :

Satz 4.4

Ist $(p_n)_{n \in \mathbb{N}} \subset [0, 1]$ eine Nullfolge mit der Eigenschaft

$$\lim_{n \rightarrow \infty} np_n = \lambda \in (0, \infty),$$

so gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Beweis

Eine einfache Umformung liefert

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \frac{(np_n)^k}{k!} \left(1 - \frac{np_n}{n}\right)^n.$$

Bei festem k ergibt sich mit $n \rightarrow \infty$ für den ersten Faktor der Grenzwert 1, für den zweiten $\frac{\lambda^k}{k!}$. Beim Nenner des letzten Faktors erhält man den Limes 1, beim Zähler verwendet man die Monotonie von $x \mapsto (1 - \frac{x}{n})^n$, $x > 0$, in Verbindung mit einem Einschachtelungsargument und der bekannten Aussage

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x,$$

um den Grenzwert $e^{-\lambda}$ zu erhalten. □

In Worten besagt dieser Satz, dass bei einer großen Anzahl n von Wiederholungen mit kleiner Erfolgswahrscheinlichkeit p die Zahl X der Erfolge näherungsweise Poisson-verteilt ist mit Parameter $\lambda = np$. Diese Verteilung taucht daher häufig im Zusammenhang mit seltenen Ereignissen auf, beispielsweise bei der Anzahl der Druckfehler pro Seite in einem Buch, der Anzahl emittierter Partikel pro Zeiteinheit bei radioaktivem Material, bei der Anzahl der durch Hufschlag ihres Pferdes ums Leben gekommenen Soldaten eines Kavallerieregiments etc. Satz 4.4 ist daher auch als **Gesetz der seltenen Ereignisse** bekannt.

4.2.3

Angenommen, wir werfen einen fairen Würfel solange, bis eine Sechs erscheint. Es sei X die hierfür notwendige Anzahl der Würfe, einschließlich des Wurfes, der die erste Sechs liefert. Offensichtlich gilt $X = n$ (mit $n \in \mathbb{N}$) genau dann, wenn die ersten $n-1$ Versuche keine Sechs ergeben und im n -ten Versuch eine Sechs erscheint. Aufgrund der Unabhängigkeit der Würfe hat dieses Ereignis die Wahrscheinlichkeit

$$\left(1 - \frac{1}{6}\right)^{n-1} \cdot \frac{1}{6}.$$

Wenn allgemeiner X nur Werte aus \mathbb{N} annimmt und

$$P(X = n) = (1-p)^{n-1} p \text{ für alle } n \in \mathbb{N}$$

gilt, dann heißt X **geometrisch verteilt** mit Parameter p , wobei $p \in (0,1)$ ist.

Diese Verteilung tritt also als Verteilung der Anzahl der Versuche auf, wenn man ein Zufallsexperiment solange wiederholt, bis ein bestimmtes Ereignis, das die Wahrscheinlichkeit p hat, eingetreten ist. Wartet man in Verallgemeinerung hiervon auf das r -te Eintreten des Ereignisses, so erhält man eine Zufallsvariable X , die nur die Werte $r, r+1, \dots$ annimmt und für die

$$P(X = n) = \binom{n-1}{r-1} (1-p)^{n-r} p^r \text{ für alle } n \in \mathbb{N}, n \geq r,$$

gilt. Man nennt diese Verteilung die **negative Binomialverteilung** mit Parametern r und p , wobei $r \in \mathbb{N}$ und $0 < p < 1$. In der Literatur wird stattdessen häufig auch die Verteilung der Anzahl der Misserfolge bis zum r -ten Versuch (also von $Y = X - r$) so benannt.

Wir haben hier die explizite Aufgabe des Definitionsbereiches Ω der Zufallsvariablen vermieden. Ergebnisräume der Form $\{0,1\}^{\mathbb{N}}$ (unendlich oft wiederholter Münzwurf) sind überabzählbar, passen also nicht in den gegenwärtigen Rahmen. Alternativ kann man beim Warten auf den ersten Erfolg von der abzählbaren Ergebnismenge $\Omega := \{(0, \dots, 0, 1) \in \{0,1\}^k \mid k \in \mathbb{N}\}$ ausgehen.

4.2.4

Eine Urne enthalte N Kugeln, M weiße und $N - M$ schwarze. Dieser Urne werden n Kugeln ohne Zurücklegen entnommen ($n, M \leq N$), X sei die Anzahl der weißen Kugeln in der „Stichprobe“. Dann gilt, wobei wie üblich $\binom{i}{j} = 0$ für $j > i$ gesetzt wird,

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \text{ für } k = 0, \dots, n,$$

denn es gibt $\binom{M}{k}$ Möglichkeiten für die weißen und $\binom{N-M}{n-k}$ für die schwarzen Kugeln in der Stichprobe und alle $\binom{N}{n}$ möglichen Ziehungen werden als gleich wahrscheinlich vorausgesetzt. Man nennt diese Verteilung die **hypergeometrische Verteilung** mit Parametern n , N und M . Beispielsweise ist in der in Abschnitt 3.3.2 beschriebenen Situation die Anzahl der Asse, die „Nord“ erhält, hypergeometrisch verteilt mit Parame-

tern 13, 52 und 4. Ein anderes populäres Beispiel: Die Wahrscheinlichkeit für k Richtige beim Zahlenlotto „6 aus 49“ ist

$$\frac{\binom{6}{k} \binom{43}{6-k}}{\binom{49}{6}} \text{ für } k = 0, \dots, 6,$$

man erhält eine hypergeometrische Verteilung mit den Parametern 6, 49 und 6.

4.2.5

Es seien (Ω, \mathcal{A}, P) ein Zufallsexperiment und A_1, \dots, A_r eine Ereignispartition (siehe Satz 2.2 (ii)) von Ω , $p_i := P(A_i)$ für $i = 1, \dots, r$. Dieses Experiment werde n -mal unabhängig wiederholt, $X = (X_1, \dots, X_r)$ sei der Zufallsvektor, dessen l -te Komponente zählt, wie oft das Ereignis A_l eingetreten ist. Dann gilt in Verallgemeinerung von 4.2.1

$$P(X = (k_1, \dots, k_r)) = \frac{n!}{k_1! \cdot \dots \cdot k_r!} p_1^{k_1} \cdot \dots \cdot p_r^{k_r}$$

für alle $k_1, \dots, k_r \in \mathbb{N}_0$ mit $k_1 + \dots + k_r = n$. Man nennt diese Verteilung die **Multinomialverteilung** mit Parametern n und $p = (p_1, \dots, p_r)$. Hierbei muss $n \in \mathbb{N}$, $p \in [0, 1]^r$ mit $\sum_{i=1}^r p_i = 1$ erfüllt sein. Zählt man beispielsweise beim n -fachen Wurf eines fairen Würfels, wie häufig die Ergebnisse $1, \dots, 6$ eingetreten sind, so erhält man die Multinomialverteilung mit Parametern n und $p = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$.

4.3 Erwartungswert und Varianz von Zufallsvariablen

In diesem Unterabschnitt sei stets (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine (diskrete) Zufallsvariable.

Definition 4.5

Der **Erwartungswert** von X wird definiert durch

$$EX = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}),$$

vorausgesetzt, die Summe konvergiert absolut, d.h.

$$\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) < \infty.$$

Ist dies nicht der Fall, so sagen wir, dass der Erwartungswert von X **nicht existiert**.

Das folgende Resultat zeigt, dass man die Summation auf den Bildraum verlagern kann.

Satz 4.6

Zusätzlich zu (Ω, \mathcal{A}, P) und X sei $f : \mathbb{R} \rightarrow \mathbb{R}$ gegeben, $Y := f(X)$. Dann ist Y eine diskrete Zufallsvariable und mit p_X, p_Y als zugehörigen Massenfunktionen gilt

$$EX = \sum_{x \in \mathbb{R}} xp_X(x) := \sum_{\substack{x \in \mathbb{R} \\ p_X(x) > 0}} xp_X(x),$$

$$EY = \sum_{y \in \mathbb{R}} yp_Y(y) = \sum_{x \in \mathbb{R}} f(x) p_X(x),$$

vorausgesetzt, die beteiligten Summen konvergieren absolut.

Beweis

Die Mengen $A_x := \{\omega \in \Omega \mid X(\omega) = x\}$, $x \in \text{Bild}(X)$, bilden eine Ereignispartition von Ω . Da absolut konvergente Reihen beliebig umgeordnet werden können, erhalten wir

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{x \in \text{Bild}(X)} \sum_{\omega \in A_x} X(\omega) P(\{\omega\}) = \sum_{x \in \mathbb{R}} x \sum_{\omega \in A_x} P(\{\omega\}) = \sum_{x \in \mathbb{R}} xp(X = x).$$

Y ist offensichtlich wieder eine reellwertige Abbildung auf Ω , also eine (diskrete) Zufallsvariable. Es gilt

$$\begin{aligned} EY &= \sum_{\omega \in \Omega} Y(\omega) P(\{\omega\}) = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}) \\ &= \sum_{\omega \in \text{Bild}(X)} \sum_{\omega \in A_x} f(X(\omega)) P(\{\omega\}) = \sum_{x \in \mathbb{R}} f(x) P(X = x), \end{aligned}$$

denn $f \circ X$ ist auf A_x konstant. □

Wichtige Konsequenz: EX hängt von X nur über die Verteilung von X ab – insbesondere haben Zufallsvariablen mit derselben Verteilung auch denselben Erwartungswert. Für das Verständnis von Erwartungswerten ist vielleicht die folgende Analogie zur Mechanik hilfreich: Platziert man Massen $\pi_1, \pi_2, \pi_3, \dots$ auf die Punkte $x_1, x_2, x_3, \dots \in \mathbb{R}$, so ist $\sum x_i p_i$ mit $p_i := \pi_i / \sum_j \pi_j$ der Schwerpunkt des Gesamtbildes.

Beispiel 4.7

Im Falle $X \sim \text{Bin}(n, p)$ erhalten wir, da das Bild von X aus den Zahlen $0, 1, \dots, n$ besteht,

$$\begin{aligned} EX &= \sum_{k=0}^n k P(X = k) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np. \end{aligned}$$

Definiert man Y durch $Y := X(X-1)$, so ergibt sich ganz analog

$$\begin{aligned} EY &= \sum_{k=2}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!((n-2)-(k-2))!} p^{k-2} (1-p)^{(n-2)-(k-2)} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k (1-p)^{(n-2)-k} = n(n-1)p^2. \end{aligned}$$

Der folgende Satz zeigt, dass der Erwartungswertoperator linear und monoton ist.

Satz 4.8

Es seien X, Y diskrete Zufallsvariablen mit existierendem Erwartungswert und $c \in \mathbb{R}$.

(i) Linearität: Dann existieren auch $E(X+Y)$ sowie $E(cX)$ und es gilt

$$E(X+Y) = EX + EY, \quad E(cX) = cEX.$$

(ii) Monotonie: Gilt $X \leq Y$, also $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$, so folgt $EX \leq EY$.

Beweis

Die Existenz beispielsweise von $E(X+Y)$ ergibt sich leicht mit der Dreiecksungleichung:

$$\begin{aligned} \sum_{\omega \in \Omega} |(X+Y)(\omega)| P(\{\omega\}) &\leq \sum_{\omega \in \Omega} (|X(\omega)| + |Y(\omega)|) P(\{\omega\}) \\ &\leq \sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) + \sum_{\omega \in \Omega} |Y(\omega)| P(\{\omega\}) < \infty. \end{aligned}$$

Nachdem dies geklärt ist, kann man den Erwartungswert der Summe mit im Wesentlichen denselben Schritten einfach nachrechnen:

$$E(X+Y) = \sum_{\omega \in \Omega} (X+Y)(\omega) P(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) P(\{\omega\}) = EX + EY,$$

die anderen Beweisteile können genauso leicht erbracht werden. □

Mit der Monotonie folgt aus $X \leq |X|$ und $-X \leq |X|$ die wichtige Beziehung

$$|EX| \leq E|X|.$$

Der Erwartungswert von X beschreibt die **Lage** der Verteilung von X . Es folgen nun Messzahlen für die **Variabilität** der Verteilung.

Definition 4.9

Das **k -te Moment** einer Zufallsvariable X ist EX^k , vorausgesetzt, es gilt $\sum_x |x|^k P(X=x) < \infty$ (sonst sagen wir, dass das k -te Moment von X nicht existiert). Existiert das zweite Moment zu X , so nennen wir

$$\text{var}(X) := E(X - EX)^2, \quad \sigma(X) := \sqrt{\text{var}(X)}$$

die **Varianz** und die **Standardabweichung** von X .

Die Varianz ist also die mittlere quadratische Abweichung der Zufallsvariable X von ihrem Mittelwert. Durch den Übergang zur Standardabweichung erhält man eine Steuermesszahl in den gleichen Dimensionen wie X . Bei der Berechnung von Varianzen ist die folgende Formel oft hilfreich.

Lemma 4.10

Es gilt:

$$\text{var}(X) = EX^2 - (EX)^2.$$

Beweis

Mit den Rechenregeln aus Satz 4.8 erhält man

$$\begin{aligned} \text{var}(X) &= E(X^2 - 2(EX)X + (EX)^2) \\ &= EX^2 - 2(EX)EX + E((EX)^2) \\ &= EX^2 - (EX)^2, \end{aligned}$$

wobei wir gebraucht haben, dass der Erwartungswert einer konstanten Zufallsvariable gleich dieser Konstanten ist. \square

Beispiel 4.11

(i) Im Falle $X \sim \text{Bin}(n, p)$ gilt nach Beispiel 4.7

$$EX = np, \quad EX(X-1) = n(n-1)p^2,$$

also

$$EX^2 = E(X^2 - X) + EX = EX(X-1) + EX = n^2p^2 - np^2 + np$$

und damit

$$\text{var}(X) = EX^2 - (EX)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p).$$

(ii) Ist X Poisson-verteilt mit Parameter λ (siehe Absatz 4.2.2), so erhält man

$$EX = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

sowie

$$EX(X-1) = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2,$$

also

$$\text{var}(X) = EX(X-1) + EX - (EX)^2 = \lambda.$$

Bei der Poisson-Verteilung stimmen also Erwartungswert und Varianz überein.

Bemerkung 4.12

Ist M eine beliebige Menge und $A \subset M$, so heißt

$$1_A : M \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

die **Indikatorfunktion** zu A . Man kann $A \mapsto 1_A$ als Einbettung der Potenzmenge von M in den Ring der reellwertigen Funktionen auf M betrachten. So wird beispielsweise aus dem Durchschnitt die Multiplikation. Ist (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum und $A \subset \Omega$, so zeigt die Zufallsvariable $X := 1_A$ an, ob das Ereignis A eintritt (Wert 1) oder nicht (Wert 0). Offensichtlich gilt $\mathcal{L}(X) = \text{Bin}(n, p)$ mit $p = P(A)$. Mit dieser Konstruktion sieht man, dass Erwartungswerte Wahrscheinlichkeiten verallgemeinern:

$$E1_A = 0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1) = P(A),$$

d.h. die Wahrscheinlichkeit eines Ereignisses ist gleich dem Erwartungswert der zugehörigen Indikatorfunktion. Mathematisch ergeben sich Erwartungswerte als natürliche Fortsetzung von Wahrscheinlichkeiten, wenn man Ereignisse über ihre Indikatorfunktionen in den Raum der Zufallsvariablen einbettet: Die Additivität des Maßes wird zur Linearität des Erwartungswertes.

4.4 Bedingte Verteilungen und Unabhängigkeit

Sind $X : \Omega \rightarrow S_1$ und $Y : \Omega \rightarrow S_2$ Zufallsgrößen auf einem diskreten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , so ist

$$Z : \Omega \rightarrow S_1 \times S_2, \quad \omega \mapsto (X(\omega), Y(\omega)),$$

eine Zufallsgröße mit Werten in $S_1 \times S_2$. Die Verteilung P^Z von Z nennt man auch die **gemeinsame Verteilung** von X und Y .

Beispiel 4.13

In der Situation von Absatz 3.3.2 (Bridge) sei X die Anzahl der Asse von „Nord“, Y die von „Süd“. Dann ist $Z := (X, Y)$ eine Zufallsgröße mit Werten in $\{0, \dots, 4\} \times \{0, \dots, 4\}$ und die dort eingeführten Techniken führen auf

$$P(Z = (k, l)) = \frac{\binom{4}{k} \binom{48}{13-k} \binom{4-l}{l} \binom{35+k}{13-l} \binom{26}{13}}{\frac{52!}{(13!)^4}}.$$

| | | X | | | | | |
|----------------------|---|------|------|------|-----|----|---------------------|
| | | 0 | 1 | 2 | 3 | 4 | Zeilensummen |
| Y | 0 | 1150 | 2600 | 1950 | 572 | 55 | 6327 |
| | 1 | 2600 | 4225 | 2028 | 286 | 0 | 9139 |
| | 2 | 1950 | 2028 | 468 | 0 | 0 | 4446 |
| | 3 | 572 | 286 | 0 | 0 | 0 | 858 |
| | 4 | 55 | 0 | 0 | 0 | 0 | 55 |
| Spaltensummen | | 6327 | 9139 | 4446 | 858 | 55 | 20825 |

Tabelle der mit 20825 multiplizierten Werte

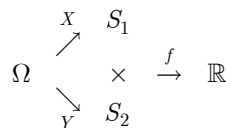
Aus den Werten in der Tabelle ergeben sich wegen

$$P(X = i) = P(X = i, Y = 0) + P(X = i, Y = 1) + \dots + P(X = i, Y = 4)$$

für $i = 0, \dots, 4$ (analog für Y) die **Marginalverteilungen** (oder auch **Randverteilungen**) der Verteilung von Z , also die Verteilungen der Komponenten X und Y von Z . Die gemeinsame Verteilung enthält i.A. mehr Information als die Randverteilungen. Man kann aus der Tabelle die Wahrscheinlichkeit von Ereignissen ablesen, die von X und Y abhängen, beispielsweise

$$\begin{aligned} P(X = Y) &= P(X = 0, Y = 0) + \dots + P(X = 4, Y = 4) \\ &= \frac{1150 + 4225 + 468 + 0 + 0}{20825} \approx 0.280576. \end{aligned}$$

Die gemeinsame Verteilung erlaubt auch eine Verlagerung der Summation bei der Berechnung von Erwartungswerten von Zufallsvariablen der Form $f(X, Y)$. In der im folgenden Diagramm zusammengefassten Situation



erhält man im Stil von Satz 4.6 die für Rechnungen häufig nützliche Formel

$$Ef(X, Y) = \sum_x \sum_y f(x, y) P(X = x, Y = y).$$

Satz 4.14

Mit (Ω, \mathcal{A}, P) , S_1 , S_2 , X und Y wie oben gilt für alle $x \in S_1$ mit $P(X = x) > 0$: Durch

$$A \mapsto P(Y \in A | X = x) \left(= \frac{P(\{\omega \in \Omega | Y(\omega) \in A \wedge X(\omega) = x\})}{P(\{\omega \in \Omega | X(\omega) = x\})} \right)$$

wird ein Wahrscheinlichkeitsmaß auf $(S_2, \mathfrak{P}(S_2))$ definiert, die **bedingte Verteilung** von Y unter X . Schreibweise: $P^{Y|X=x}$ oder $\mathcal{L}(Y | X = x)$.

Im Falle $S_2 = \mathbb{R}$ und $\sum_y |y| P^{Y|X=x}(\{y\}) < \infty$ nennen wir

$$E[Y | X = x] := \sum_{y \in \mathbb{R}} y P^{Y|X=x}(\{y\}) \left(= \frac{1}{P(X = x)} \sum_y y P(Y = y, X = x) \right)$$

den **bedingten Erwartungswert** von Y unter $X = x$.

Für die Verknüpfung der Abbildungen $X : \Omega \rightarrow S_1$ und $x \mapsto P^{Y|X=x}$ bzw. $x \mapsto E[Y | X = x]$ schreiben wir kurz $P^{Y|X}$ oder $\mathcal{L}(Y | X)$ bzw. $E[Y | X]$. Beide Abbildungen sind Zufallsgrößen, die sich von X darstellen lassen.

Beweis

Klar. □

In der Situation von Beispiel 4.13 ergibt sich beispielsweise als bedingte Erwartung der Anzahl der Asse des Partners, wenn man selbst zwei Asse hat,

$$\begin{aligned} E[Y | X = 2] &= 0 \cdot P(Y = 0 | X = 2) + \dots + 4 \cdot P(Y = 4 | X = 2) \\ &= 0 \cdot \frac{1950}{4446} + 1 \cdot \frac{2028}{4446} + 2 \cdot \frac{468}{4446} + 3 \cdot \frac{0}{4446} + 4 \cdot \frac{0}{4446} = \frac{2964}{4446} = \frac{2}{3}. \end{aligned}$$

Als Erwartungswert für Y , also ohne die Zusatzinformation $X = 2$, erhält man den Wert 1 – was man übrigens auch begründen kann, ohne zu rechnen. In den Übungen werden einige Eigenschaften bedingter Erwartungswerte behandelt und es wird gezeigt, dass der bedingte Erwartungswert $E[Y | X]$ die Funktion von X ist, die die Zufallsvariable Y in einem gewissen Sinn optimal vorhersagt.

Beispiel 4.15

Es sei $(\Omega', \mathcal{A}', P')$ das Modell für ein Zufallsexperiment, in dem ein bestimmtes Ereignis A mit Wahrscheinlichkeit $p > 0$ eintritt. Unser Modell für das n -malige unabhängige Wiederholen des Ausgangsexperiments ist (Ω, \mathcal{A}, P) mit $\Omega = (\Omega')^n$, $\mathcal{A} = \mathfrak{P}(\Omega)$ und

$$P(\{(\omega_1, \dots, \omega_n)\}) = P'(\{\omega_1\}) \cdot \dots \cdot P'(\{\omega_n\}).$$

Es sei

$$X : \Omega \rightarrow \mathbb{R}, \omega \mapsto \#\{1 \leq i \leq n \mid \omega_i \in A\}$$

die Anzahl der Einzelexperimente mit Resultat in A ,

$$Y : \Omega \rightarrow \mathfrak{P}(\{1, \dots, n\}), \omega \mapsto \{1 \leq i \leq n \mid \omega_i \in A\}$$

die Menge der Versuchsnummern, in denen A eintritt. Die gemeinsame Verteilung von X und Y ist offensichtlich auf

$$\{(k, B) \mid k \in \{0, \dots, n\}, B \subset \{1, \dots, n\} \text{ mit } \#B = k\}$$

konzentriert und für jedes Element dieser Menge gilt

$$P(X = k, Y = B) = \prod_{j \in B} p \prod_{j \notin B} (1 - p) = p^k (1 - p)^{n-k}.$$

Aus Abschnitt 4.2.1 ist bereits $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ bekannt, also folgt

$$P^{Y|X=k}(\{B\}) = \frac{p^k (1 - p)^{n-k}}{\binom{n}{k} p^k (1 - p)^{n-k}} = \frac{1}{\binom{n}{k}}.$$

Die bedingte Verteilung von Y unter $X = k$ ist also die Gleichverteilung (auch Laplace-Verteilung genannt) auf der Menge der Teilmengen vom Umfang k von $\{1, \dots, n\}$: Alle möglichen Anordnungen für die „Erfolge“ sind gleich wahrscheinlich. In der Statistik wird es sich als wichtig erweisen, dass in dieser bedingten Verteilung der Parameter p nicht auftaucht – im Gegensatz zur Verteilung von Y selbst, gilt doch beispielsweise $P(Y = \{1, \dots, n\}) = p^n$.

Beispiel 4.16*

Auf dem Produktraum (Ω, \mathcal{A}, P) sind die Projektionen $X_i : \Omega \rightarrow \Omega_i$, $X_i((\omega_1, \dots, \omega_n)) = \omega_i$, die Resultate der Einzelexperimente. Für $i < j$ gilt:

$$\begin{aligned} P(X_i = \omega_i, X_j = \omega_j) &= P(\Omega_1 \times \dots \times \Omega_{i-1} \times \{\omega_j\} \times \Omega_{j+1} \times \dots \times \Omega_n) \\ &= P_i(\{\omega_i\}) \cdot P_j(\{\omega_j\}) = P(X_i = \omega_i) \cdot P(X_j = \omega_j). \end{aligned}$$

Hieraus folgt sofort: $P^{X_i|X_j} = P^{X_i}$ und umgekehrt.

Wir dehnen nun den Unabhängigkeitsbegriff auf Zufallsgrößen aus.

Definition 4.16

Für jedes $i \in I$ sei $X_i : \Omega \rightarrow S_i$ eine diskrete Zufallsgröße. Die Familie $\{X_i \mid i \in I\}$ heißt **stochastisch unabhängig**, wenn für jede Wahl von $A_i \subset S_i$, $i \in I$, die Ereignisfamilie $\{X_i^{-1}(A_i) \mid i \in I\}$ stochastisch unabhängig ist im Sinne von Definition 2.5.

Satz 4.17

Eine Familie $\{X_i \mid i \in I\}$ von diskreten Zufallsgrößen ist genau dann unabhängig, wenn für alle $\{i_1, \dots, i_n\} \subset I$, $x_{i_1} \in S_{i_1}, \dots, x_{i_n} \in S_{i_n}$, gilt:

$$P(X_{i_1} = x_{i_1}, \dots, X_{i_n} = x_{i_n}) = P(X_{i_1} = x_{i_1}) \cdot \dots \cdot P(X_{i_n} = x_{i_n}).$$

Beweis

Für beliebige $A_i \subset S_i$ und $\{i_1, \dots, i_n\} \subset I$ gilt

$$\begin{aligned} P\left(\bigcap_{j=1}^n X_{i_j}^{-1}(A_{i_j})\right) &= \sum_{x_{i_1} \in A_{i_1}, \dots, x_{i_n} \in A_{i_n}} P(X_{i_1} = x_{i_1}, \dots, X_{i_n} = x_{i_n}) \\ &= \sum_{x_{i_1} \in A_{i_1}} P(X_{i_1} = x_{i_1}) \cdot \dots \cdot \sum_{x_{i_n} \in A_{i_n}} P(X_{i_n} = x_{i_n}) \\ &= P(X_{i_1} \in A_{i_1}) \cdot \dots \cdot P(X_{i_n} \in A_{i_n}), \end{aligned}$$

also ist die Bedingung hinreichend. Wählt man Elementarereignisse in Definition 4.16, so folgt auch die Notwendigkeit. □

Bei einer Familie X_1, \dots, X_n hat man also Unabhängigkeit genau dann, wenn die gemeinsame Massenfunktion p mit

$$p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n),$$

sich als Produkt der marginalen Massenfunktionen p_i , $p_i(x_i) = P(X_i = x_i)$ für $1 \leq i \leq n$, schreiben lässt, also

$$p(x_1, \dots, x_n) = p_1(x_1) \cdot \dots \cdot p_n(x_n)$$

gilt für alle $x_1 \in S_1, \dots, x_n \in S_n$. Bei Unabhängigkeit ergibt sich daher die gemeinsame Verteilung aus den Randverteilungen. I.A. ist dies nicht der Fall.

4.5 Reellwertige diskrete Zufallsgrößen

Mit \mathbb{R} als Wertebereich hat man zusätzliche Strukturen und damit spezielle Probleme und Konzepte.

Satz 4.18 (Multiplikationsregel für Erwartungswerte)

Sind X und Y unabhängige Zufallsvariablen mit existierenden Erwartungswerten, so existiert auch der Erwartungswert zu $X \cdot Y$ und es gilt $EXY = EXEY$.

Beweis

Die Mengen

$$A_{xy} := \{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}, \quad x \in \text{Bild}(X), \quad y \in \text{Bild}(Y),$$

bilden eine Partition von Ω , also folgt wie im Beweis zu Satz 4.6 (Verlagerung der Summation) unter Ausnutzung der Unabhängigkeit

$$\begin{aligned} \sum_{\omega \in \Omega} |(X \cdot Y)(\omega)| P(\{\omega\}) &= \sum_x \sum_y \sum_{\omega \in A_{x,y}} |(X \cdot Y)(\omega)| P(\{\omega\}) = \sum_x \sum_y |xy| P(X = x, Y = y) \\ &= \sum_x \sum_y |x||y| P(X = x) P(Y = y) = \left(\sum_x |x| P(X = x)\right) \left(\sum_y |y| P(Y = y)\right) \\ &= \left(\sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\})\right) \left(\sum_{\omega \in \Omega} |Y(\omega)| P(\{\omega\})\right). \end{aligned}$$

Wegen der vorausgesetzten Existenz der einzelnen Erwartungswerte ist dies endlich, also existiert auch EXY . Wiederholt man nun die Rechnung ohne Betragsstriche, so erhält man $EXY = EXEY$. □

Im Allgemeinen folgt die Existenz von EXY nicht aus der von EX und EY . Man hat jedoch:

Satz 4.19 (Cauchy-Schwarz-Ungleichung)

Existiert zu den Zufallsvariablen X und Y das zweite Moment, so existiert auch EXY und es gilt

$$(EXY)^2 \leq EX^2EY^2.$$

Beweis

Wegen

$$|(X \cdot Y)(\omega)| = |X(\omega)||Y(\omega)| \leq X(\omega)^2 + Y(\omega)^2 \quad \text{für alle } \omega \in \Omega$$

gilt

$$\sum_{\omega \in \Omega} |(X \cdot Y)(\omega)| P(\{\omega\}) \leq \sum_{\omega \in \Omega} X(\omega)^2 P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega)^2 P(\{\omega\}),$$

also existiert der Erwartungswert zu XY . Für beliebiges $t \in \mathbb{R}$ existiert dann auch das zweite Moment zu $X + tY$ (Satz 4.8) und ist nichtnegativ:

$$0 \leq E(X + tY)^2 = EX^2 + t^2EY^2 + 2tEXY \text{ für alle } t \in \mathbb{R}.$$

Im Falle $EY^2 = 0$ kann die Gerade auf der rechten Seite nur dann oberhalb von 0 bleiben, wenn $EXY = 0$. In diesem Falle gilt also die behauptete Ungleichung. Im Falle $EY^2 > 0$ erhält man als kleinsten Wert der Parabel auf der rechten Seite

$$\frac{1}{EY^2}(EX^2EY^2 - (EXY)^2).$$

Dies ist nur dann nichtnegativ, wenn die behauptete Ungleichung gilt. □

Varianten der Cauchy-Schwarz-Ungleichung tauchen auch in anderen Vorlesungen auf, oft im Zusammenhang mit Begriffen wie Orthogonalität und Projektion. In der folgenden Bemerkung stellen wir die Verbindung her und erhalten gleichzeitig eine geometrische Interpretation bedingter Erwartungswerte. Details sind Gegenstand einer Übungsaufgabe.

Bemerkung 4.20

Ist (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum mit der Eigenschaft

$$P(\{\omega\}) > 0 \text{ für alle } \omega \in \Omega,$$

so ist

$$H := \{X : \Omega \rightarrow \mathbb{R} \mid EX^2 < \infty\} \text{ mit } \langle X, Y \rangle := EXY$$

ein Hilbert-Raum. Mit $\|X\| := \sqrt{\langle X, X \rangle}$ wird die Cauchy-Schwarz-Ungleichung zu

$$|\langle X, Y \rangle| \leq \|X\| \|Y\|.$$

Ist Z eine Zufallsgröße auf diesem Wahrscheinlichkeitsraum und mit Werten in irgendeiner Menge S , so wird durch

$$H(Z) := \{X \in H \mid X = \phi(Z) \text{ für ein } \phi : S \rightarrow \mathbb{R}\}$$

ein Unterraum von H definiert. Die Abbildung

$$H \rightarrow H(Z), X \mapsto E[X \mid Z]$$

ist die Orthogonalprojektion auf diesen Unterraum.

Definition 4.21

Es seien X und Y Zufallsvariablen mit endlichem zweiten Moment und den Standardabweichungen σ_X, σ_Y . Dann heißt

$$\text{cov}(X, Y) := E(X - EX)(Y - EY)$$

die **Kovarianz** von X und Y . Im Falle $\text{cov}(X, Y) = 0$ nennt man X und Y **unkorreliert**. Ist $\sigma_X \cdot \sigma_Y > 0$, so nennt man

$$\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

den **Korrelationskoeffizienten** von X und Y .

Satz 4.22

Es seien X und Y Zufallsvariablen mit existierendem zweiten Moment. Dann gilt:

- (i) $\text{cov}(X, Y) = EXY - EXEY$.
- (ii) Sind X und Y unabhängig, so sind sie auch unkorreliert.
- (iii) Ist $\rho(X, Y)$ definiert, so gilt $-1 \leq \rho(X, Y) \leq 1$.

Beweis

- (i) Mit der Linearität des Erwartungswertoperators (Satz 4.8) folgt

$$\begin{aligned} \text{cov}(X, Y) &= E(XY - (EX)Y - X(EY) + (EX)(EY)) \\ &= EXY - (EX)(EY) - (EX)(EY) + (EX)(EY) = EXY - EXEY. \end{aligned}$$

- (ii) Folgt unmittelbar aus (i) und Satz 4.18.
- (iii) Satz 4.19 liefert

$$\begin{aligned} \text{var}(X) \text{var}(Y) \rho(X, Y)^2 &= (E(X - EX)(Y - EY))^2 \leq E(X - EX)^2 E(Y - EY)^2 \\ &= \text{var}(X) \text{var}(Y). \end{aligned}$$

□

Gemäß Teil (ii) des Satzes sind unabhängige Zufallsvariablen unkorreliert – die Umkehrung hiervon gilt nicht! Kovarianz und Korrelation können als Maß für die lineare Abhängigkeit von Zufallsvariablen betrachtet werden. Auch dies wird in den Übungsaufgaben weiter ausgeführt. Mit Hilfe dieser Begriffe lässt sich etwas über die Varianz einer Summe von Zufallsvariablen aussagen. Die zweite Aussage des folgenden Satzes ist auch als **Gleichheit von Bienaymé** bekannt.

Satz 4.23 (Gleichheit von Bienaymé)

Es seien X_1, \dots, X_n Zufallsvariablen mit existierendem zweitem Moment. Dann gilt

$$\text{var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{var}(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \text{cov}(X_i, X_j).$$

Sind die Zufallsvariablen X_1, \dots, X_n darüber hinaus unabhängig, so gilt

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

Beweis

Unter Verwendung von Satz 4.22 und Lemma 4.10 folgt

$$\begin{aligned} \text{var}\left(\sum_{i=1}^n X_i\right) &= E\left(\sum_{i=1}^n X_i\right)^2 - \left(E\sum_{i=1}^n X_i\right)^2 = \sum_{i,j=1}^n EX_iX_j - \sum_{i,j=1}^n EX_iEX_j \\ &= \sum_{i=1}^n (EX_i^2 - (EX_i)^2) + \sum_{i \neq j} (EX_iX_j - EX_iEX_j) = \sum_{i=1}^n \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j). \end{aligned}$$

Der zweite Teil folgt hieraus sofort mit Satz 4.22 (ii). □

Beispiel 4.24

- (i) In einem Zufallsexperiment sei A ein Ereignis mit der Wahrscheinlichkeit p . Das Experiment werde n -mal unabhängig wiederholt. X_i zeige an, ob das Ereignis in der i -ten Wiederholung eintritt ($X_i = 1$) oder nicht ($X_i = 0$). Dann sind X_1, \dots, X_n unabhängig mit

$$EX_i = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = p, \quad EX_i^2 = EX_i = p, \quad \text{var}(X_i) = p - p^2 = p(1 - p).$$

Somit gilt für $S_n := X_1 + \dots + X_n$

$$ES_n = \sum_{i=1}^n EX_i = np, \quad \text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) = np(1 - p).$$

Wegen $S_n \sim \text{Bin}(n, p)$ ist dies ein alternativer Beweis für die Formeln aus Beispiel 4.11 (i).

- (ii) Es sei X hypergeometrisch verteilt, also

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad \text{für } k = 0, \dots, n.$$

Wir kürzen dies ab zu $X \sim \text{HypGeo}(N; M, n)$. Bei dieser Reihenfolge der Parameter darf man die letzten beiden vertauschen. Wie in Abschnitt 4.2.4 erklärt, entsteht dies als Verteilung der Anzahl der weißen Kugeln, wenn man einer Urne mit N Kugeln eine Stichprobe vom Umfang n entnimmt. Hierbei wird vorausgesetzt, dass M der Kugeln in der Urne weiß sind. Setzt man $X_i = 1$, wenn im i -ten Zug eine weiße Kugel gezogen wird und $X_i = 0$ sonst, so gilt offensichtlich $X = X_1 + \dots + X_n$. Im Gegensatz zu der unter (i) betrachteten Situation sind die Summanden nun allerdings nicht mehr unabhängig, wir benötigen also eine Hilfsüberlegung. Hierzu stellen wir uns die Kugeln als mit den Zahlen 1 bis N nummeriert vor. Sind Y_1, \dots, Y_n die (Nummern der) gezogenen Kugeln, so gilt $X_i = \phi(Y_i)$ mit

$$\phi(i) := \begin{cases} 1, & i\text{-te Kugel weiß,} \\ 0, & \text{sonst} \end{cases}$$

und mit den in Abschnitt 3 besprochenen Techniken erhält man

$$P(Y_1 = i_1, \dots, Y_n = i_n) = \frac{(N - n)!}{N!}$$

für alle n -Permutationen (i_1, \dots, i_n) ohne Wiederholung von $\{1, \dots, N\}$. Es sei S_n die Menge der Permutationen von $\{1, \dots, n\}$. Für beliebiges $\pi \in S_n$ und (i_1, \dots, i_n) wie oben ergibt sich

$$P(Y_{\pi(1)} = i_1, \dots, Y_{\pi(n)} = i_n) = P(Y_1 = i_{\pi^{-1}(1)}, \dots, Y_n = i_{\pi^{-1}(n)}) = \frac{(N - n)!}{N!} = P(Y_1 = i_1, \dots, Y_n = i_n),$$

also gilt $\mathcal{L}((Y_1, \dots, Y_n)) = \mathcal{L}((Y_{\pi(1)}, \dots, Y_{\pi(n)}))$ und damit auch

$$\mathcal{L}((X_1, \dots, X_n)) = \mathcal{L}((X_{\pi(1)}, \dots, X_{\pi(n)})) \text{ für alle } \pi \in S_n$$

(man spricht dann von **vertauschbaren** Zufallsvariablen). Dies impliziert, dass die Verteilung von X_i nicht von i abhängt. Man sieht leicht, dass $X_1 \sim \text{Bin}(1, \frac{M}{N})$ gilt, erhält also

$$EX = \sum_{i=1}^n EX_i = nEX_1 = \frac{nM}{N}.$$

Bei der Varianz argumentiert man analog und benutzt nun, dass $\mathcal{L}((X_i, X_j)) = \mathcal{L}((X_1, X_2))$ für alle i, j mit $i \neq j$ gilt. Wegen $X_1 + X_2 \sim \text{HypGeo}(2; N, M)$ bedeutet dies

$$EX_1X_2 = P(X_1 + X_2 = 2) = \frac{\binom{M}{2} \binom{N-M}{0}}{\binom{N}{2}} = \frac{M(M-1)}{N(N-1)}.$$

Mit Satz 4.23 folgt nun

$$\begin{aligned} \text{var}(X) &= n \text{var}(X_1) + n(n-1) \text{cov}(X_1, X_2) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) + n(n-1) \left(\frac{M(M-1)}{N(N-1)} - \frac{M^2}{N^2}\right) \\ &= \frac{nM(N-n)(N-M)}{N^2(N-1)}. \end{aligned}$$

Beide Formeln kann man natürlich auch im Stil von Beispiel 4.7 „zu Fuß“ erhalten.

Satz 4.25

(i) Es seien P und Q Wahrscheinlichkeitsmaße auf \mathbb{Z} mit Massenfunktionen p und q . Dann ist auch

$$r : \mathbb{Z} \rightarrow \mathbb{R}, r_n := \sum_{k \in \mathbb{Z}} p_k q_{n-k}$$

eine Wahrscheinlichkeitsmassenfunktion. Das zugehörige Wahrscheinlichkeitsmaß R nennen wir die **Faltung** von P und Q . Schreibweise: $R = P \star Q$.

(ii) Sind X und Y unabhängige Zufallsvariablen mit Werten in \mathbb{Z} . so ist auch $X + Y$ eine Zufallsvariable mit Werten in \mathbb{Z} und es gilt $P^{X+Y} = P^X \star P^Y$.

Beweis

(i) Offensichtlich hat man $r_n \geq 0$ für alle $n \in \mathbb{Z}$ sowie

$$\sum_{n \in \mathbb{Z}} r_n = \sum_{n \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} p_k q_{n-k} = \sum_{k \in \mathbb{Z}} p_k \sum_{n \in \mathbb{Z}} q_{n-k} = \sum_{k \in \mathbb{Z}} p_k \cdot 1 = 1,$$

also definiert r ein Wahrscheinlichkeitsmaß auf \mathbb{Z} (durch $R(A) := \sum_{k \in A} r_k$).

(ii) Wir zerlegen nach dem Wert X :

$$\begin{aligned} P(X + Y = n) &= \sum_{k \in \mathbb{Z}} P(X = k, X + Y = n) = \sum_{k \in \mathbb{Z}} P(X = k, Y = n - k) \\ &= \sum_{k \in \mathbb{Z}} P(X = k) P(Y = n - k). \end{aligned}$$

Verwende nun Teil (i) mit $p_k = P(X = k)$, $q_k = P(Y = k)$ und $r_k = P(X + Y = k)$. □

Beispiel 4.26

Es seien X und Y unabhängige Zufallsvariablen. X sei Poisson-verteilt mit Parameter λ und Y sei Poisson-verteilt mit Parameter μ . Dann gilt für alle $n \in \mathbb{N}_0$

$$\begin{aligned} P(X + Y = n) &= \sum_{k \in \mathbb{Z}} P(X = k) P(Y = n - k) = \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}, \end{aligned}$$

$X + Y$ ist also wieder Poisson-verteilt und zwar mit Parameter $\lambda + \mu$. Die Poisson-Verteilungen bilden eine so genannte **Faltungshalbgruppe**.

Was ist die bedingte Verteilung von X unter $X + Y$? Für alle $n \in \mathbb{N}_0$, $k \in \{0, \dots, n\}$, erhält man

$$\begin{aligned} P(X = k | X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} = \frac{P(X = k) P(Y = n - k)}{P(X + Y = n)} \\ &= \frac{e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!}}{e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}} = \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{n-k}, \end{aligned}$$

also gilt $\mathcal{L}(X | X + Y) = \text{Bin}(X + Y, \frac{\lambda}{\lambda + \mu})$. Konkret: Angenommen, ein Buch von 100 Seiten hat auf Seite k X_k Druckfehler, wobei X_1, \dots, X_{100} unabhängig und Poisson-verteilt sind mit Parameter $\lambda > 0$ (diese An-

nahmen sind natürlich bestenfalls näherungsweise erfüllt). Enthält das Buch insgesamt 10 Druckfehler, so ist die bedingte Verteilung der Anzahl der Druckfehler auf der dritten Seite $\text{Bin}(10, \frac{1}{100})$.

4.6 Wahrscheinlichkeitserzeugende Funktionen

Ist $(a_n)_{n \in \mathbb{N}_0}$ eine Folge reeller Zahlen, so nennt man bekanntlich die Potenzreihe

$$\hat{a}(z) := \sum_{n=0}^{\infty} a_n z^n$$

die zugehörige **erzeugende Funktion**. Ist die Folge beschränkt, so darf \hat{a} in einer Nullumgebung beliebig oft gliedweise differenziert werden und man kann dann insbesondere die Folge aus ihrer erzeugenden Funktion zurückerhalten:

$$a_n = \frac{1}{n!} \frac{d^n}{dz^n} \hat{a}(z) \Big|_{z=0}.$$

Manche Probleme, insbesondere die Behandlung von Differenzgleichungen, können durch den Übergang zu erzeugenden Funktionen vereinfacht werden.

Beispiel 4.27 (Ein Ruin-Problem)

Spieler 1 besitzt $n\text{€}$, Spieler 2 $N - n\text{€}$. In jeder Runde gewinnt S1 von S2 1€ mit Wahrscheinlichkeit p und verliert 1€ sonst. Das Spiel wird fortgesetzt, bis einer der Spieler sein gesamtes Geld verloren hat. Mit welcher Wahrscheinlichkeit gewinnt S1 das Spiel?

Sei $N \in \mathbb{N}$ fest, A_n bezeichne das Ereignis, dass S1 bei Anfangskapital n gewinnt, B das Ereignis, dass S1 die erste Runde gewinnt. Das Gesetz von der totalen Wahrscheinlichkeit (Satz 2.2 (ii)) liefert

$$P(A_n) = P(A_n | B)P(B) + P(A_n | B^c)P(B^c) \text{ für } 0 < n < N.$$

Sei $p_n := P(A_n)$. Wir nehmen an, dass die Runden voneinander unabhängig sind und erhalten dann für (p_0, \dots, p_N) die folgende Differenzgleichung zweiter Ordnung mit zwei Randbedingungen:

$$p_n = pp_{n+1} + (1 - p)p_{n-1} \text{ für } 1 \leq n \leq N - 1, \quad p_0 = 0, \quad p_N = 1. \tag{4.1}$$

Mit erzeugenden Funktionen lassen sich solche Gleichungen häufig routinemäßig lösen (oft geht es natürlich auch, wie übrigens auch hier, direkt mit irgendwelchen Tricks oder geschickten Umformungen – diese muss man allerdings erst einmal finden). Sei $r := \frac{1-p}{p}$, wir setzen (zunächst) $r \neq 1$ voraus (also $p \neq \frac{1}{2}$). Löst man (4.1) nach p_{n+1} auf, so erhält man

$$p_{n+1} = (1 + r)p_n - rp_{n-1}.$$

Multiplikation mit z^{n+1} und Summation über $n \in \mathbb{N}$ liefert unter Beachtung von $p_0 = 0$ für

$$\hat{p}(z) = \sum_{n=0}^{\infty} p_n z^n$$

die Beziehung

$$\hat{p}(z) - p_1 z = (1 + r)z\hat{p}(z) - rz^2\hat{p}(z).$$

Löst man dies nach $\hat{p}(z)$ auf und führt man dann eine Partialbruchzerlegung durch, so ergibt sich

$$\hat{p}(z) = \frac{p_1 z}{1 - (1 + r)z + rz^2} = \frac{p_1}{r - 1} \left(\frac{1}{1 - rz} - \frac{1}{1 - z} \right).$$

Erinnert man sich nun an die Formel für die geometrische Reihe, so erhält man hieraus

$$p_n = \frac{p_1}{r - 1} (r^n - 1).$$

Die übrige Randbedingung, $p_N = 1$, führt auf $p_1 = \frac{r-1}{r^N-1}$, also folgt insgesamt

$$p_n = \frac{r^n - 1}{r^N - 1}, \quad n = 0, \dots, N.$$

Ähnlich erhält man bei $r = 1$ das Resultat $p_n = \frac{n}{N}$, $n = 0, \dots, N$. Konkret: Ich betrete ein Kasino mit 100€ Kapital und setze bei Roulette in jeder Runde einen Euro auf Rot. Rot erscheint mit Wahrscheinlichkeit $18/37$ und bringt 2€. Ich höre auf, wenn ich 100€ gewonnen oder aber alles verloren habe. Dies passt in die obige Situation mit $p = 18/37$, $N = 200$ und $n = 100$. Die zugehörige Erfolgswahrscheinlichkeit ist

$$\frac{\left(\frac{19}{18}\right)^{100} - 1}{\left(\frac{19}{18}\right)^{200} - 1} \approx 0.00447.$$

In dieser Situation ist es offensichtlich geschickter, alles auf einen Schlag auf Rot zu setzen, denn dann ist die Erfolgswahrscheinlichkeit gleich $18/37 \approx 0.4865$.

Definition 4.28

Ist X eine \mathbb{N}_0 -wertige Zufallsvariable, so heißt

$$\hat{p}_X(z) := \sum_{k=0}^{\infty} P(X = k) z^k \quad (= Ez^X)$$

die **wahrscheinlichkeitserzeugende Funktion** zu(r Verteilung von) X .

Satz 4.29

- (i) Ist X eine \mathbb{N}_0 -wertige Zufallsvariable mit wahrscheinlichkeitserzeugender Funktion \hat{p} , so gilt für alle $k \in \mathbb{N}$: Das k -te faktorielle Moment $E(X(X-1) \cdots (X-k+1))$ existiert genau dann, wenn $\lim_{z \uparrow 1} \hat{p}^{(k)}(z)$ existiert und dann gilt

$$EX(X-1) \cdots (X-k+1) = \lim_{z \uparrow 1} \hat{p}^{(k)}(z).$$

- (ii) Sind X und Y unabhängige, \mathbb{N}_0 -wertige Zufallsvariablen mit wahrscheinlichkeitserzeugenden Funktionen \hat{p}_X und \hat{p}_Y , so gilt für die wahrscheinlichkeitserzeugende Funktion \hat{p}_{X+Y} zur Summe $X+Y$:

$$\hat{p}_{X+Y}(z) = \hat{p}_X(z) \hat{p}_Y(z) \quad \text{für alle } z \text{ mit } |z| \leq 1.$$

Beweis

- (i) Innerhalb des Konvergenzradius ist die Vertauschung von Summation und Differentiation erlaubt, d.h. es gilt

$$\hat{p}^{(k)}(z) = \sum_{n=k}^{\infty} n(n-1) \cdots (n-k+1) P(X = n) z^{n-k}.$$

Nach dem aus der Analysis bekannten Satz von Abel gilt für Potenzreihen $\sum_{n=0}^{\infty} a_n z^n$ mit nichtnegativen Koeffizienten

$$\lim_{z \uparrow 1} \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n,$$

wobei bestimmte Divergenz zugelassen ist (d.h. genau dann kommt auf der einen Seite ∞ heraus, wenn dies auch für die andere Seite gilt). Schließlich gilt nach der letzten Formel in Satz 4.6

$$EX(X-1) \cdots (X-k+1) = \sum_{n=0}^{\infty} n(n-1) \cdots (n-k+1) P(X = n).$$

- (ii) Es gilt

$$\hat{p}_{X+Y}(z) = Ez^{X+Y} = Ez^X z^Y = Ez^X Ez^Y = \hat{p}_X(z) \hat{p}_Y(z).$$

Hierbei haben wir verwendet, dass bei festem $|z| \leq 1$ mit X und Y auch die Zufallsvariablen z^X und z^Y unabhängig sind (hierzu später mehr) und somit Satz 4.19 angewendet werden kann. \square

Beispiel 4.30

Ist X Poisson-verteilt mit Parameter $\lambda > 0$, so erhält man

$$\hat{p}_X(z) = \sum_{n=0}^{\infty} z^n e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda z)^n = e^{\lambda(z-1)}.$$

Hieraus folgt

$$\hat{p}'_X(z) = \lambda \hat{p}_X(z), \quad \hat{p}''_X(z) = \lambda^2 \hat{p}_X(z),$$

mit Satz 4.29 (i) also

$$EX = \lim_{z \uparrow 1} \lambda e^{\lambda(z-1)} = \lambda, \quad EX(X-1) = \lim_{z \uparrow 1} \lambda^2 e^{\lambda(z-1)} = \lambda^2,$$

in Übereinstimmung mit Beispiel 4.11 (ii). Ist Y eine weitere, von X unabhängige und mit Parameter μ Poisson-verteilte Zufallsvariable, so folgt mit Satz 4.29 (ii)

$$\hat{p}_{X+Y}(z) = \hat{p}_X(z) \hat{p}_Y(z) = e^{\lambda(z-1)} e^{\mu(z-1)} = e^{(\lambda+\mu)(z-1)}.$$

Dies ist die wahrscheinlichkeitserzeugende Funktion zur Poisson-Verteilung mit Parameter $\lambda + \mu$. Da p durch \hat{p} festgelegt ist, muss also die Zufallsvariable $X+Y$ wieder Poisson-verteilt sein, und zwar mit Parameter $\lambda + \mu$. Insgesamt haben wir damit einen alternativen Beweis für einen bereits in Beispiel 4.26 hergeleiteten Sachverhalt.

4.7 Das schwache Gesetz der großen Zahlen

Nach den Resultaten aus Abschnitt 4.5 gilt für den Mittelwert $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ von n unabhängigen Zufallsvariablen X_1, \dots, X_n , die alle den Erwartungswert μ und die Varianz σ^2 haben,

$$E\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \mu = \mu, \quad \text{var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}.$$

(wir haben hier die Rechenregel $\text{var}(\alpha X) = \alpha^2 \text{var}(X)$ benutzt, siehe Aufgabe 48a). Für große n ist also die Verteilung von \bar{X}_n mit kleiner Variabilität um den Mittelwert herum konzentriert. Präzisere Aussagen ermöglichen Ungleichungen vom folgenden Typ.

Satz 4.31

(i) (Die Markovsche Ungleichung) Es sei $p > 0$ und $E|X|^p < \infty$. Dann gilt:

$$P(|X| \geq \alpha) \leq \frac{1}{\alpha^p} E|X|^p \quad \text{für alle } \alpha > 0.$$

(ii) (Die Chebyshevsche Ungleichung) Es sei $EX^2 < \infty$. Dann gilt

$$P(|X - EX| \geq \alpha) \leq \frac{1}{\alpha^2} \text{var}(X) \quad \text{für alle } \alpha > 0.$$

Beweis

(i) Wir definieren eine neue (diskrete) Zufallsvariable Y durch

$$Y(\omega) := \begin{cases} \alpha, & |X(\omega)| \geq \alpha, \\ 0, & |X(\omega)| < \alpha. \end{cases}$$

Offensichtlich gilt $|Y(\omega)|^p \leq |X(\omega)|^p$ für alle $\omega \in \Omega$, die Monotonieeigenschaft des Erwartungswertes (Satz 4.8) liefert also $E|Y|^p \leq E|X|^p$. Da Y nur die beiden Werte α und 0 annimmt, gilt gemäß Satz 4.6

$$E|Y|^p = 0^p P(|X| < \alpha) + \alpha^p P(|X| \geq \alpha).$$

Insgesamt erhält man also $\alpha^p P(|X| \geq \alpha) \leq E|X|^p$.

(ii) Sei $Y = X - EX$. Wir verwenden Teil (i) mit $p = 2$:

$$P(|X - EX| \geq \alpha) = P(|Y| \geq \alpha) \leq \frac{1}{\alpha^2} EY^2 = \frac{1}{\alpha^2} \text{var}(X) \quad \square$$

Der folgende Satz ist eine einfache Version des schwachen Gesetzes der großen Zahlen.

Satz 4.32

Es sei X_1, X_2, \dots eine Folge von paarweise unkorrelierten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 , $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$. Dann gilt

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \rightarrow 0 \quad \text{mit } n \rightarrow \infty \quad \text{für alle } \varepsilon > 0.$$

Beweis

Mit Satz 4.23 erhält man $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$, also folgt mit Chebyshev (Satz 4.31 (ii))

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{var}(\bar{X}_n) \rightarrow 0$$

mit $n \rightarrow \infty$ für jedes feste $\varepsilon > 0$. □

Nimmt man also ein festes $\varepsilon > 0$ (wie klein auch immer), so geht die Wahrscheinlichkeit dafür, dass der Mittelwert der Beobachtungen vom gemeinsamen Erwartungswert um mehr als ε abweicht, mit wachsendem n gegen 0. Ein Spezialfall ist der, bei dem X_i anzeigt, ob im i -ten Experiment ein bestimmtes Ereignis A eingetreten ist. Der obige Satz besagt dann, dass die relative Häufigkeit von A bei n Wiederholungen mit $n \rightarrow \infty$ in einem gewissen Sinn gegen die Wahrscheinlichkeit von A konvergiert: Die Wahrscheinlichkeit dafür, dass relative Häufigkeit und Wahrscheinlichkeit um mehr als ε ($\varepsilon > 0$ fest) voneinander abweichen, wird bei hinreichend großer Anzahl von Versuchswiederholungen beliebig klein. Man kann dieses Resultat als eine (erste) Bestätigung des axiomatischen Aufbaus der Wahrscheinlichkeitstheorie durch die Kolmogorov-Axiome ansehen.

Beispiel 4.33 (Eine Anwendung in der Analysis)

Der Approximationssatz von Weierstraß besagt, dass eine stetige reellwertige Funktion auf einem kompakten Intervall $[a, b] \subset \mathbb{R}$ gleichmäßig durch Polynome approximiert werden kann. Wir wollen diesen Satz mit den Mitteln der Stochastik beweisen – sogar konstruktiv! Wir können $[a, b] = [0, 1]$ annehmen. Sei hierzu

$$p_n : [0, 1] \rightarrow \mathbb{R}, \quad p_n(x) := \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

das n -te **Bernstein-Polynom** zu f . Wir behaupten:

$$\forall \varepsilon > 0 : \exists n_0 \in \mathbb{N} : \forall n \geq n_0, \forall x \in [0, 1] : |f(x) - p_n(x)| \leq \varepsilon \quad (4.2)$$

Sei also $\varepsilon > 0$. Da eine stetige Funktion auf einem kompakten Intervall gleichmäßig stetig ist, existiert ein $\delta = \delta(\varepsilon) > 0$ mit

$$\forall x, y \in [0, 1] : |x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{2}.$$

Außerdem sind stetige Funktionen auf kompakten Intervallen beschränkt, d.h. es gibt ein $K < \infty$ mit $|f(x)| \leq K$ für alle $x \in [0, 1]$. Nach diesen analytischen Vorbereitungen stellen wir nun wie folgt die Verbindung zur Stochastik her: Wähle $x \in [0, 1]$. Wir betrachten den n -fach wiederholten Wurf einer Münze, die mit Wahrscheinlichkeit x das Resultat 1 und sonst 0 liefert. Bezeichnet X_i das Resultat des i -ten Wurfs, so ist $n\bar{X}_n$ die Anzahl der 1-Ereignisse, also $\text{Bin}(n, x)$ -verteilt und es folgt

$$Ef(\bar{X}_n) = \sum_{k=0}^n f\left(\frac{k}{n}\right) P(n\bar{X}_n = k) = p_n(x).$$

Wie im Beweis zu Satz 4.32 erhalten wir

$$P_n(|\bar{X}_n - x| \geq \delta) \leq \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2},$$

denn $x(1-x) \leq \frac{1}{4}$. Wähle nun $n_0 \in \mathbb{N}$ so groß, dass die Ungleichung

$$\frac{2K}{4n_0\delta^2} < \frac{\varepsilon}{2}$$

erfüllt ist. Für alle $n \geq n_0$ gilt dann

$$\begin{aligned} |f(x) - p_n(x)| &= |Ef(\bar{X}_n) - f(x)| \leq E|f(\bar{X}_n) - f(x)| \mathbf{1}_{\{|\bar{X}_n - x| < \delta\}} + E|f(\bar{X}_n) - f(x)| \mathbf{1}_{\{|\bar{X}_n - x| \geq \delta\}} \\ &\leq \frac{\varepsilon}{2} P(|\bar{X}_n - x| < \delta) + 2KP(|\bar{X}_n - x| \geq \delta) < \varepsilon. \end{aligned}$$

Damit ist (4.2) bewiesen. □

Kapitel 5

Allgemeine Wahrscheinlichkeitsräume

5.1 Mengensysteme

In Abschnitt 3.4.3 haben wir gesehen, dass man bei überabzählbarem Ergebnisraum Ω in der Regel nicht mehr allen Teilmengen A von Ω eine Wahrscheinlichkeit zuordnen kann. Der Definitionsbereich von P soll aber häufig zumindest bestimmte Mengen enthalten, beispielsweise die Intervalle im Falle $\Omega = \mathbb{R}$. Wir beschäftigen uns in diesem Unterabschnitt zunächst ganz allgemein mit Mengensystemen.

Definition 5.1

Es sei $\Omega \neq \emptyset$ und $\mathcal{E} \subset \mathfrak{P}(\Omega)$. Dann heißt

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{A} \supset \mathcal{E} \\ \mathcal{A} \text{ } \sigma\text{-Algebra}}} \mathcal{A}$$

die von \mathcal{E} erzeugte σ -Algebra. \mathcal{E} nennt man ein **Erzeugendensystem** zu \mathcal{A} .

In dieser Definition haben wir stillschweigend von der (trivialen) Tatsache Gebrauch gemacht, dass der Durchschnitt von beliebig vielen σ -Algebren über derselben Grundmenge wieder eine σ -Algebra ist. Der obige Durchschnitt ist übrigens nichtleer, denn es gilt $\mathcal{E} \subset \mathfrak{P}(\Omega)$ und $\mathfrak{P}(\Omega)$ ist eine σ -Algebra. Der für uns vorläufig wichtigste Fall ist $\Omega = \mathbb{R}$.

Definition 5.2

Die von den „LORA-Intervallen“ $(a, b]$, $-\infty < a < b < \infty$, erzeugte σ -Algebra heißt die σ -Algebra der **Borel-Mengen von \mathbb{R}** . Schreibweisen: \mathcal{B} , $\mathcal{B}(\mathbb{R})$ oder $\mathcal{B}_{\mathbb{R}}$.

Eine σ -Algebra \mathcal{A} kann durchaus verschiedene Erzeugendensysteme haben, größere Mengensysteme erzeugen größere σ -Algebren und trivialerweise gilt $\sigma(\mathcal{A}) = \mathcal{A}$. Als „general abstract nonsense“ formuliert: Die Abbildung $\mathcal{E} \mapsto \sigma(\mathcal{E})$ ist isoton und idempotent, aber nicht injektiv.

Satz 5.3

Die σ -Algebra $\mathcal{B}(\mathbb{R})$ wird auch erzeugt von den Mengensystemen

$$\mathcal{E}_1 := \{(a, b) \mid -\infty < a < b < \infty\} \text{ (den „ROLA-Intervallen“),}$$

$$\mathcal{E}_2 := \{(-\infty, a] \mid -\infty < a < \infty\},$$

$$\mathcal{E}_3 := \{U \subset \mathbb{R} \mid U \text{ offen}\}.$$

Beweis

Es sei $\mathcal{E} := \{(a, b) \mid -\infty < a < b < \infty\}$ das Erzeugendensystem aus der Definition von \mathcal{B} . Es reicht, jeweils $\mathcal{E}_i \subset \mathcal{B}$ und $\mathcal{E} \subset \sigma(\mathcal{E}_i)$ zu zeigen: Die erste Inklusion impliziert $\sigma(\mathcal{E}_i) \subset \mathcal{B}$, die zweite $\mathcal{B} \subset \sigma(\mathcal{E}_i)$. Hierbei können wir die mengenalgebraischen Abgeschlossenheitseigenschaften von σ -Algebren gegenüber endlichen und abzählbar unendlichen Vereinigungen und Durchschnitts sowie Komplementen verwenden. In diesem Sinne ergibt sich $\sigma(\mathcal{E}_1) = \mathcal{B}$ aus

$$[a, b) = \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \left(a - \frac{1}{n}, b - \frac{1}{m} \right), \quad (a, b) = \bigcup_{n=1}^{\infty} \bigcap_{m=1}^{\infty} \left[a + \frac{1}{n}, b + \frac{1}{m} \right)$$

und $\sigma(\mathcal{E}_2) = \mathcal{B}$ folgt aus

$$(-\infty, a] = \bigcup_{n=1}^{\infty} (a - n, a], \quad (a, b) = (-\infty, b] \cap (-\infty, a]^c.$$

Bei \mathcal{E}_3 verwenden wir, dass es zu jedem x aus einer offenen Menge U ein x enthaltendes Intervall $(a, b] \subset U$ gibt, von dem wir annehmen können, dass die Endpunkte rationale Zahlen sind:

$$U = \bigcup_{\{(a,b) \in \mathbb{Q} \times \mathbb{Q} \mid (a,b] \subset U\}} (a, b].$$

Dies zeigt, dass jede offene Menge $U \subset \mathbb{R}$ als abzählbare Vereinigung von LORA-Intervallen dargestellt werden kann, also $\sigma(\mathcal{E}_3) \subset \mathcal{B}$. Die Gegenrichtung folgt aus der Darstellung

$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n} \right)$$

und der bekannten Tatsache, dass offene Intervalle offene Mengen sind. □

Dieser Satz impliziert, dass die Intervalle $[a, b)$, $(-\infty, a]$ Borel-Mengen sind, ebenso wie alle offenen Mengen. Wegen

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a \right]$$

sind auch alle Einpunktmengen und somit alle abzählbaren Mengen wie beispielsweise \mathbb{Q} Borel-Mengen, damit auch kompakte Intervalle, die irrationalen Zahlen etc.. \mathcal{B} ist für alle praktischen Zwecke reichhaltig genug.

Ist A eine nichtleere Teilmenge von \mathbb{R} , so wird durch

$$\mathcal{B}_A := \{B \cap A \mid B \in \mathcal{B}\}$$

eine σ -Algebra über A definiert (Übungsaufgabe), die **Spur von \mathcal{B} auf A** . Wir nennen \mathcal{B}_A auch das System der Borel-Mengen von A . In der Maßtheorie wird der folgende wichtige Satz bewiesen.

Satz 5.4

Es gibt ein Wahrscheinlichkeitsmaß P auf $([0,1], \mathcal{B}_{[0,1]})$ mit der Eigenschaft

$$P([a,b]) = b - a \text{ für alle } a, b \text{ mit } 0 \leq a < b < 1. \tag{5.1}$$

Bemerkung 5.5

- (i) Man kann zeigen, dass (5.1) auf die Eigenschaft (3.1) aus Abschnitt 3.4.3 führt. Wir werden später sehen, dass (mit $\mathcal{B}_{[0,1]}$ anstelle von \mathcal{A}) auch die Gegenrichtung gilt. Satz 5.4 zeigt also, dass durch eine Verkleinerung des Definitionsbereichs, die für praktische Anwendungen bedeutungslos ist, tatsächlich das in Abschnitt 3.4.3 angesprochene Problem gelöst wird.
- (ii) Man kann P auf $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ fortsetzen durch

$$P_{\mathbb{R}}(B) := P(B \cap [0,1]) \text{ für alle } B \in \mathcal{B}_{\mathbb{R}}.$$

Umgekehrt erhält man aus einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ ein Wahrscheinlichkeitsmaß $P_{[0,1]}$ auf $([0,1], \mathcal{B}_{[0,1]})$ durch

$$P_{[0,1]}(B) := P(B \cap [0,1]),$$

wenn nur $P([0,1]) = 1$ gilt. Das Intervall $[0,1]$ lässt sich hierbei durch ein $A \in \mathcal{B}$ mit $P(A) = 1$ ersetzen. In diesem Sinne nennt man das Wahrscheinlichkeitsmaß P aus Satz 5.4 die **Gleichverteilung auf dem Einheitsintervall**, ohne i.A. zu spezifizieren, ob man $[0,1)$, $(0,1]$, $(0,1)$ oder $[0,1]$ meint, denn wegen

$$P(\{x\}) = \lim_{n \rightarrow \infty} P\left(\left[x, x + \frac{1}{n}\right)\right) = \lim_{n \rightarrow \infty} \left(x + \frac{1}{n} - x\right) = 0$$

spielen die Randpunkte keine Rolle. Man schreibt für P auch $\text{unif}(0,1)$, die „uniforme“ Verteilung. Eine weitere Bezeichnung, deren Sinn später klar werden wird, ist **Rechteckverteilung**.

- (iii) In der Maßtheorie nennt man ein Paar (Ω, \mathcal{A}) , $\Omega \neq \emptyset$ und \mathcal{A} eine σ -Algebra über Ω , einen **messbaren Raum** und eine Abbildung $\mu : \mathcal{A} \rightarrow [0, \infty]$ ein **Maß**, wenn

$$\mu(\emptyset) = 0, \mu\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

für alle paarweise disjunkten $A_1, A_2, \dots \in \mathcal{A}$ gilt. In diesem Sinne sind Wahrscheinlichkeiten ganz einfach normierte Maße. Die geometrische Variante des Problems aus Abschnitt 3.4.3 lautet: Lässt sich allen Teilmengen von \mathbb{R} (oder allgemeiner \mathbb{R}^d) sinnvoll eine Länge (allgemeiner ein Volumen) zuordnen? Es ist wieder eine Einschränkung des Definitionsbereichs nötig und man erhält dann: Es gibt ein Maß ℓ (das **Lebesgue-Maß**) auf $(\mathbb{R}, \mathcal{B})$ mit

$$\ell((a,b]) = b - a \text{ für alle } a < b, a, b \in \mathbb{R}.$$

Man kann $\text{unif}(0,1)$ als Einschränkung von ℓ auf das Einheitsintervall auffassen.

Wir müssen uns nun mit dem Problem der Eindeutigkeit auseinandersetzen – ist beispielsweise $\text{unif}(0,1)$ durch (5.1) eindeutig bestimmt? Hierzu verwenden wir ein auch später sehr nützliches Hilfsmittel.

Definition 5.6

Es sei Ω eine nichtleere Menge. Dann heißt $\mathcal{D} \subset \mathfrak{P}(\Omega)$ ein **Dynkin-System**, wenn gilt

- (i) $\Omega \in \mathcal{D}$,
- (ii) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$,
- (iii) $A_1, A_2, \dots \in \mathcal{D}$ mit $A_i \cap A_j = \emptyset$ für $i \neq j \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$

Im Vergleich zu σ -Algebren wird also die Forderung der Abgeschlossenheit gegenüber beliebigen abzählbaren Vereinigungen auf disjunkte Vereinigungen abgeschwächt. Der Durchschnitt von beliebig vielen Dynkin-Systemen ist offensichtlich wieder ein Dynkin-System, wir können also von

$$\delta(\mathcal{E}) := \bigcap_{\substack{\mathcal{D} \supset \mathcal{E} \\ \mathcal{D} \text{ Dynkin-System}}} \mathcal{D}$$

als dem von \mathcal{E} erzeugten Dynkin-System sprechen.

Dynkin-Systeme sind „fast“ σ -Algebren. Um dies präzisieren zu können, benötigen wir den folgenden Begriff: Wir nennen ein Mengensystem \mathcal{E} **durchschnittsstabil** und schreiben kurz \cap -stabil, wenn gilt

$$A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}.$$

Der folgende Satz zeigt, dass genau diese Eigenschaft den Schritt vom Dynkin-System zur σ -Algebra ermöglicht.

Satz 5.7

- (i) Ein \cap -stabiles Dynkin-System ist eine σ -Algebra.
- (ii) Ist \mathcal{E} \cap -stabil, so gilt $\delta(\mathcal{E}) = \sigma(\mathcal{E})$.

Beweis

- (i) Es seien $A_1, A_2, \dots \in \mathcal{D}$ (nicht notwendigerweise disjunkt). Wir wollen zeigen, dass $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$ gilt und setzen hierzu $B_1 := A_1$,

$$B_n := A_n \cap A_1^c \cap \dots \cap A_{n-1}^c = A_n \setminus (A_1 \cup \dots \cup A_{n-1})$$

für alle $n > 1$. Durchschnittsstabilität und Eigenschaft (ii) liefern $B_n \in \mathcal{D}$ für alle $n \in \mathbb{N}$. Offensichtlich sind die B_n disjunkt, also gilt nach Eigenschaft (iii) $\bigcup_{n=1}^{\infty} B_n \in \mathcal{D}$. Mit

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

folgt nun die gewünschte Aussage (eine ähnliche Konstruktion wurde bereits im Beweis von Satz 1.7 verwendet).

- (ii) Da jede σ -Algebra ein Dynkin-System ist, folgt $\delta(\mathcal{E}) \subset \sigma(\mathcal{E})$ unmittelbar aus den beteiligten Definitionen. Es sei nun für jedes $A \in \delta(\mathcal{E})$

$$\mathcal{D}_A := \{B \subset \Omega \mid B \cap A \in \delta(\mathcal{E})\}.$$

Dann ist \mathcal{D}_A ein Dynkin-System: (i) und (iii) sind trivial, (ii) folgt mit

$$B^c \cap A = (A^c + B \cap A + \Omega^c + \Omega^c + \dots)^c.$$

Da \mathcal{E} \cap -stabil ist, gilt $E' \in \mathcal{D}_E$ für alle $E, E' \in \mathcal{E}$, also $\mathcal{E} \subset \mathcal{D}_E$ und damit $\delta(\mathcal{E}) \subset \mathcal{D}_E$ für alle $E \in \mathcal{E}$, denn \mathcal{D}_E ist ja ein Dynkin-System. Dies heißt

$$D \in \delta(\mathcal{E}), E \in \mathcal{E} \Rightarrow D \cap E \in \delta(\mathcal{E}),$$

also $E \in \mathcal{D}_D$ für alle $E \in \mathcal{E}$, $D \in \delta(\mathcal{E})$. Dies wiederum liefert $\mathcal{E} \subset \mathcal{D}_D$, also $\delta(\mathcal{E}) \subset \mathcal{D}_D$ für alle $D \in \delta(\mathcal{E})$ und damit

$$A \in \delta(\mathcal{E}), D \in \delta(\mathcal{E}) \Rightarrow A \cap D \in \delta(\mathcal{E}).$$

Also ist $\delta(\mathcal{E})$ \cap -stabil und $\delta(\mathcal{E}) \supset \sigma(\mathcal{E})$ folgt mit Teil (i). □

Satz 5.8

Es sei \mathcal{A} eine σ -Algebra mit \cap -stabilem Erzeuger \mathcal{E} . Sind dann P und Q Wahrscheinlichkeitsmaße auf \mathcal{A} mit der Eigenschaft

$$P(E) = Q(E) \text{ für alle } E \in \mathcal{E},$$

so gilt

$$P(A) = Q(A) \text{ für alle } A \in \mathcal{A}.$$

Beweis

Es sei

$$\mathcal{D} := \{A \in \mathcal{A} \mid P(A) = Q(A)\}.$$

Dann gilt $\mathcal{E} \subset \mathcal{D}$ und \mathcal{D} ist, wie man leicht überprüft, ein Dynkin-System. Satz 5.7 (ii) liefert nun

$$\mathcal{D} \supset \delta(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}. \quad \square$$

Stimmen also zwei Wahrscheinlichkeitsmaße auf einem \cap -stabilen Erzeuger überein, so sind sie gleich. Die Mengen $[a, b)$, $0 \leq a < b < 1$, bilden ein Erzeugendensystem von $\mathcal{B}_{[0,1]}$. Dieses ist offensichtlich \cap -stabil. Insbesondere gibt es also nur ein Wahrscheinlichkeitsmaß auf $\mathcal{B}_{[0,1]}$ mit der Eigenschaft (5.1) und wir können von der Gleichverteilung auf dem Einheitsintervall sprechen.

5.2 Zufallsgrößen und Verteilungen

Oft interessiert man sich nicht für das exakte Resultat $\omega \in \Omega$ eines Zufallsexperiments, sondern nur für den Wert $X(\omega)$ einer Funktion X hiervon. Wie im diskreten Fall, geht es dann um die Wahrscheinlichkeit, dass X in einer bestimmten Menge landet. Da unser Wahrscheinlichkeitsmaß nun u.U. nicht mehr auf der gesamten Potenzmenge des Ergebnisraums definiert ist, ist nicht mehr automatisch gewährleistet, dass $P(X \in A)$ überhaupt erlaubt ist. Wir schreiben weiterhin $X \in A$ oder $X^{-1}(A)$ für $\{\omega \in \Omega \mid X(\omega) \in A\}$.

Definition 5.9

Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und (Ω', \mathcal{A}') ein messbarer Raum. Eine Abbildung $X : \Omega \rightarrow \Omega'$ heißt **Zufallsgröße** (auf (Ω, \mathcal{A}, P) und mit Werten in (Ω', \mathcal{A}')), wenn X $(\mathcal{A}, \mathcal{A}')$ -**messbar** ist, d.h. wenn gilt:

$$X^{-1}(A') \in \mathcal{A} \text{ für alle } A' \in \mathcal{A}'.$$

Für eine Zufallsgröße sind also die Wahrscheinlichkeiten dafür, dass ein Wert in einer messbaren Menge des Bildraumes angenommen wird, definiert. Der Begriff Messbarkeit stammt (natürlich) aus der Maßtheorie. Die folgende Analogie zur Topologie ist gelegentlich hilfreich: Auf einer Menge M wird eine Topologie durch das System $\mathcal{U} \subset \mathfrak{P}(U)$ der offenen Mengen beschrieben. Eine Abbildung $f : M \rightarrow M'$ von einem topologischen Raum (M, \mathcal{U}) in einen weiteren topologischen Raum (M', \mathcal{U}') heißt stetig, wenn $f^{-1}(U') \in \mathcal{U}$ für alle $U' \in \mathcal{U}'$ gilt. Also: Messbarkeit heißt, dass die Urbilder messbarer Mengen messbar sind, Stetigkeit heißt, dass die Urbilder offener Mengen offen sind. Natürlich ist im Falle $\mathcal{A} = \mathfrak{P}(\Omega)$ die Bedingung $X^{-1}(A') \in \mathcal{A}$ sogar für alle $A' \in \mathfrak{P}(\Omega')$ erfüllt – dies ist der Grund dafür, dass wir bei diskreten Wahrscheinlichkeitsräumen bisher ohne den Messbarkeitsbegriff ausgekommen sind.

Es ist bekannt, dass Verknüpfungen stetiger Funktionen wieder stetig sind. Der folgende Satz enthält den entsprechenden maßtheoretischen Sachverhalt.

Satz 5.10

Es seien (Ω, \mathcal{A}) , (Ω', \mathcal{A}') , $(\Omega'', \mathcal{A}'')$ messbare Räume sowie $X : \Omega \rightarrow \Omega'$, $Y : \Omega' \rightarrow \Omega''$ $(\mathcal{A}, \mathcal{A}')$ - bzw. $(\mathcal{A}', \mathcal{A}'')$ -messbare Abbildungen. Dann ist $Z := Y \circ X$ $(\mathcal{A}, \mathcal{A}'')$ -messbar.

Beweis

Für alle $A'' \in \mathcal{A}''$ gilt

$$Z^{-1}(A'') = \{\omega \in \Omega \mid Y(X(\omega)) \in A''\} = X^{-1}(\{\omega' \in \Omega' \mid Y(\omega') \in A''\}) = X^{-1}(Y^{-1}(A'')) \in \mathcal{A},$$

denn $A' := Y^{-1}(A'') \in \mathcal{A}'$, $X^{-1}(A') \in \mathcal{A}$ gilt aufgrund der vorausgesetzten Messbarkeiten. □

Beim Nachweis der Messbarkeit kann man sich auf Erzeugendensysteme beschränken:

Satz 5.11

Es seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') messbare Räume und $X : \Omega \rightarrow \Omega'$ eine Abbildung. Ist $\mathcal{E}' \subset \mathfrak{P}(\Omega')$ ein Erzeugendensystem von \mathcal{A}' und gilt

$$X^{-1}(E') \in \mathcal{A} \text{ für alle } E' \in \mathcal{E}',$$

so ist X $(\mathcal{A}, \mathcal{A}')$ -messbar.

Beweis

Es sei $\mathcal{A}_0 := \{A' \subset \Omega' \mid X^{-1}(A') \in \mathcal{A}\}$. Dann ist \mathcal{A}_0 eine σ -Algebra über Ω' : $X^{-1}(\Omega') = \Omega \in \mathcal{A}$, also gilt $\Omega' \in \mathcal{A}_0$. $X^{-1}(A^c) = \{\omega \in \Omega \mid X(\omega) \notin A\} = (\{\omega \in \Omega \mid X(\omega) \in A\})^c = (X^{-1}(A))^c$, also gilt

$$A \in \mathcal{A}_0 \Rightarrow X^{-1}(A) \in \mathcal{A} \Rightarrow (X^{-1}(A))^c \in \mathcal{A} \Rightarrow X^{-1}(A^c) \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}_0.$$

Analog erhält man mit

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n)$$

die dritte definierende Eigenschaft einer σ -Algebra. Nach Voraussetzung gilt $\mathcal{E}' \subset \mathcal{A}_0$, also $\mathcal{A}' = \sigma(\mathcal{E}') \subset \mathcal{A}_0$ und damit $X^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$. □

Schließlich haben wir die folgende Verallgemeinerung von Satz 4.2.

Satz 5.12

Ist X eine (Ω', \mathcal{A}') -wertige Zufallsgröße auf (Ω, \mathcal{A}, P) , so wird durch

$$A' \in \mathcal{A}', A' \mapsto P(X \in A') \quad (= P(\{\omega \in \Omega \mid X(\omega) \in A'\}))$$

ein Wahrscheinlichkeitsmaß auf (Ω', \mathcal{A}') definiert. Dieses Wahrscheinlichkeitsmaß heißt die **Verteilung** von X . Schreibweisen: P^X oder $\mathcal{L}(X)$.

Bei Beachtung der Messbarkeit ist der Beweis identisch zum Beweis im diskreten Fall. In der Sprache der Maßtheorie ist die Verteilung einer Zufallsgröße das durch die messbare Abbildung auf dem Bildraum induzierte Bildmaß.

Beispiel 5.13

Es sei $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}_{[0,1]}, \text{unif}(0,1))$. Für jedes $x \in \Omega$ werde $T_x : \Omega \rightarrow \Omega$ definiert durch

$$T_x(y) := \begin{cases} y - x, & \text{wenn } y \geq x, \\ y - x + 1, & \text{wenn } y < x. \end{cases}$$

Für alle $A \in \mathcal{A}$ gilt dann

$$T_x^{-1}(A) = \{y \in \Omega \mid y - x \in A \text{ oder } y - x + 1 \in A\} = x + A \pmod{1},$$

insbesondere also

$$T_x^{-1}([0,a)) = \begin{cases} [x, x+a), & \text{wenn } x+a \leq 1, \\ [0, x+a-1) \cup [x,1), & \text{wenn } x+a > 1 \end{cases} \in \mathcal{A}.$$

Mit $\sigma(\{[0,a) \mid 0 < a \leq 1\}) = \mathcal{A}$ und Satz 5.11 folgt hieraus die $(\mathcal{A}, \mathcal{A})$ -Messbarkeit von T_x . Man sieht auch, dass

$$P(T_x^{-1}([0,a))) = a = P([0,a))$$

für alle $a \in (0,1]$ gilt, mit Satz 5.8 folgt also $P^{T_x} = P$. Dies wiederum liefert

$$P(x + A) = P(A) \text{ für alle } A \in \mathcal{A},$$

d.h. das Wahrscheinlichkeitsmaß $\text{unif}(0,1)$ hat die Eigenschaft (5.1) (Translationsinvarianz modulo 1).

5.3 Reellwertige Zufallsgrößen

Wie in der in Abschnitt 4 behandelten diskreten Situation verdient der Fall, in dem \mathbb{R} der Wertebereich der Zufallsgrößen ist, besondere Beachtung. Eine reellwertige Zufallsgröße nennen wir auch Zufallsvariable (kurz: ZV). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Als σ -Algebra auf \mathbb{R} werden wir grundsätzlich die σ -Algebra \mathcal{B} der Borel-Mengen nehmen. Aus Satz 5.3 und Satz 5.11 folgt unmittelbar, dass $X : \Omega \rightarrow \mathbb{R}$ genau dann eine Zufallsvariable, also $(\mathcal{A}, \mathcal{B})$ -messbar ist, wenn $X^{-1}((-\infty, a]) \in \mathcal{A}$ für alle $a \in \mathbb{R}$ erfüllt ist. Den einfachsten Fall solcher Abbildungen liefern die Indikatorfunktionen: Wegen

$$1_A^{-1}((-\infty, a]) = \begin{cases} \emptyset, & a < 0, \\ A^c, & 0 \leq a < 1, \\ \Omega, & a \geq 1, \end{cases}$$

ist 1_A genau dann eine Zufallsvariable, wenn $A \in \mathcal{A}$ gilt. Durch den Übergang $A \mapsto 1_A$ werden also die messbaren Mengen in den Raum der messbaren Abbildungen eingebettet.

Häufig werden mit einer Zufallsvariable X Operationen ausgeführt, im Zusammenhang mit der Streuung ist beispielsweise X^2 interessant. Ist X^2 wieder eine Zufallsvariable?

Satz 5.14

Ist $g : \mathbb{R} \rightarrow \mathbb{R}$ stetig oder (schwach) monoton steigend oder fallend, so ist g $(\mathcal{B}, \mathcal{B})$ -messbar.

Beweis

Ist g stetig, so ist $g^{-1}(U)$ für jede offene Menge offen, also in \mathcal{B} . Hieraus folgt die Behauptung mit Satz 5.3 und Satz 5.11. Der Beweis für monotone Funktionen g ist Gegenstand einer Aufgabe. \square

Ist X eine Zufallsvariable, so kann X^2 als Verknüpfung der $(\mathcal{A}, \mathcal{B})$ -messbaren Abbildung X und der $(\mathcal{B}, \mathcal{B})$ -messbaren, weil stetigen, Abbildung $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(x) = x^2$, angesehen werden, ist nach Satz 5.10 also $(\mathcal{A}, \mathcal{B})$ -messbar und damit wieder eine Zufallsvariable. Wird eine neue Abbildung aus mehreren Zufallsvariablen zusammengesetzt, so lässt sich häufig der folgende Satz anwenden.

Satz 5.15

(i) Sind X und Y Zufallsvariablen auf (Ω, \mathcal{A}, P) , so liegen die Mengen $\{X < Y\}$, $\{X \leq Y\}$, $\{X = Y\}$ und $\{X \neq Y\}$ in \mathcal{A} (hierbei steht $\{X < Y\}$ für die Menge $\{\omega \in \Omega \mid X(\omega) < Y(\omega)\}$ etc.).

(ii) Sind X, Y Zufallsvariablen auf (Ω, \mathcal{A}, P) und $\alpha, \beta \in \mathbb{R}$, so sind auch

$$\alpha X + \beta, X + Y, X \cdot Y, X \wedge Y, X \vee Y$$

Zufallsvariablen. ($a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$)

(iii) Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen auf (Ω, \mathcal{A}, P) , so sind auch

$$\sup_{n \in \mathbb{N}} X_n, \inf_{n \in \mathbb{N}} X_n, \limsup_{n \rightarrow \infty} X_n, \liminf_{n \rightarrow \infty} X_n$$

Zufallsvariablen (vorausgesetzt, diese Größen sind \mathbb{R} -wertig). Gilt $X_n(\omega) \rightarrow X(\omega)$ für $\omega \in \Omega$, so ist auch X eine Zufallsvariable.

Beweis

(i) Durch $\{X < Y\} = \bigcup_{q \in \mathbb{Q}} \{X < q\} \cap \{Y > q\}$ wird die Menge $\{X < Y\}$ als zugelassene Kombination messbarer Mengen dargestellt. Wegen $\{X \leq Y\} = \{Y < X\}^c$, $\{X = Y\} = \{X \leq Y\} \cap \{X < Y\}^c$, $\{X \neq Y\} = \{X = Y\}^c$ liegen dann auch die anderen Mengen in \mathcal{A} .

(ii) Die Abbildung $x \rightarrow \alpha x + \beta$ ist stetig, also ist $\alpha X + \beta$ als Verknüpfung messbarer Abbildungen messbar (siehe auch das obige Argument für X^2). Weiter erhält man mit dem bereits bewiesenen Teil (i)

$$\{X + Y \leq a\} = \{X \leq a - Y\} \in \mathcal{A} \text{ für alle } a \in \mathbb{R},$$

denn $a - Y$ ist eine Zufallsvariable, folglich ist $X + Y$ messbar. Mit

$$X \cdot Y = \frac{1}{4}((X + Y)^2 - (X - Y)^2)$$

folgt dann auch die Messbarkeit von $X \cdot Y$, mit

$$\{X \vee Y \leq a\} = \{X \leq a\} \cap \{Y \leq a\}, \{X \wedge Y \leq a\} = \{X \leq a\} \cup \{Y \leq a\}$$

die von $X \vee Y$ und $X \wedge Y$ (hierbei haben wir wiederholt verwendet, dass X $(\mathcal{A}, \mathcal{B})$ -messbar ist, wenn $\{X \leq a\} \in \mathcal{A}$ gilt für alle $a \in \mathbb{R}$).

(iii) Ähnlich wie bei Teil (ii) erhält man

$$\{\sup_{n \in \mathbb{N}} X_n \leq a\} = \bigcap_{n=1}^{\infty} \{X_n \leq a\} \in \mathcal{A}.$$

Die Messbarkeit der anderen Abbildungen ergibt sich nun mit

$$\inf_{n \in \mathbb{N}} X_n = -\sup_{n \in \mathbb{N}} (-X_n), \limsup_{n \rightarrow \infty} X_n = \inf_{n \in \mathbb{N}} \sup_{m \geq n} X_m, \liminf_{n \rightarrow \infty} X_n = \sup_{n \in \mathbb{N}} \inf_{m \geq n} X_m.$$

Konvergiert X_n mit $n \rightarrow \infty$ punktweise gegen X , so gilt $X = \limsup_{n \rightarrow \infty} X_n$, also ist X eine Zufallsvariable. \square

Im Teil (iii) lässt sich die Einschränkung auf \mathbb{R} -wertige Abbildungen beseitigen, wenn man \mathbb{R} zu

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\} (= [-\infty, \infty])$$

erweitert und auch \mathcal{B} passend ergänzt zu $\mathcal{B}(\overline{\mathbb{R}}) := \sigma(\mathcal{B} \cup \{\{-\infty\}, \{\infty\}\})$.

5.4 Verteilungsfunktionen

Die Verteilung einer reellwertigen Zufallsgröße ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$, also eine Abbildung von \mathcal{B} nach $[0,1]$. Wir wollen nun zeigen, dass sich solche Wahrscheinlichkeitsmaße durch Abbildungen von \mathbb{R} nach $[0,1]$ beschreiben lassen.

Definition 5.16

Die **Verteilungsfunktion** F zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B})$ wird definiert durch

$$F : \mathbb{R} \rightarrow \mathbb{R}, \quad F(x) := P((-\infty, x]) \text{ für alle } x \in \mathbb{R}.$$

Ist P die Verteilung einer Zufallsvariable X , so nennen wir F auch die Verteilungsfunktion zu X .

Da die Mengen $(-\infty, x]$, $x \in \mathbb{R}$, ein \cap -stabiles Erzeugendensystem von \mathcal{B} bilden (Satz 5.3), wird P durch das zugehörige F eindeutig festgelegt (Satz 5.8).

Satz 5.17

Ist F die Verteilungsfunktion zu einem Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B})$, so hat F die folgenden Eigenschaften:

- (i) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow \infty} F(x) = 1$,
- (ii) F ist (schwach) monoton steigend,
- (iii) F ist stetig von rechts.

Beweis

- (i) Sei $(x_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ mit $\lim_{n \rightarrow \infty} x_n = -\infty$ (d.h. $\forall c > 0 : \exists n_0 \in \mathbb{N} : \forall n \geq n_0 : x_n \leq c$). Setze

$$y_n := \sup_{m \geq n} x_m.$$

Dann gilt $y_n \downarrow -\infty$, also $(-\infty, y_n] \downarrow \emptyset$ und es folgt mit der Stetigkeit von P in \emptyset (Satz 1.7 (iv))

$$0 \leq F(x_n) = P((-\infty, x_n]) \leq P((-\infty, y_n]) \rightarrow 0$$

mit $n \rightarrow \infty$. Die andere Aussage erhält man analog mit der Stetigkeit von P von unten (in \mathbb{R} , Satz 1.7 (ii)).

- (ii) folgt unmittelbar aus der Monotonie von P (siehe Satz 1.6 (iv)).
- (iii) Ist $(x_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ mit $x_n \geq x$ für alle $n \in \mathbb{N}$ und $x_n \rightarrow x$, so gilt $y_n \downarrow x$ für $y_n := \sup_{m \geq n} x_m$, also

$$F(x) = P((-\infty, x]) \leq P((-\infty, x_n]) = F(x_n) \leq P((-\infty, y_n]) \rightarrow P((-\infty, x]) = F(x),$$

wobei wir wieder eine Stetigkeitseigenschaft von P verwendet haben. □

Wir wollen nun zeigen, dass die obige Liste vollständig ist, d.h. dass zu jeder Funktion F mit den Eigenschaften (i)-(iii) ein Wahrscheinlichkeitsmaß P existiert, dessen Verteilungsfunktion F ist.

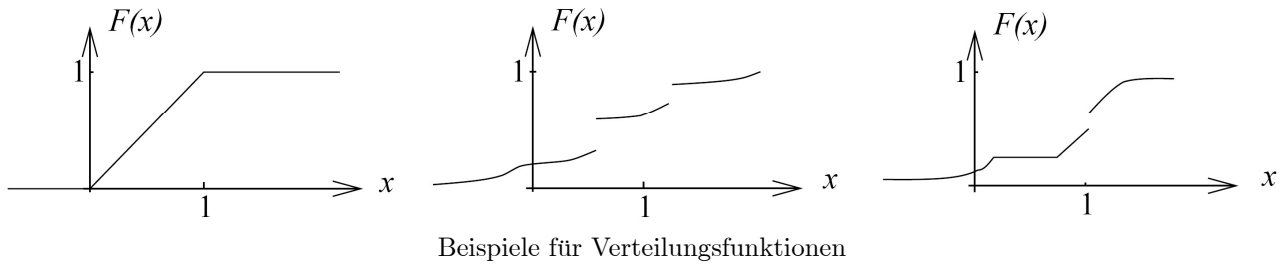
Definition 5.18

Sei F eine Funktion mit den Eigenschaften (i)-(iii) aus Satz 5.17. Dann definieren wir die **Quantilfunktion** Q zu F durch

$$Q : (0,1) \rightarrow \mathbb{R}, \quad Q(y) := \inf \{x \in \mathbb{R} \mid F(x) \geq y\}.$$

Wir schreiben auch F^{-1} für die Quantilfunktion zu F .

Ist X eine Zufallsvariable mit Verteilungsfunktion F , so nennt man $F^{-1}(\alpha)$, $0 < \alpha < 1$, das α -**Quantil** zu X (bzw. $\mathcal{L}(X)$ oder F). Es ist dies der kleinste Wert q_α mit der Eigenschaft, dass der Wert von X mit Mindestwahrscheinlichkeit α nicht größer ist. Nur wenn F stetig und streng monoton wachsend ist, ist F^{-1} die Umkehrfunktion von F im üblichen Sinne.



Lemma 5.19

Es gilt:

$$y \leq F(x) \iff F^{-1}(y) \leq x.$$

Beweis

„ \Rightarrow “ Folgt unmittelbar aus der Definition von F^{-1} . Da außerdem

$$\begin{aligned} F(x) < y &\Rightarrow F\left(x + \frac{1}{n}\right) < y \text{ f\"ur ein } n \in \mathbb{N} \text{ (denn } F \text{ ist stetig von rechts)} \\ &\Rightarrow F^{-1}(y) \geq x + \frac{1}{n} \text{ (denn } F \text{ ist schwach monoton steigend)} \\ &\Rightarrow F^{-1}(y) > x \end{aligned}$$

gilt, hat man auch die Gegenrichtung. □

Satz 5.20

Es sei $F : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion mit den Eigenschaften (i)-(iii) aus Satz 5.17. Dann existiert ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B})$ mit Verteilungsfunktion F .

Beweis

Es sei $\Omega = (0,1)$, $\mathcal{A} = \mathcal{B}_{(0,1)}$ und $P_0 = \text{unif}(0,1)$. Wir definieren $X : \Omega \rightarrow \mathbb{R}$ durch $X(\omega) := F^{-1}(\omega)$. Dann ist X eine Zufallsvariable (Messbarkeit folgt aus der Monotonie von F^{-1}) und Lemma 5.19 liefert für $P := \mathcal{L}(X)$

$$P((-\infty, x]) = P_0(X \leq x) = P_0(\{\omega \in \Omega \mid F^{-1}(\omega) \leq x\}) = P_0((0, F(x)]) = F(x). \quad \square$$

Der Übergang von $P : \mathcal{B} \rightarrow \mathbb{R}$ zu $F : \mathbb{R} \rightarrow \mathbb{R}$, der letztlich durch die spezielle Struktur von $(\mathbb{R}, \mathcal{B})$ ermöglicht wird, bedeutet eine erhebliche Vereinfachung. Satz 5.20 zeigt auch, dass es zu jedem Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ eine Zufallsvariable mit diesem Wahrscheinlichkeitsmaß als Verteilung gibt.

In den Übungen wird gezeigt, dass Verteilungsfunktionen linksseitige Limiten haben, d.h. für alle $x \in \mathbb{R}$ existiert

$$F(x-) := \lim_{\substack{y \uparrow x \\ y < x}} F(y),$$

und dass die Wahrscheinlichkeit, mit der X einen Wert x annimmt, durch die Sprunghöhe $F(x) - F(x-)$ von F in x gegeben wird. Insbesondere besteht die Verteilungsfunktion zu einer diskreten Zufallsvariable nur aus Sprüngen. Ist $f : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion mit

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

so wird nach den obigen Resultaten durch

$$P((-\infty, x]) := \int_{-\infty}^x f(y) dy \text{ f\"ur alle } x \in \mathbb{R}$$

ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ definiert, das Wahrscheinlichkeitsmaß mit der **Riemann-Dichte** f . Hat die Zufallsvariable X eine solche Verteilung P , so nennen wir f eine **Wahrscheinlichkeitsdichte** von X . Zufallsvariablen mit einer Dichte werden gelegentlich „stetig“ genannt (als Gegensatz zu „diskret“) – dies bezieht sich nicht auf X als Abbildung, sondern ist nur als Abkürzung von „ X ist absolutstetig verteilt“ zu verstehen. Ist f stetig in x , so ist die zugehörige Verteilungsfunktion F ,

$$F(x) = \int_{-\infty}^x f(y) dy \text{ f\"ur alle } x \in \mathbb{R},$$

in x differenzierbar und es gilt $F'(x) = f(x)$.

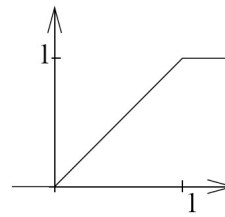
Beispiel 5.21

Im Falle $P = \text{unif}(0,1)$ hat man

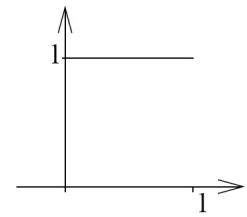
$$P((-\infty, x]) = \int_{-\infty}^x f(y) dy \text{ für alle } x \in \mathbb{R}$$

mit

$$f(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{sonst} \end{cases} \quad (= 1_{(0,1)}(y)).$$



Verteilung



Dichte

Wahrscheinlichkeiten sind in mancher Hinsicht ein infinitesimales Analogon zu Wahrscheinlichkeitsmassenfunktionen, können aber durchaus Werte größer als 1 annehmen. Ganz allgemein gilt für eine Zufallsvariable X mit Dichte f :

$$P(X \in A) = \int_A f(x) dx,$$

die Wahrscheinlichkeiten ergeben sich also als Fläche unter der Dichtefunktion. Da wir hier nur das Riemann-Integral voraussetzen, macht die rechte Seite nicht für alle Borel-Mengen Sinn – dies wird erst durch den (in der Maßtheorie bzw. der Stochastik II ausgeführten) Übergang zum Lebesgue-Integral erreicht.

5.5 Einige wichtige Verteilungen mit Riemann-Dichten

5.5.1 Die Funktion

$$f_{a,b} : \mathbb{R} \rightarrow \mathbb{R}, f_{a,b}(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{sonst,} \end{cases}$$

hat für alle $a, b \in \mathbb{R}$ mit $a < b$ die Eigenschaften

$$f_{a,b}(x) \geq 0 \text{ für alle } x \in \mathbb{R}, \int_{-\infty}^{\infty} f_{a,b}(x) dx = 1,$$

ist also Dichte eines Wahrscheinlichkeitsmaßes auf $(\mathbb{R}, \mathcal{B})$. Wir nennen dieses Wahrscheinlichkeitsmaß die **Gleich-** oder **Rechteckverteilung** auf dem Intervall (a, b) (die Randpunkte spielen keine Rolle) und schreiben hierfür $\text{unif}(a, b)$. Offensichtlich verallgemeinert dies die zu Beginn dieses Abschnitts eingeführte Gleichverteilung auf dem Einheitsintervall. Alle diese Verteilungen gehen durch affine Transformationen auseinander hervor: Hat X die Verteilung $\text{unif}(0,1)$, so gilt für die Zufallsvariable $Y := a + (b - a)X$

$$P(Y \leq y) = P\left(X \leq \frac{y-a}{b-a}\right) = \frac{y-a}{b-a} \text{ für } a < y < b,$$

$$P(Y \leq y) = 0 \text{ für } y \leq a, P(Y \leq y) = 1 \text{ für } y \geq b,$$

also insgesamt

$$P(Y \leq y) = \int_{-\infty}^y f_{ab}(x) dx \text{ für alle } y \in \mathbb{R},$$

d.h. $Y \sim \text{unif}(a, b)$. (Wir haben Satz 5.15 (ii) verwendet.)

Beispiel 5.22

Ein Stab der Länge 1 zerbricht an einer zufälligen Stelle. Wir machen die (einigermaßen unrealistische) Annahme, dass alle Bruchpositionen gleich wahrscheinlich sind und erhalten dann als Modell für dieses Zufallsexperiment den Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $\Omega = (0,1)$, $\mathcal{A} = \mathcal{B}_{(0,1)}$ und $P = \text{unif}(0,1)$. Die Länge des kürzesten Stücks ist $X(\omega) = \min\{\omega, 1 - \omega\}$, nach Satz 5.15 ist dies eine Zufallsvariable. Welche Verteilung hat X ? Offensichtlich gilt $P(X \leq x) = 0$ für $x < 0$ und $P(X \leq x) = 1$ für $x \geq \frac{1}{2}$ und für $x \in (0, \frac{1}{2})$ erhält man

$$P(X \leq x) = P(\{\omega \in (0,1) \mid \omega \leq x \text{ oder } 1 - \omega \leq x\}) = P((0, x] \cup [1 - x, 1)) = 2x.$$

Dies ist die Verteilungsfunktion zu $\text{unif}(0, \frac{1}{2})$, also ist X wieder gleichverteilt, nun auf dem Intervall $(0, \frac{1}{2})$.

5.5.2 Die Gamma-Verteilung mit Parametern α und λ ($\alpha > 0, \lambda > 0$) ist die Verteilung mit der Dichte

$$f_{\alpha, \lambda}(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \lambda^\alpha e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

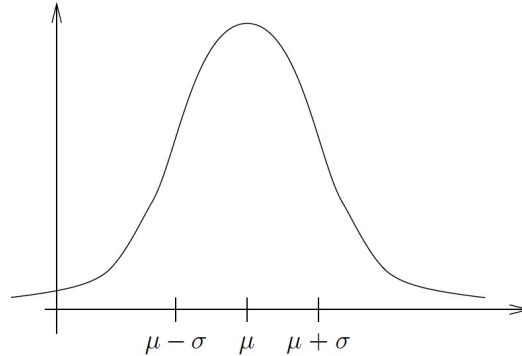
wobei

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

die Gamma-Funktion bezeichnet. Wir schreiben hierfür auch $\Gamma(\alpha, \lambda)$ und kurz $X \sim \Gamma(\alpha, \lambda)$, wenn die Zufallsvariable X diese Verteilung hat. Diese Klasse von Wahrscheinlichkeitsmaßen taucht in verschiedenen Zusammenhängen auf. Besonders wichtig ist der Fall $\alpha = 1$, der auf die **Exponentialverteilungen** führt.

5.5.3 Die Normalverteilung mit Parametern μ und σ^2 , kurz $N(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ beliebig und $\sigma^2 > 0$, ist die Verteilung mit der Dichte

$$\phi_{\mu, \sigma^2}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \quad x \in \mathbb{R}.$$



Als Graph erhält man die berühmte **Gaußsche Glockenkurve**. Die Parameter μ und σ beschreiben die Lage und Breite von ϕ . Im Falle $\mu = 0, \sigma^2 = 1$ spricht man von den **Standardparametern**, $N(0,1)$ ist die **Standardnormalverteilung**. Offensichtlich gilt

$$\phi_{\mu, \sigma^2}(x) = \frac{1}{\sigma} \phi_{0,1}\left(\frac{x - \mu}{\sigma}\right) \text{ für alle } x \in \mathbb{R}.$$

Die Verteilungsfunktion zu $N(0,1)$ ist Φ ,

$$\Phi : \mathbb{R} \rightarrow [0,1], \quad \Phi(x) := \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Eine Variante hiervon ist auch als „Fehlerfunktion“ bekannt. Die Funktion Φ ist vertafelt und in gängigen Softwarepaketen enthalten. Für die statistischen Anwendungen sind die zugehörige α -Quantile von Bedeutung. Für $\alpha = 0.9, 0.95, 0.975, 0.99, 0.995$ erhält man die Werte 1.2816, 1.6449, 1.9600, 2.3263, 2.5758.

Lemma 5.23

- (i) $\int_{-\infty}^{\infty} \phi_{\mu, \sigma^2}(x) dx = 1$ für alle $\mu \in \mathbb{R}, \sigma^2 > 0$,
- (ii) $\Phi(x) = 1 - \Phi(-x)$ für alle $x \in \mathbb{R}$,
- (iii) $X \sim N(\mu, \sigma^2), a \neq 0, b \in \mathbb{R} \Rightarrow Y := aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Beweis

(i) Substitution $y = \sigma^{-1}(x - \mu)$ zeigt, dass es reicht, den Fall $\mu = 0, \sigma^2 = 1$ zu behandeln. Standardtechniken der Analysis (Transformation auf Polarkoordinaten) ergeben

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy = \int_0^{2\pi} \int_0^{\infty} r e^{-\frac{r^2}{2}} dr d\phi = \int_0^{2\pi} \left(-e^{-\frac{r^2}{2}}\right)\Big|_0^{\infty} d\phi = 2\pi.$$

- (ii) Folgt mit $\phi(-x) = \phi(x)$.
- (iii) Im Falle $a > 0$ erhält man mit der Substitution $x' = ax + b$

$$\begin{aligned} P(Y \leq y) &= P\left(X \leq \frac{y - b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp\left(-\frac{1}{2\sigma^2 a^2}(x' - (a\mu + b))^2\right) dx'. \end{aligned}$$

Dies zeigt, dass die Verteilungsfunktion zu Y die Verteilungsfunktion zu $N(a\mu + b, a^2\sigma^2)$ ist, also $Y \sim N(a\mu + b, a^2\sigma^2)$ gilt. □

Teil (i) ist ein Nachtrag: ϕ_{μ, σ^2} ist tatsächlich eine Wahrscheinlichkeitsdichte. Wegen (ii) und (iii) reicht es, die Verteilungsfunktionen zu $N(\mu, \sigma^2)$ für die Standardparameter und Argumente ≥ 0 zu vertafeln. Beispielsweise gilt $u_\alpha = -u_{1-\alpha}$ für die Quantile u_α zu $N(0,1)$. In Kombination mit den oben genannten Quantilen ergibt sich als typische Anwendung von Lemma 5.23 (ii) und (iii) die Aussage, dass

$$P(|X - \mu| > 1.96\sigma) \approx 0.05$$

gilt, wenn X normalverteilt ist mit Parametern μ und σ^2 .

5.6 Erwartungswerte

Die „offizielle“ Verallgemeinerung erfordert das allgemeine Lebesgue-Integral, das beispielsweise zu Beginn der Vorlesung Stochastik II besprochen wird. Wir begnügen uns hier mit Andeutungen.

Ist X eine Zufallsvariable mit Dichte f und setzt man für alle $x \in \mathbb{R}$

$$\lceil x \rceil := \min \{k \in \mathbb{Z} \mid k \geq x\}, \lfloor x \rfloor := \max \{k \in \mathbb{Z} \mid k \leq x\},$$

so wird durch

$$\underline{X}_n := 2^{-n} \lfloor 2^n X \rfloor, \bar{X}_n := 2^{-n} \lceil 2^n X \rceil$$

eine Familie von diskreten Zufallsvariablen definiert, für die $\underline{X}_n \uparrow X, \bar{X}_n \downarrow X$ mit $n \rightarrow \infty$ gilt. Bei diesen können wir die bereits vorhandene Definition des Erwartungswertes verwenden:

$$\begin{aligned} E\underline{X}_n &= \sum_{k \in \mathbb{Z}} k 2^{-n} P(\underline{X}_n = k 2^{-n}) = \sum_{k \in \mathbb{Z}} k 2^{-n} \int_{k 2^{-n}}^{(k+1)2^{-n}} f(x) dx = \sum_{k \in \mathbb{Z}} \int_{k 2^{-n}}^{(k+1)2^{-n}} \frac{\lfloor 2^n x \rfloor}{2^n} f(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\lfloor 2^n x \rfloor}{2^n} f(x) dx \leq \int_{-\infty}^{\infty} x f(x) dx \leq \int_{-\infty}^{\infty} \frac{\lceil 2^n x \rceil}{2^n} f(x) dx = \dots = E\bar{X}_n. \end{aligned}$$

Wegen $\bar{X}_n - \underline{X}_n \leq 2^{-n}$ gilt

$$E\bar{X}_n - E\underline{X}_n = E(\bar{X}_n - \underline{X}_n) \leq 2^{-n},$$

es liegt also nahe, den Erwartungswert von X im Falle $\int \lceil x \rceil f(x) dx < \infty$ durch

$$EX = \int x f(x) dx$$

zu definieren. Obwohl dies für praktische Zwecke (Rechnungen) i.A. reicht, ist es doch mathematisch unbefriedigend: Eine nützliche Formel wie

$$Eg(X) = \int g(x) f(x) dx,$$

die wir im Folgenden häufig verwenden werden, ergibt sich nicht ohne Weiteres.

Beispiel 5.24

Im Falle $X \sim N(\mu, \sigma^2)$ erhält man

$$EX = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sqrt{2\mu\sigma^2}} e^{-\frac{(x-\mu)^2}{2}} dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\mu\sigma^2}} e^{-\frac{(x-\mu)^2}{2}} dx = \mu,$$

denn das erste Integral hat aus Symmetriegründen den Wert 0 und das zweite Integral ist als Integral über eine Wahrscheinlichkeitsdichte gleich 1.

5.7 Unabhängigkeit

Bisher sind uns σ -Algebren nur als „notwendiges Übel“ begegnet. Sie spielen aber in der Stochastik eine weitaus wichtigere Rolle, beispielsweise als natürliche Heimat des Unabhängigkeitsbegriffs und als Repräsentanten von Teilinformation.

Satz 5.25

Es sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Werten in dem messbaren Raum (Ω', \mathcal{A}') . Dann ist $\{X^{-1}(A) \mid A \in \mathcal{A}'\}$ eine σ -Algebra. Diese nennt man die **von X erzeugte σ -Algebra**. Schreibweise: $\sigma(X)$.

Beweis

Aufgabe 58 (a). □

Kennen wir das Resultat ω des Zufallsexperiments, so können wir von jedem Ereignis $A \in \mathcal{A}$ sagen, ob es eingetreten ist oder nicht. Die von X erzeugte σ -Algebra $\sigma(X)$ ist die Menge der Ereignisse, für die wir diese Entscheidung treffen können, wenn uns nur $X(\omega)$ bekannt ist.

Wir haben in Abschnitt 2 der Vorlesung zwei Ereignisse A und B unabhängig genannt, wenn

$$P(A \cap B) = P(A)P(B)$$

gilt und in Aufgabe 7 (d) gesehen, dass dann auch A^c und B^c unabhängig sind. Es gilt sogar, dass dann zwei beliebige Mengen aus den jeweiligen erzeugten σ -Algebren

$$\sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}, \quad \sigma(\{B\}) = \{\emptyset, B, B^c, \Omega\}$$

in diesem Sinne unabhängig sind. Dies führt auf:

Definition 5.26

Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $I \neq \emptyset$.

- (i) Eine Familie $\{A_i \mid i \in I\}$ von Unter- σ -Algebren von \mathcal{A} heißt **stochastisch unabhängig**, wenn für jede endliche Teilmenge $J = \{j_1, \dots, j_n\}$ von I und alle $A_{j_1} \in \mathcal{A}_{j_1}, \dots, A_{j_n} \in \mathcal{A}_{j_n}$ gilt:

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \tag{5.2}$$

- (ii) Ist X_i für jedes $i \in I$ eine Zufallsgröße auf (Ω, \mathcal{A}, P) mit Werten in einem messbaren Raum $(\Omega_i, \mathcal{A}_i)$, so heißt die Familie $\{X_i \mid i \in I\}$ **stochastisch unabhängig** (kurz: die Zufallsgrößen $X_i, i \in I$, sind unabhängig), wenn die Familie $\{\sigma(X_i) \mid i \in I\}$ der erzeugten σ -Algebren im Sinne von (i) unabhängig ist.

Der folgende Satz zeigt, dass man sich beim Nachweis der entscheidenden Eigenschaft (5.2) aus der Definition auf \cap -stabile Erzeugendensysteme beschränken kann.

Satz 5.27

Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $I \neq \emptyset$ und \mathcal{A}_i für jedes $i \in I$ eine Unter- σ -Algebra von \mathcal{A} mit \cap -stabilem Erzeugendensystem \mathcal{E}_i . Gilt dann

$$P\left(\bigcap_{k=1}^n E_{j_k}\right) = \prod_{k=1}^n P(E_{j_k})$$

für alle endlichen $J = \{j_1, \dots, j_n\} \subset I$ und alle $E_{j_k} \in \mathcal{E}_{j_k}, k = 1, \dots, n$, so sind $\mathcal{A}_i, i \in I$, stochastisch unabhängig.

Beweis

Sei $J = \{j_1, \dots, j_n\} \subset I$. Sei \mathcal{D}_{j_1} die Menge aller $A \in \mathcal{A}_{j_1}$ mit

$$P(A \cap E_{j_2} \cap \dots \cap E_{j_n}) = P(A)P(E_{j_2}) \cdot \dots \cdot P(E_{j_n})$$

für alle $E_{j_2} \in \mathcal{E}_{j_2}, \dots, E_{j_n} \in \mathcal{E}_{j_n}$. Man sieht leicht, dass \mathcal{D}_{j_1} ein Dynkin-System ist. Da \mathcal{D}_{j_1} den \cap -stabilen Erzeuger \mathcal{E}_{j_1} von \mathcal{A}_{j_1} enthält, gilt also $\mathcal{D}_{j_1} = \mathcal{A}_{j_1}$ nach Satz 5.7 (ii). Im zweiten Schritt sei \mathcal{D}_{j_2} die Menge aller $A \in \mathcal{A}_{j_2}$ mit

$$P(A_{j_1} \cap A \cap E_{j_3} \cap \dots \cap E_{j_n}) = P(A_{j_1})P(A)P(E_{j_3}) \cdot \dots \cdot P(E_{j_n})$$

für alle $E_{j_3} \in \mathcal{E}_{j_3}, \dots, E_{j_n} \in \mathcal{E}_{j_n}$. Man sieht wieder, dass \mathcal{D}_{j_2} ein Dynkin-System ist, das nach dem bereits bewiesenen Teil \mathcal{E}_{j_2} enthält und es folgt wieder $\mathcal{D}_{j_2} = \mathcal{A}_{j_2}$. Nach insgesamt n Schritten dieser Art erhält man die gewünschte Beziehung

$$P(A_{j_1} \cap \dots \cap A_{j_n}) = P(A_{j_1}) \cdot \dots \cdot P(A_{j_n})$$

für alle $A_{j_1} \in \mathcal{A}_{j_1}, \dots, A_{j_n} \in \mathcal{A}_{j_n}$. □

Bei einer diskreten Zufallsgröße X bilden die Mengen $X^{-1}(\{x\}), x \in \text{Bild}(X)$, ein \cap -stabiles Erzeugendensystem von $\sigma(X)$. Satz 5.7 zeigt also, dass Teil (ii) von Definition 5.26 zu Definition 4.16 „abwärtskompatibel“ ist.

Der Zugang über σ -Algebren bietet Vorteile, beispielsweise beim Beweis des folgenden Satzes, der grob gesprochen besagt, dass Funktionen unabhängiger Zufallsgrößen wieder unabhängig sind.

Satz 5.28

Für jedes $i \in I$ seien X_i eine Zufallsgröße mit Werten in $(\Omega_i, \mathcal{A}_i)$, $(\Omega'_i, \mathcal{A}'_i)$ ein weiterer messbarer Raum und $g_i : \Omega_i \rightarrow \Omega'_i$ eine $(\mathcal{A}_i, \mathcal{A}'_i)$ -messbare Abbildung. Ist dann $\{X_i \mid i \in I\}$ eine unabhängige Familie, so ist auch $\{Y_i \mid i \in I\}$ mit $Y_i := g_i(X_i)$ unabhängig.

Beweis

$\sigma(Y_i) \subset \sigma(X_i)$. □

Beispiel 5.29

Es sei $(\Omega, \mathcal{A}, P) = ([0,1], \mathcal{B}_{(0,1)}, \text{unif}(0,1))$. Für jedes $n \in \mathbb{N}$ werde $X_n = \Omega \rightarrow \{0,1\}$ definiert durch

$$X_n(\omega) := \lfloor 2^n \omega \rfloor - 2 \lfloor 2^{n-1} \omega \rfloor.$$

Dann gilt $\omega = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega)$ – die Folge $0.X_1(\omega)X_2(\omega)X_3(\omega)\dots$ ist also eine (mehr oder weniger: die) Binärdarstellung von ω .

Für alle $k_1, \dots, k_n \in \{0,1\}$ gilt

$$P(X_1 = k_1, \dots, X_n = k_n) = P\left(\sum_{l=1}^n 2^{-l} k_l \leq \omega < \sum_{l=1}^n 2^{-l} k_l + 2^{-n}\right) = 2^{-n},$$

denn das Intervall besteht aus allen $\omega \in [0,1)$, deren Binärdarstellung mit den Ziffern (Bits) k_1, \dots, k_n beginnt. Für beliebige $i_1 < i_2 < \dots < i_n$ erhält man somit

$$\begin{aligned} P(X_{i_1} = 1, \dots, X_{i_n} = 1) &= \sum_{\substack{(k_1, \dots, k_n) \in \{0,1\}^n \\ k_{i_j} = 1 \text{ für } j=1, \dots, n}} P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) \\ &= 2^{-i_n} \# \{(k_1, \dots, k_n) \in \{0,1\}^n \mid k_{i_j} = 1 \text{ für } j = 1, \dots, n\} \\ &= 2^{-i_n} 2^{i_n - n} \text{ (denn genau } n \text{ Positionen sind festgelegt)} \\ &= 2^{-n}. \end{aligned}$$

Insbesondere folgt $P(X_{i_j} = 1) = \frac{1}{2}$ und damit insgesamt

$$P(X_{i_1} = 1, \dots, X_{i_n} = 1) = P(X_{i_1} = 1) \cdot \dots \cdot P(X_{i_n} = 1).$$

Da $\{X_i^{-1}(\{1\})\}$ ein \cap -stabiles Erzeugendensystem von $\sigma(X_i)$ ist, haben wir damit die Unabhängigkeit der Zufallsvariablen X_1, X_2, X_3, \dots gezeigt. Außerdem gilt $\mathcal{L}(X_i) = \text{Bin}(1, \frac{1}{2})$, die gesamte Konstruktion kann also als Modell für den unendlich oft wiederholten Wurf einer fairen Münze dienen. Umgekehrt ließe sich aus einer unendlichen Folge von Münzwürfen k_1, k_2, \dots eine auf $[0,1)$ gleichverteilte Zahl x durch $x := \sum_{i=1}^{\infty} k_i 2^{-i}$ konstruieren!

Wir betrachten nun den Fall reellwertiger Zufallsgrößen etwas näher. Sind X und Y unabhängige Zufallsvariablen mit Verteilungsfunktionen F_X und F_Y , so gilt

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$$

für alle $x, y \in \mathbb{R}$. Definiert man die **gemeinsame Verteilungsfunktion** von zwei (beliebigen) Zufallsvariablen X und Y durch

$$F_{X,Y} = \mathbb{R}^2 \rightarrow \mathbb{R}, \quad F_{X,Y}(x,y) := P(X \leq x, Y \leq y),$$

so erhält man, dass bei Unabhängigkeit die gemeinsame Verteilungsfunktion das Produkt der einzelnen Verteilungsfunktionen ist, d.h.

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ für alle } x, y \in \mathbb{R}.$$

Die Mengen $(-\infty, x]$, $x \in \mathbb{R}$, bilden nach Satz 5.3 ein \cap -stabiles Erzeugendensystem von $\mathcal{B}(\mathbb{R})$, also folgt mit Satz 5.27 auch umgekehrt die Unabhängigkeit von X und Y aus dieser Darstellung.

Sind X und Y stetige Zufallsvariablen mit Dichten f_X, f_Y , d.h. insbesondere

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad F_Y(y) = \int_{-\infty}^y f_Y(z) dz,$$

so erhält man bei Unabhängigkeit

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) du dv.$$

In nahe liegender Verallgemeinerung des eindimensionalen Falles nennt man $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ eine **gemeinsame Dichte** von X und Y , wenn

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y) dx dy$$

für „hinreichend viele“ $A \subset \mathbb{R}^2$ gilt (in der Vorlesung Stochastik II wird dies präzisiert). Insbesondere hat man bei unabhängigen Zufallsvariablen X, Y mit Dichten f_X, f_Y

$$f_{X,Y}(x,y) = f_X(x) f_Y(y),$$

die Analogie zum diskreten Fall (Satz 4.17) ist offensichtlich.

Mit gemeinsamen Dichtefunktionen lassen sich auch beispielsweise Erwartungswerte von Funktionen von mehreren Zufallsvariablen ausrechnen. Wir beschränken uns wie oben auf den Fall von zwei Zufallsvariablen X und Y . Erinnerung: Sind X und Y diskrete Zufallsgrößen mit gemeinsamer Massenfunktion

$$p_{X,Y}(x,y) = P(X = x, Y = y),$$

so gilt unter der Voraussetzung, dass die Summe absolut konvergiert,

$$Eg(X,Y) = \sum_{x \in \text{Bild}(X)} \sum_{y \in \text{Bild}(Y)} g(x,y) p_{X,Y}(x,y)$$

(siehe Beispiel 4.13). Ganz analog hat man in der stetigen Situation

$$Eg(X,Y) = \iint g(x,y) f_{X,Y}(x,y) dx dy$$

(genauer, beispielsweise zur Messbarkeit von g , wird in der Vorlesung Stochastik II besprochen). Hiermit erhält man u.A. eine Variante der Multiplikationsregel für unabhängige stetige Zufallsvariablen X, Y :

$$EXY = \iint xyf_X(x)f_Y(y)dx dy = \left(\int xf_X(x)dx\right)\left(\int yf_Y(y)dy\right) = (EX)(EY),$$

man vergleiche dies mit Satz 4.18. Auch Begriffe wie Kovarianz etc. lassen sich auf diese Weise auf den stetigen Fall übertragen.

In der Maßtheorie (siehe die Vorlesung mit diesem Namen, aber auch den Beginn der Stochastik II) wird gezeigt, dass sowohl der diskrete als auch der stetige Fall Spezialfälle einer allgemeinen Theorie sind. Es gibt übrigens durchaus auch Zufallsvariablen, die weder diskret noch stetig sind – ein Beispiel wird in den Übungen behandelt.

Mit dem Obenstehenden sind die möglichen Analogiebetrachtungen bei Weitem nicht erschöpft. Siehe beispielsweise die Aufgaben 70 und 74 zur Faltung.

Beispiel 5.30

Die Lebensdauer X einer Glühbirne vom Typ A sei exponentialverteilt mit Parameter λ_A , Y sei die Lebensdauer einer Glühbirne vom Typ B, ebenfalls exponentialverteilt, nun mit Parameter λ_B . Wir setzen voraus, dass die Zufallsvariablen X und Y unabhängig sind. Mit welcher Wahrscheinlichkeit brennt die B-Birne länger als die A-Birne? Die obigen Überlegungen führen auf

$$\begin{aligned} P(X < Y) &= P((X, Y) \in \{(x, y) \in \mathbb{R}^2 \mid x < y\}) = \iint_{\{(x, y) \in \mathbb{R}^2 \mid x < y\}} f_{X, Y}(x, y) dy dx \\ &= \iint_{\{(x, y) \in \mathbb{R}^2 \mid x < y\}} \lambda_A e^{-\lambda_A y} \lambda_B e^{-\lambda_B x} dy dx = \int_0^\infty \left(\int_x^\infty \lambda_B e^{-\lambda_B y} dy \right) \lambda_A e^{-\lambda_A x} dx \\ &= \lambda_A \int_0^\infty e^{-\lambda_B x} e^{-\lambda_A x} dx = \frac{\lambda_A}{\lambda_A + \lambda_B}. \end{aligned}$$

Kapitel 6

Verteilungskonvergenz und Normalapproximation

6.1 Verteilungskonvergenz

Wir hatten bereits in Satz 4.4 gesehen, dass man die Verteilung $\text{Bin}(n, p)$ bei großem n und kleinem p durch die Poisson-Verteilung mit Parameter $\lambda = np$ approximieren kann. In diesem Absatz geht es allgemein um Approximationen von Verteilungen und zugehörige Grenzwertsätze. Zur Einführung betrachten wir ein weiteres (nun nicht-diskretes) Beispiel.

Beispiel 6.1

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und $\text{unif}(0,1)$ -verteilt. Wir interessieren uns für die Verteilung von $Y_n := \min\{X_1, \dots, X_n\}$, also des Minimums der ersten n Variablen. Die zugehörige Verteilungsfunktion ist

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) = 1 - P(X_1 > y, \dots, X_n > y) = 1 - \prod_{i=1}^n P(X_i > y), \text{ da } X_1, \dots, X_n \text{ unabhängig} \\ &= 1 - \prod_{i=1}^n (1 - F_{X_i}(y)) = \begin{cases} 1 - (1 - y)^n, & 0 \leq y \leq 1, \\ 0, & y < 0, \\ 1, & y > 1. \end{cases} \end{aligned}$$

Mit $n \rightarrow \infty$ erhält man als (punktweisen) Limes den Wert 1 bei $y > 0$ und den Wert 0 bei $y \leq 0$. Man beachte, dass diese Grenzfunktion keine Verteilungsfunktion ist, da sie in 0 nicht rechtsstetig ist. Diese Rechnung liefert nur die auch intuitiv klare und nicht sonderlich interessante Aussage, dass das Minimum „in gewisser Weise“ gegen 0 strebt. Eine erheblich interessantere Aussage erhält man, wenn man geeignet umskaliert: Für alle $y > 0$, n groß genug, gilt

$$F_{nY_n}(y) = P(nY_n \leq y) = 1 - \left(1 - \frac{y}{n}\right)^n \rightarrow 1 - e^{-y}$$

mit $n \rightarrow \infty$, d.h. bei großem n ist nY_n näherungsweise exponentialverteilt mit Parameter 1.

Für Wahrscheinlichkeitsmaße gibt es eine ganze Reihe von Konvergenzbegriffen. Der wichtigste ist der folgende.

Definition 6.2

Es seien P, P_n , $n \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ mit Verteilungsfunktionen F, F_n , $n \in \mathbb{N}$,

$$C(F) := \{x \in \mathbb{R} \mid F \text{ stetig in } x\}$$

bezeichne die Menge der Stetigkeitspunkte von F . Gilt dann

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ für alle } x \in C(F),$$

so sagt man, dass P_n **schwach** gegen P . Schreibweise: $P_n \xrightarrow{w} P$ (w - weak). Sind X, X_n , $n \in \mathbb{N}$, Zufallsvariablen mit den Verteilungen P, P_n , $n \in \mathbb{N}$, so sagt man in dieser Situation, dass X_n **in Verteilung** gegen X konvergiert. Schreibweise: $X_n \xrightarrow{d} X$ (d - distribution).

Warum die Einschränkung auf $C(F)$? Ohne diese Einschränkung würde man mit $X_n \equiv \frac{1}{n}$ nicht den Limes $X \equiv 0$ erhalten, da dann $F_n(0) = 0 \not\rightarrow 1 = F(0)$.

6.2 Normalapproximation bei Poisson-Verteilungen

Es sei X Poisson-verteilt mit Parameter λ . Aus Beispiel 4.11 (ii) ist $EX = \text{var}(X) = \lambda$ bekannt, also folgt mit der Ungleichung von Chebyshev (Satz 4.31 (ii))

$$P_\lambda(|X - \lambda| \geq C\sqrt{\lambda}) \leq \frac{1}{C^2},$$

die Hauptmasse der Verteilung liegt also im Intervall

$$I(C; \lambda) := [\lambda - C\sqrt{\lambda}, \lambda + C\sqrt{\lambda}] \cap \mathbb{N}_0.$$

Es sei

$$\phi(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

die Dichte zu $N(\mu, \sigma^2)$. Wir behaupten nun, dass $P_\lambda(X = k)$ bei großem λ auf $I(C, \lambda)$ gleichmäßig durch $\phi(k | \lambda, \lambda)$ approximiert werden kann:

Satz 6.3 (Normalapproximation für Poisson-Verteilungen, lokale Form)

Für alle $C > 0$ gilt

$$\lim_{\lambda \rightarrow \infty} \sup_{k \in I(C, \lambda)} \left| \frac{e^{-\lambda} \frac{\lambda^k}{k!}}{\frac{1}{\sqrt{2\pi\lambda}} \exp\left(-\frac{1}{2\lambda}(k - \lambda)^2\right)} - 1 \right| = 0.$$

Beweis

Sei $C > 0$. Wir setzen für $k \in \mathbb{N}_0$, $\lambda > 0$

$$g(k | \lambda) := \log P_\lambda(X = k) = -\lambda + k \log \lambda - \log k!, \quad g(k, \lambda) := \frac{1}{\sqrt{\lambda}}(k - \lambda).$$

Dann gilt

$$g(k | \lambda) - g(k - 1 | \lambda) = \log \lambda - \log k \quad \text{für alle } k \in \mathbb{N}$$

sowie

$$|z(k, \lambda)| \leq C \quad \text{für alle } k \in I(C, \lambda).$$

Mit $\log(1 + x) = x + O(x^2)$, $x \rightarrow 0$ erhält man

$$\log k = \log(\lambda + \sqrt{\lambda}z(k, \lambda)) = \log \lambda + \log\left(1 + \frac{1}{\sqrt{\lambda}}z(k, \lambda)\right) = \log \lambda + \frac{1}{\sqrt{\lambda}}z(k, \lambda) + O\left(\frac{1}{\lambda}\right)$$

mit $\lambda \rightarrow \infty$, gleichmäßig in $k \in I(C, \lambda)$, also

$$g(k | \lambda) - g(\lfloor \lambda \rfloor | \lambda) = \sum_{j=\lfloor \lambda \rfloor+1}^k (g(j | \lambda) - g(j-1 | \lambda)) = -\frac{1}{\sqrt{\lambda}} \sum_{j=\lfloor \lambda \rfloor+1}^k z(j | \lambda) + O\left(\frac{1}{\sqrt{\lambda}}\right),$$

denn die Anzahl der Summanden ist wegen $k \in I(C, \lambda)$ maximal von der Ordnung $\sqrt{\lambda}$ und es gilt

$$\sqrt{\lambda}O\left(\frac{1}{\lambda}\right) = O\left(\frac{1}{\sqrt{\lambda}}\right).$$

Setzt man die Definition von z ein, so ergibt dies

$$g(k | \lambda) - g(\lfloor \lambda \rfloor | \lambda) = -\frac{1}{\lambda} \sum_{j=\lfloor \lambda \rfloor+1}^k (j - \lambda) + O\left(\frac{1}{\sqrt{\lambda}}\right) = -\frac{1}{2\lambda}(k - \lambda)^2 + O\left(\frac{1}{\sqrt{\lambda}}\right)$$

(die „Ungenauigkeiten“ landen alle im O -Term). Definiert man

$$h(\lambda) := \exp(g(\lfloor \lambda \rfloor | \lambda)),$$

so haben wir gezeigt:

$$P_\lambda(X = k) = h(\lambda) \exp\left(-\frac{1}{2\lambda}(k - \lambda)^2\right) \left(1 + O\left(\frac{1}{\sqrt{\lambda}}\right)\right)$$

mit $\lambda \rightarrow \infty$ und gleichmäßig über $I(C, \lambda)$ – die Abhängigkeit von k ist weitgehend verschwunden.

Es bleibt noch zu zeigen, dass

$$\lim_{\lambda \rightarrow \infty} \sqrt{2\pi\lambda} h(\lambda) = 1 \tag{6.1}$$

gilt. Hierfür summieren wir die bereits erhaltene Approximation über $I(C, \lambda)$:

$$\begin{aligned} 1 &\geq P_\lambda(X \in I(C, \lambda)) = h(\lambda) \left(\sum_{k \in I(C, \lambda)} \exp\left(-\frac{(k - \lambda)^2}{2\lambda}\right) \right) \left(1 + O\left(\frac{1}{\sqrt{\lambda}}\right)\right) \\ &= \sqrt{2\pi\lambda} h(\lambda) \left(1 + O\left(\frac{1}{\sqrt{\lambda}}\right)\right) \sum_{|k - \lambda| \leq C\sqrt{\lambda}} \frac{1}{\sqrt{\lambda}} \phi\left(\frac{k - \lambda}{\sqrt{\lambda}}\right). \end{aligned}$$

Die letzte Summe ist eine Riemann-Summe, die mit $\lambda \rightarrow \infty$ gegen $\int_{-C}^C \phi(x) dx$ konvergiert. Damit folgt

$$1 \geq (\limsup_{\lambda \rightarrow \infty} \sqrt{2\pi\lambda} h(\lambda)) \int_{-C}^C \phi(x) dx.$$

Wegen $\int_{-\infty}^{\infty} \phi(x) dx = 1$ erhält man daher mit $C \rightarrow \infty$

$$\limsup_{\lambda \rightarrow \infty} \sqrt{2\pi\lambda} h(\lambda) \leq 1.$$

Andererseits hat man nach Chebyshev

$$1 - \frac{1}{C^2} \leq P_\lambda(X \in I(C, \lambda)),$$

erhält also mit denselben Rechnungen

$$(\limsup_{\lambda \rightarrow \infty} \sqrt{2\pi\lambda} h(\lambda)) \int_{-C}^C \phi(x) dx \geq 1 - \frac{1}{C^2},$$

woraus mit $C \rightarrow \infty$

$$\liminf_{\lambda \rightarrow \infty} \sqrt{2\pi\lambda} h(\lambda) \geq 1$$

und damit insgesamt (6.1) folgt. □

Als Korollar ergibt sich eine (be)merkwürdige Formel:

Korollar 6.4 (Stirling-Formel)

Es gilt:

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} e^{-n} n^{n+\frac{1}{2}}} = 1.$$

Beweis

Setze $\lambda = \lambda_n = n$, $k = n$ in Satz 6.3. □

Die Gleichmäßigkeit in Satz 6.3 erlaubt den Übergang zu Summen.

Satz 6.5 (Normalapproximation für Poisson-Verteilungen, kumulative Form)

Für alle $\lambda > 0$ sei X_λ eine mit Parameter λ Poisson-verteilte Zufallsvariable. Dann gilt

$$\mathcal{L}\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}}\right) \xrightarrow{w} N(0,1) \text{ mit } \lambda \rightarrow \infty.$$

Beweis

Für beliebige $a, b \in \mathbb{R}$ mit $a < b$ gibt es ein $C > 0$ mit $[a, b] \subset [-C, C]$, also folgt wie im Beweis zu Satz 6.3

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} P\left(a \leq \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) &= \lim_{\lambda \rightarrow \infty} \sum_{\lambda + a\sqrt{\lambda} \leq k \leq \lambda + b\sqrt{\lambda}} P(X_\lambda = k) = \lim_{\lambda \rightarrow \infty} \sum_{a \leq \frac{k-\lambda}{\sqrt{\lambda}} \leq b} \frac{1}{\sqrt{\lambda}} \phi\left(\frac{k-\lambda}{\sqrt{\lambda}}\right) \\ &= \int_a^b \phi(x) dx = \Phi(b) - \Phi(a), \end{aligned}$$

wobei Φ wieder die (stetige) Verteilungsfunktion zu $N(0,1)$ bezeichnet. Es bleibt zu zeigen, dass a durch $-\infty$ ersetzt werden kann.

Sei hierzu $\varepsilon > 0$. Dann existiert ein $a > -\infty$ mit $\Phi(a) < \frac{\varepsilon}{2}$ und

$$P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} < a\right) < \frac{\varepsilon}{2} \text{ für alle } \lambda > 0.$$

Zur Begründung verwendet man Satz 5.17 (i) bei Φ und die mit der Ungleichung von Chebyshev erhältliche, von λ unabhängige Oberschranke $\frac{1}{a^2}$ bei X_λ . Die Dreiecksungleichung liefert

$$\left|P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) - \Phi(b)\right| \leq \left|P\left(a \leq \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) - (\Phi(b) - \Phi(a))\right| + P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq a\right) + \Phi(a).$$

Hieraus und aus den Bedingungen an a folgt

$$\limsup_{\lambda \rightarrow \infty} \left|P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) - \Phi(b)\right| \leq \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, ergibt dies

$$\lim_{\lambda \rightarrow \infty} P\left(\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \leq b\right) = \Phi(b) \text{ für alle } b \in \mathbb{R}. \quad \square$$

6.3 Normalapproximation bei Binomialverteilungen

Wir betrachten nun das Verhalten von $\text{Bin}(n, p)$ bei $n \rightarrow \infty$ und festem $p \in (0, 1)$. (Insbesondere gilt dann $np \rightarrow \infty$ im Gegensatz zu dem in Abschnitt 4.2.2 besprochenen Gesetz der seltenen Ereignisse.)

Satz 6.6 (Normalapproximation für Binomialverteilungen, lokale Form)

Es sei $p \in (0, 1)$ und für alle $C > 0, n \in \mathbb{N}$,

$$I(C, n) := [np - C\sqrt{n}, np + C\sqrt{n}] \cap \{0, \dots, n\}.$$

Dann gilt für alle $C > 0$

$$\lim_{n \rightarrow \infty} \sup_{k \in I(C, n)} \left| \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{1}{2np(1-p)}(k-np)^2\right)} - 1 \right| = 0.$$

Beweis

Setzt man $\lambda = np, \mu = n(1-p)$, so gilt

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-\lambda} \lambda^k}{k!} \frac{e^{-\mu} \mu^{n-k}}{(n-k)!} = \frac{e^{-(\lambda+\mu)} (\lambda+\mu)^n}{n!}.$$

Wendet man nun Satz 6.3 drei Mal an, so ergibt sich die Behauptung nach etwas Rechenarbeit. □

Der Satz besagt, dass für eine $\text{Bin}(n, p)$ -verteilte Zufallsvariable $X_{n,p}$ bei großem n (und festem p) die Wahrscheinlichkeiten $P(X_{n,p} = k)$ gleichmäßig auf $I(C, n)$ durch $\phi(k | np, np(1-p))$ approximiert werden können. Es gibt auch hier eine kumulative Form, die, entsprechend formuliert, auf einen der frühen Klassiker der Wahrscheinlichkeitstheorie führt.

Satz 6.7 (de Moivre-Laplace)

Es sei X_1, X_2, \dots eine Folge von unabhängigen Zufallsvariablen mit

$$P(X_i = 1) = 1 - P(X_i = 0) = p \text{ für alle } i \in \mathbb{N}$$

mit einem festen $p \in (0, 1)$. Für alle $n \in \mathbb{N}$ sei $S_n := \sum_{i=1}^n X_i$. Dann gilt

$$\mathcal{L}\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right) \xrightarrow{w} N(0, 1) \text{ mit } n \rightarrow \infty.$$

Beweis

Wegen $S_n \sim \text{Bin}(n, p)$ ist dies die kumulative Form von Satz 6.6 und kann, aufbauend auf diesem, ganz analog wie Satz 6.5 bewiesen werden. □

Bemerkung 6.8

Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit endlichem zweiten Moment. Dann nennt man $(S_n)_{n \in \mathbb{N}}$ mit $S_n := \sum_{i=1}^n X_i$ die Folge der **Partialsommen** und

$$S_n^* := \frac{1}{\sqrt{\text{var } S_n}} (S_n - ES_n)$$

die n -te **standardisierte** Partialsumme (klar: $ES_n^* = 0, \text{var}(S_n^*) = 1$, daher standardisiert). Man sagt, dass $(X_n)_{n \in \mathbb{N}}$ dem **zentralen Grenzwertsatz** (ZGWS) genügt, wenn S_n^* mit $n \rightarrow \infty$ in Verteilung gegen $N(0, 1)$ konvergiert, wenn also gilt

$$\lim_{n \rightarrow \infty} P(S_n^* \leq y) = \Phi(y) \text{ für alle } y \in \mathbb{R}.$$

Satz 6.7 zeigt, dass $(X_n)_{n \in \mathbb{N}}$ dem zentralen Grenzwertsatz genügt, wenn die Variablen unabhängig und $\text{Bin}(1, p)$ -verteilt sind. Auch die kumulative Form der Normalapproximation von Poisson-Verteilungen führt auf eine solche Aussage: Sind die Zufallsvariablen $X_n, n \in \mathbb{N}$, unabhängig und Poisson-verteilt mit Parameter $\lambda_n = n\lambda$, also folgt aus Satz 6.5, dass $(X_n)_{n \in \mathbb{N}}$ dem zentralen Grenzwertsatz genügt.

Bemerkung 6.9*

Wir haben die jeweilige Aussage zur Verteilungskonvergenz aus einer entsprechenden lokalen Aussage hergeleitet, die stärker ist: Nach Satz 6.6 kann die Wahrscheinlichkeitsmassenfunktion zu $\text{Bin}(n, p)$ durch die

Wahrscheinlichkeitsdichte der Normalverteilung approximiert werden, deren Lage- und Breiteparameter mit dem Erwartungswert und der Varianz von $\text{Bin}(n, p)$ übereinstimmen.

Beispiel 6.9

Mit welcher Wahrscheinlichkeit erscheint beim 600-maligen Wurf eines Würfels mindestens 90-mal und höchstens 105-mal eine 6? Als tatsächlicher Wert ergibt sich

$$\sum_{k=90}^{105} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{600-k} = 0.60501\dots,$$

Satz 6.6 führt auf den Näherungswert

$$\sum_{k=90}^{105} \phi\left(k \mid 100, \frac{500}{6}\right) = 0.601769\dots,$$

also eine recht gute Annäherung (der relative Fehler beträgt etwa 0.54%). Bei der Verwendung der kumulativen Form, also des Satzes von de Moivre-Laplace, erhält man

$$\begin{aligned} P(90 \leq S_{600} \leq 105) &= P(S_{600} \leq 105) - P(S_{600} \leq 89) = P\left(S_{600}^* \leq \frac{105 - 100}{\sqrt{\frac{500}{6}}}\right) - P\left(S_{600}^* \leq \frac{89 - 100}{\sqrt{\frac{500}{6}}}\right) \\ &\approx \Phi\left(\frac{5}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(\frac{-11}{\sqrt{\frac{500}{6}}}\right) \approx 0.5939551\dots \end{aligned}$$

Man kann diese Approximation mit der so genannten **Stetigkeitskorrektur** verbessern: Die Verteilungsfunktion zu S_n ist vom reinen Sprungtyp und auf Intervallen der Länge 1 konstant, Φ ist stetig. Ersetzt man $P(S_n \leq k)$ durch $P(S_n \leq k + 0.5)$, wählt man also den Mittelpunkt des Konstantheitsintervalls, so erhält man

$$\begin{aligned} P(90 \leq S_{600} \leq 105) &= P(S_{600} \leq 105.5) - P(S_{600} \leq 89.5) = \dots \\ &\approx \Phi\left(\frac{5.5}{\sqrt{\frac{500}{6}}}\right) - \Phi\left(\frac{-10.5}{\sqrt{\frac{500}{6}}}\right) \approx 0.6015504\dots, \end{aligned}$$

also eine Verringerung des relativen Fehlers von 1.83% auf 0.57%.

6.4 Normalapproximation bei Gamma-Verteilungen

Als Beispiel für die Normalapproximation in einem nichtdiskreten Fall betrachten wir die in Abschnitt 5.5.2 eingeführte Gamma-Verteilung $\Gamma(\alpha, \lambda)$ mit der Dichtefunktion

$$f(x \mid \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0.$$

Ist $X_{\alpha, \lambda}$ eine Zufallsvariable mit dieser Verteilung, so gilt $EX_{\alpha, \lambda} = \frac{\alpha}{\lambda}$, $\text{var}(X_{\alpha, \lambda}) = \frac{\alpha}{\lambda^2}$. Wir vergleichen zunächst wieder die Dichte der Gamma-Verteilung mit der Dichte derjenigen Normalverteilung, die bezüglich Erwartungswert und Varianz passt.

Satz 6.10 (Normalapproximation für Gamma-Verteilungen, lokale Form)

Es sei $\lambda > 0$ fest und für alle $C > 0$ sei

$$I(C, \alpha) := \left[\frac{\alpha}{\lambda} - C \frac{\sqrt{\alpha}}{\lambda}, \frac{\alpha}{\lambda} + C \frac{\sqrt{\alpha}}{\lambda} \right] \cap (0, \infty).$$

Dann gilt

$$\lim_{\alpha \rightarrow \infty} \sup_{x \in I(C, \alpha)} \left| \frac{\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}}{\frac{\lambda}{\sqrt{2\pi\alpha}} \exp\left(-\frac{\lambda^2}{2\alpha} \left(x - \frac{\alpha}{\lambda}\right)^2\right)} - 1 \right| = 0.$$

Beweis

Ersetzt man λx durch x , so erkennt man, dass es genügt, die Behauptung für den Fall $\lambda = 1$ zu beweisen. Diese Überlegung ist wegen der Skalierungseigenschaften der Gamma-Verteilungen nahe liegend und erleichtert die Schreibearbeit. Sei $C > 0$. Wir verwenden

$$\log(1+z) = z - \frac{z^2}{2} + o(z^2) \quad \text{für } z \rightarrow \infty$$

sowie die Stirling-Formel in der Form

$$\log \Gamma(\alpha) = \frac{1}{2} \log(2\pi) + \left(\alpha - \frac{1}{2}\right) \log \alpha - \alpha + o(1) \quad \text{für } \alpha \rightarrow \infty$$

(man beachte $\Gamma(n+1) = n!$) und erhalten

$$\begin{aligned} \log f(x | \alpha, 1) - \log \phi(x | \alpha, \alpha) &= -\log \Gamma(\alpha) + (\alpha - 1) \log x - x + \frac{1}{2} \log(2\pi\alpha) + \frac{1}{2\alpha} (x - \alpha)^2 \\ &= \alpha + (\alpha - 1) \log\left(1 + \frac{x - \alpha}{\alpha}\right) - x + \frac{1}{2\alpha} (x - \alpha)^2 + o(1) \\ &= -(x - \alpha) + (\alpha - 1) \left(\frac{x - \alpha}{\alpha} - \frac{1}{2} \left(\frac{x - \alpha}{\alpha} \right)^2 + o\left(\frac{1}{\alpha}\right) \right) + \frac{1}{2\alpha} (x - \alpha)^2 + o(1) \\ &= o(1), \end{aligned}$$

gleichmäßig auf $x \in I(C, \alpha)$. Hieraus folgt die gewünschte Aussage. \square

Wie in den vorangegangenen beiden Unterabschnitten ist die auch hier die lokale Form als rein analytische Aussage formuliert. Bei der kumulativen Form wählen wir eine Darstellung, die den stochastischen Kern der Aussage, die Approximation von auf Erwartungswert 0 und Varianz 1 standardisierten Zufallsvariablen (bzw. deren Verteilungen) durch die Standardnormalverteilung, herausstellt.

Satz 6.11 (Normalapproximation für Gamma-Verteilungen, kumulative Form)

Es sei $\lambda > 0$. Für alle $\alpha > 0$ sei X_α eine $\Gamma(\alpha, \lambda)$ -verteilte Zufallsvariable. Dann gilt

$$\mathcal{L}\left(\frac{X_\alpha - EX_\alpha}{\sqrt{\text{var}(X_\alpha)}}\right) \xrightarrow{w} N(0, 1) \text{ mit } \alpha \rightarrow \infty.$$

Beweis

Wie zu Satz 6.5. \square

Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, alle mit derselben Verteilung $\Gamma(\alpha, \lambda)$, so hat die n -Partialsomme $S_n = \sum_{k=1}^n X_k$ die Verteilung $\Gamma(n\alpha, \lambda)$ (Aufgabe 70). Satz 6.10 zeigt, dass auch solche Folgen dem zentralen Grenzwertsatz genügen. In der Vorlesung Stochastik II wird gezeigt, dass dies ganz allgemein für Folgen von unabhängigen, identisch verteilten Zufallsvariablen gilt, wenn nur das zugehörige zweite Moment endlich ist.

Kapitel 7

Grundbegriffe der mathematischen Statistik

7.1 Allgemeines

In der Wahrscheinlichkeitstheorie geht man von einem Modell (Ω, \mathcal{A}, P) für ein Zufallsexperiment aus und berechnet beispielsweise die Wahrscheinlichkeit eines Ereignisses A . In der Statistik soll man, ausgehend von den bei der Ausführung des Experiments gewonnenen Daten, eine Aussage über das zugehörige P machen (P ist also unbekannt). Beim zehnfachen Münzwurf ist beispielsweise eine typische wahrscheinlichkeitstheoretische Frage:

Mit welcher Wahrscheinlichkeit kommt acht Mal Kopf, wenn die Münze fair ist?

Typische statistische Fragestellungen wären in dieser Situation:

Es kam acht Mal Kopf. Welchen Wert hat p , die Wahrscheinlichkeit für Kopf?

Ist die Münze fair, d.h. gilt $p = \frac{1}{2}$?

Allgemein gehen wir von einem messbaren Raum $(\mathcal{X}, \mathcal{A})$, dem **Stichprobenraum**, der die möglichen Datenwerte x erhält. Auf $(\mathcal{X}, \mathcal{A})$ hat man eine Familie \mathcal{P} von Wahrscheinlichkeitsmaßen, die in Frage kommenden Verteilungen für die Daten. Diese Familie kann die Klasse aller Wahrscheinlichkeitsmaße auf dem Stichprobenraum sein, hat aber meistens eine bestimmte Struktur. Häufig ist $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ mit $\Theta \subset \mathbb{R}^d$ eine d -dimensionale **parametrische Familie**, Θ heißt dann die **Parametermenge**. Die Daten $x \in \mathcal{X}$ können als Realisierungen einer Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ mit unbekannter Verteilung $\mathcal{L}(X) \in \mathcal{P}$ betrachtet werden. Wird beispielsweise beim zehnfachen Münzwurf nur die Anzahl der „Kopf“-Würfe beobachtet, so könnte man

$$\mathcal{X} = \{0, 1, \dots, 10\}, \quad \mathcal{A} = \mathcal{P}(\mathcal{X}), \quad \Theta = [0, 1], \quad P_\theta = \text{Bin}(10, \theta)$$

wählen. Klar: Die Beobachtung $x = 8$ lässt die exakte Bestimmung des unbekanntes Parameters θ nicht zu – auf der Basis von zufälligen Beobachtungen lassen sich i.A. keine absolut sicheren (nicht-trivialen) Schlüsse ziehen („you can't make a silk purse out of a sow's ear“).

Einen besonders wichtigen Spezialfall der obigen Situation erhält man, wenn die Daten durch unabhängige Wiederholungen eines Zufallsexperiments gewonnen werden, also $x = (x_1, \dots, x_n)$ gilt, wobei x_i das Ergebnis der i -ten Wiederholung ist. Man spricht dann von (den Werten) einer **Stichprobe vom Umfang n** aus einer Verteilung.

Wir betrachten die drei hauptsächlichen statistischen Verfahren: Schätzer, Tests und Konfidenzbereiche.

7.2 Schätztheorie

Ein **Schätzer** (auch **Schätzfunktion**) ist eine Abbildung $\hat{\theta} : \mathcal{X} \rightarrow \Theta$, die jeder Beobachtung x einen Schätzwert $\hat{\theta} = \hat{\theta}(x)$ für den unbekanntes Parameter θ zuordnet. Im Münzwurfbeispiel ist $\hat{\theta} := \frac{x}{10}$ ein nahe liegender Schätzer.

Wie erhält man (gute) Schätzfunktionen? Ein plausibles und sehr wichtiges Prinzip besteht darin, dass man den Wert $\hat{\theta}$ wählt, unter dem die Beobachtung x die größte (infinitesimale) Wahrscheinlichkeit hat. Dies ist die **Maximum-Likelihood-Methode**.

Definition 7.1

Es sei $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf dem messbaren Raum $(\mathcal{X}, \mathcal{A})$. Wir setzen voraus, dass entweder \mathcal{X} abzählbar ist und dann $\mathcal{A} = \mathfrak{P}(\mathcal{X})$ gilt (der diskrete Fall) oder dass $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$ gilt und dann jedes P_θ eine Dichtefunktion hat (der stetige Fall). Im diskreten Fall sei $p(\cdot \mid \theta)$ die Massenfunktion zu P_θ , im stetigen Fall $f(\cdot \mid \theta)$ eine (geeignete) Dichtefunktion zu P_θ . Dann heißt

$$l(\cdot | x) : \Theta \rightarrow \mathbb{R}, \quad l(\theta | x) := \begin{cases} p(x | \theta), & \text{im diskreten,} \\ f(x | \theta), & \text{im stetigen Fall,} \end{cases}$$

die **Likelihood-Funktion** bei Vorliegen der Beobachtung x . $\log l(\cdot | x)$ ist die **Log-Likelihood-Funktion**. Hat $\hat{\theta} : \mathcal{X} \rightarrow \Theta$ die Eigenschaft

$$l(\hat{\theta}(x) | x) = \sup \{ l(\theta | x) \mid \theta \in \Theta \} \text{ f\u00fcr alle } x \in \mathcal{X},$$

so nennen wir $\hat{\theta}$ einen **Maximum-Likelihood-Sch\u00e4tzer** f\u00fcr θ .

Bei einer Stichprobe vom Umfang n aus einer Verteilung mit Massenfunktion $p(\cdot | \theta)$ bzw. Dichtefunktion $f(\cdot | \theta)$ erh\u00e4lt man mit den Resultaten aus Abschnitt 5.7

$$l(\theta | x) = l(\theta | x_1, \dots, x_n) = \begin{cases} \prod_{i=1}^n p(x_i | \theta), & \text{im diskreten Fall,} \\ \prod_{i=1}^n f(x_i | \theta), & \text{im absolutstetigen Fall.} \end{cases}$$

Es k\u00f6nnen nat\u00fcrlich allerlei Schwierigkeiten auftreten, beispielsweise wird das Supremum m\u00f6glicherweise nicht angenommen oder ist nicht eindeutig.

Beispiel 7.2

Ein Zufallsexperiment, in dem ein bestimmtes Ereignis A die Wahrscheinlichkeit θ hat, wird n -mal unabh\u00e4ngig wiederholt, θ ist zu sch\u00e4tzen.

Schreiben wir 1 f\u00fcr das Eintreten von A und sonst 0, so sind die gewonnenen Daten Elemente von $\mathcal{X} = \{0, 1\}^n$ und als Klasse von m\u00f6glichen Verteilungen ergibt sich $\mathcal{P} = \{P_\theta \mid 0 \leq \theta \leq 1\}$, wobei zu P_θ die Massenfunktion

$$p((x_1, \dots, x_n) | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^k (1 - \theta)^{n-k}$$

mit $k := \#\{1 \leq i \leq n \mid x_i = 1\}$ geh\u00f6rt. Zu gegebener Zahl k von Erfolgen erh\u00e4lt man also die Likelihood-Funktion $l(\theta) = \theta^k (1 - \theta)^{n-k}$. Wir betrachten die Randf\u00e4lle separat: Bei $k = 0$ erh\u00e4lt man das (eindeutige, globale) Maximum in $\hat{\theta} = 0$, bei $k = n$ in $\hat{\theta} = 1$. In den F\u00e4llen $k \in \{1, \dots, n-1\}$ ist $l(0) = l(1) = 0$, $l(\theta | x) > 0$ auf $0 < \theta < 1$ und das Maximum kann \u00fcber die Ableitung der Log-Likelihood-Funktion gefunden werden: Mit

$$\frac{\partial}{\partial \theta} \log l(\theta) = -\frac{n-k}{1-\theta} + \frac{k}{\theta}$$

f\u00fchrt dies auf den Maximum-Likelihood-Sch\u00e4tzer $\hat{\theta} = \frac{k}{n}$. Es ist auch intuitiv nahe liegend, die unbekannte Wahrscheinlichkeit von A durch die relative H\u00e4ufigkeit des Eintretens von A zu sch\u00e4tzen.

Beispiel 7.3

Als Beispiel f\u00fcr eine stetige Situation mit mehrdimensionalem Parameterraum betrachten wir eine Stichprobe X_1, \dots, X_n aus der Normalverteilung $N(\mu, \sigma^2)$ mit unbekanntem $\mu \in \mathbb{R}$ und unbekanntem $\sigma^2 > 0$. Wir haben

$$f_{X_i}(x_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right),$$

erhalten also als gemeinsame Dichte in $x = (x_1, \dots, x_n)$

$$f(x | \mu, \sigma^2) = \prod_{i=1}^n f_{X_i}(x_i | \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^n \frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

und damit

$$\log l(\mu, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

F\u00fcr jedes feste $\sigma^2 > 0$ wird dies als Funktion von μ durch den Stichprobenmittelwert $\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i$ maximiert. Die Funktion

$$\sigma^2 \rightarrow -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

wiederum wird maximal in $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$. Damit erh\u00e4lt man die Maximum-Likelihood-Sch\u00e4tzer

$$\hat{\mu} = \bar{x}_n, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Weitere Beispiele werden in den \u00dcbungen besprochen.

Wie beurteilt man die Qualität von Schätzfunktionen? Unser formales Modell geht von einem „Hintergrundwahrscheinlichkeitsraum“ (Ω, \mathcal{A}, P) aus. Die beobachteten Daten x werden als Werte (Realisierungen) einer Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ betrachtet (also: Großbuchstaben stehen für die Abbildung selbst, kleine Buchstaben für ihre Werte – eine Konvention, die wir allerdings nicht stets einhalten werden ...). Die Verteilung $\mathcal{L}(X)$ von X ist ein unbekanntes Element P von $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. Schätzfunktionen sind Abbildungen vom Datenraum \mathcal{X} in den Parameterraum Θ . Im Falle $\Theta \subset \mathbb{R}$ ist $\hat{\theta}(X)$ in der Regel messbar (wir setzen dies in Zukunft stillschweigend voraus), also eine Zufallsvariable, deren Erwartungswert die Lage der Verteilung des Schätzers beschreibt. Verteilung und damit auch Erwartungswert hängen natürlich von der unbekanntem Verteilung von X ab: Wir schreiben $E_\theta \hat{\theta}(X)$ oder kurz $E_\theta \hat{\theta}$ für den Erwartungswert von $\hat{\theta}(X)$ unter der Voraussetzung, dass $\mathcal{L}(X) = P_\theta$ gilt, also θ der wahre Parameter ist.

Ist $\Theta \subset \mathbb{R}$ oder betrachtet man allgemeiner eine reellwertige Parameterfunktion $g(\theta)$, so kann man die Differenz $\hat{\theta} - \theta$ bzw. $g(\hat{\theta}) - g(\theta)$ bilden. Wünschenswerte Eigenschaften eines Schätzers beziehen sich darauf, dass diese Differenz – die ja eine Zufallsgröße ist – in irgendeinem Sinne klein ist.

Definition 7.4

Es sei $\hat{\eta}$ ein (messbarer) Schätzer für eine reellwertige Parameterfunktion $\eta = g(\theta)$. Wir setzen voraus, dass die im Folgenden verwendeten Erwartungswerte existieren.

- (i) Der Schätzer $\hat{\eta}$ heißt **erwartungstreu** (Englisch: **unbiased**) für $\eta = g(\theta)$, wenn gilt:

$$E_\theta \hat{\eta} = g(\theta) \text{ für alle } \theta \in \Theta,$$

die Differenz $E_\theta \hat{\eta} - g(\theta)$ ist der **systematische Fehler** oder **Bias** von $\hat{\eta}$.

- (ii) Die **mittlere quadratische Abweichung** $\text{MSE}(\cdot; \hat{\eta})$ von $\hat{\eta}$ wird definiert durch

$$\text{MSE}(\theta; \hat{\eta}) := E_\theta (\hat{\eta} - g(\theta))^2.$$

(MSE ist die Abkürzung für „mean squared error“).

Bei einem erwartungstreuen Schätzer ist der mittlere quadratische Fehler offensichtlich gleich der Varianz. Allgemein gilt

$$\text{MSE}(\theta; \hat{\theta}) = (E_\theta \hat{\theta} - \theta)^2 + \text{var}_\theta(\hat{\theta}).$$

Beispiel 7.5

Es seien $\mathcal{X} = \{0, \dots, n\}$, $\Theta = (0, 1)$ und $P_\theta = \text{Bin}(n, \theta)$. (Dies ist die aus Beispiel 7.2 bekannte Situation, wenn man dort nur die Anzahl k der Erfolge festhält.) Der Schätzer $\hat{\theta} = \frac{X}{n}$ ist offensichtlich erwartungstreu, denn X hat unter P_θ den Erwartungswert $n\theta$. Als mittleren quadratischen Fehler erhält man

$$\text{MSE}(\theta; \hat{\theta}) = \text{var}_\theta(\hat{\theta}) = \frac{1}{n^2} \text{var}_\theta(X) = \frac{1}{n^2} n\theta(1 - \theta) = \frac{\theta(1 - \theta)}{n}.$$

Man kann zeigen, dass dieser Schätzer unter allen erwartungstreuen Schätzern für θ gleichmäßig in $\theta \in (0, 1)$ die kleinste mittlere quadratische Abweichung hat. (Dies gilt sogar im Rand: im Falle $\theta = 0$, $\theta = 1$ hat $\hat{\theta}$ den MSE 0, was nicht zu unterbieten ist.)

Was passiert, wenn man auch nichterwartungstreue Schätzer in die Konkurrenz aufnimmt? Klar: der „entartete“ Schätzer $\theta \equiv \theta_0$ für ein festes $\theta_0 \in \Theta$ hat MSE 0 in θ_0 (eine stehen gebliebene Uhr zeigt zwei Mal am Tag die genaue Zeit an). Interessanter ist der Schätzer $\hat{\theta}_A := \frac{X+1}{n+2}$, der vermeidet, dass die Wahrscheinlichkeit durch 0 bzw. 1 geschätzt wird, wenn das interessierende Ereignis gar nicht bzw. immer eintritt. Man erhält

$$E_\theta \hat{\theta}_A = \frac{1}{n+2} (E_\theta X + 1) = \frac{n\theta + 1}{n+2},$$

insbesondere ist $\hat{\theta}_A$ nicht erwartungstreu. Eine etwas längere Rechnung (oder Maple) liefert

$$E_\theta (\hat{\theta}_A - \theta)^2 = \frac{1 + (n-4)\theta(1-\theta)}{(n+2)^2}$$

und ein Vergleich der Funktionen zeigt, dass keiner der beiden Schätzer einen gleichmäßig kleineren mittleren quadratischen Fehler hat als der andere.

7.3 Tests

Es sei wieder \mathcal{P} eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$. Oft soll anhand der Daten entschieden werden, ob die tatsächliche Verteilung P ein einer vorgegebenen Teilfamilie \mathcal{P}_0 von \mathcal{P} liegt, d.h. man will die Hypothese $H : P \in \mathcal{P}_0$ testen. Bei einer parametrisierten Familie $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ lässt sich die Teilfamilie über eine Teilmenge Θ_0 des Parameterraums Θ charakterisieren. Die Hypothese lautet dann

$H : \theta \in \Theta_0$, wobei θ für den „wahren“ Parameter steht. $K : \theta \in \Theta - \Theta_0$ (bzw. $K : \mathcal{P} - \mathcal{P}_0$) bezeichnet man als **Alternative**. Man kann H und K auch als Zerlegung von Θ auffassen. H heißt **einfach** im Falle $\#\mathcal{P}_0 = 1$ bzw. $\#\Theta_0 = 1$ und **zusammengesetzt** sonst. Analoge Bezeichnungen werden auch bei K verwendet.

Definition 7.6

Eine (messbare) Funktion $\phi : \mathcal{X} \rightarrow [0,1]$ heißt (**randomisierte**) **Testfunktion zum Signifikanzniveau α** , kurz: **Test zum Niveau α** , wenn gilt:

$$E_P \phi(X) \leq \alpha \text{ für alle } P \in \mathcal{P}_0.$$

Die Abbildung $P \rightarrow E_P \phi(X)$ ist die **Gütefunktion** oder auch **Operationscharakteristik** des Tests. Im parametrischen Fall ist dies

$$\beta : \Theta \rightarrow [0,1], \beta(\theta) := E_\theta \phi(X).$$

Interpretation: Bei Vorliegen der Beobachtung x wird H mit Wahrscheinlichkeit $\phi(x)$ verworfen, also wird bei einem Test zum Niveau α die Wahrscheinlichkeit für eine irrtümliche Verwerfung der Hypothese nicht größer als α . Für α sind die Werte 0.1, 0.05, 0.01, 0.001 gebräuchlich. Bei Tests geht es also darum, eine vorgegebene Hypothese anhand der Daten entweder zu verwerfen (beachte: „nicht verwerfen“ ist nicht dasselbe wie „als richtig bewiesen“!). In der Regel wird man nichtrandomisierte Tests verwenden, bei denen also ϕ nur die Werte 0 und 1 annimmt. Die Menge $\{x \in \mathcal{X} \mid \phi(x) = 1\}$ ist dann der **Ablehnungsbereich** eines solchen Tests. Dieser wird häufig über eine **Testgröße** (auch: **Teststatistik**) T beschrieben, die die Eigenschaft hat, dass große Werte von T gegen H sprechen. In der Tat liefert eine solche Testgröße gleich eine ganze Familie von nichtrandomisierten Tests ϕ_c über

$$\phi_c(X) = \begin{cases} 1, & T(x) \geq c, \\ 0, & T(x) < c. \end{cases}$$

Man nennt in dieser Situation c den **kritischen Wert**. Bei Tests geht es um nur zwei Entscheidungen: H wird verworfen oder H wird nicht verworfen. Als Folge hiervon gibt es zwei Fehlerarten:

- **Fehler 1. Art:** Die Hypothese wird verworfen, obwohl sie richtig ist.
- **Fehler 2. Art:** Die Hypothese wird nicht verworfen, obwohl sie falsch ist.

Die Wahrscheinlichkeit für eine falsche Entscheidung hängt natürlich von dem unbekanntem wahren Parameter θ ab. Bei einem Test zum Niveau α darf die Wahrscheinlichkeit für einen Fehler 1. Art den Wert α nicht übersteigen. Alle Fehlerwahrscheinlichkeiten lassen sich aus der Gütefunktion ablesen. Man wird nun versuchen, bei einer vorgegebenen Schranke für den Fehler 1. Art einen Test zu finden, bei dem die Wahrscheinlichkeiten für einen Fehler 2. Art möglichst gleichmäßig minimiert werden. Bei einfacher Hypothese und einfacher Alternative (also bei $\#\mathcal{P} = 2$) kann man dieses Optimierungsproblem leicht lösen.

Satz 7.7 (Das Neyman-Pearson-Lemma)

Es sei $\mathcal{P} = \{P_0, P_1\}$ und $\alpha \in (0,1)$. Wir setzen voraus, dass P_0 und P_1 entweder beide diskret sind oder beide eine Dichte haben und schreiben p_0, p_1 für die Massenfunktionen im ersten und f_0, f_1 für die Dichten im zweiten Fall. Dann existieren ein $c \geq 0$ und ein $\gamma \in [0,1]$ mit

$$P_0(p_1 > cp_0) + \gamma P_0(p_1 = cp_0) = \alpha \text{ bzw. } P_0(f_1 > cf_0) + \gamma P_0(f_1 = cf_0) = \alpha$$

im diskreten bzw. stetigen Fall und der **Neyman-Pearson-Test** $\phi : \mathcal{X} \rightarrow [0,1]$,

$$\phi(x) = \begin{cases} 1, & > \\ \gamma, & p_1(x) = cp_0(x) \text{ bzw. } \phi(x) = \begin{cases} 1, & > \\ \gamma, & f_1(x) = cf_0(x) \\ 0, & < \end{cases} \\ 0, & < \end{cases}$$

im diskreten bzw. stetigen Fall ist ein Test zum Niveau α für $H : P = P_0$, der unter allen solchen Tests die kleinste Wahrscheinlichkeit für einen Fehler 2. Art hat.

Beweis

Wir betrachten nur den diskreten Fall. Der Beweis für den stetigen Fall verläuft sehr ähnlich, im Wesentlichen müssen einige Summen durch Integrale ersetzt werden.

Wir können p_0 und p_1 als Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\mathcal{X}, \mathcal{A}, P_0)$ auffassen und erhalten beispielsweise

$$P_0(p_0 > 0) = \sum_{\substack{x \in \mathcal{X} \\ p_0(x) > 0}} p_0(x) = \sum_{x \in \mathcal{X}} p_0(x) = P_0(\mathcal{X}) = 1.$$

Es sei c das $(1 - \alpha)$ -Quantil zur Verteilung von q ,

$$q(x) := \begin{cases} \frac{p_1(x)}{p_0(x)}, & \text{falls } p_0(x) > 0, \\ 0, & \text{sonst.} \end{cases}$$

Aus unseren allgemeinen Betrachtungen zu Quantilfunktionen (Lemma 5.19, Aufgabe 63) folgt dann, dass

$$P_0(q > c) \leq \alpha \leq P_0(q \geq c)$$

gilt. Wir setzen $\gamma := 0$ im Falle $P_0(q = c) = 0$ und

$$\gamma := \frac{\alpha - P_0(q > c)}{P_0(q = c)}$$

sonst. Mit diesen Werten erhält man

$$\begin{aligned} P_0(p_1 > cp_0) + \gamma P(p_1 = cp_0) &= P_0(p_1 > cp_0, p_0 > 0) + \gamma P(p_1 = cp_0, p_0 > 0) \\ &= P_0(q > c) + \gamma P(q = c) = \alpha, \end{aligned}$$

womit der erste Teil der Behauptung bewiesen wäre.

Für den Beweis des zweiten (und interessanteren) Teils sei $\tilde{\phi}$ irgendein Test zum Niveau α für $H : P = P_0$.

Wir setzen

$$A := \{x \in \mathcal{X} \mid \phi(x) > \tilde{\phi}(x)\}, \quad B := \{x \in \mathcal{X} \mid \phi(x) < \tilde{\phi}(x)\}.$$

Auf A ist $\phi > 0$, also $p_1 \geq cp_0$, auf B ist $\phi(x) < 1$, also $p_1 \leq cp_0$. Damit folgt

$$\begin{aligned} E_1\phi(X) - E_1\tilde{\phi}(X) &= \sum_{x \in \mathcal{X}} (\phi(x) - \tilde{\phi}(x)) p_1(x) = \sum_{x \in A} (\phi(x) - \tilde{\phi}(x)) p_1(x) + \sum_{x \in B} (\phi(x) - \tilde{\phi}(x)) p_1(x) \\ &\geq \sum_{x \in A} (\phi(x) - \tilde{\phi}(x)) cp_0(x) + \sum_{x \in B} (\phi(x) - \tilde{\phi}(x)) cp_0(x) \\ &= c \sum_{x \in \mathcal{X}} (\phi(x) - \tilde{\phi}(x)) p_0(x) = c(E_0\phi(X) - E_0\tilde{\phi}(X)) \geq 0, \end{aligned}$$

denn $E_0\phi(X) = \alpha$, $E_0\tilde{\phi}(X) \leq \alpha$. □

Der optimale Test hängt also nur über das Verhältnis $\frac{p_1}{p_0}$ bzw. $\frac{f_1}{f_0}$, den so genannten **Likelihood-Quotienten**, von x ab. Der Ablehnungsbereich entsteht dadurch, dass man die x -Werte mit den größten Likelihood-Quotienten zusammenfasst, soweit dies die Fehlerschranke erlaubt. Dies ist eine auch intuitiv nahe liegende Vorgehensweise.

Beispiel 7.8

Wie in Beispiel 7.2 sei $\mathcal{X} = \{0,1\}^n$,

$$p(x \mid \theta) = \theta^{T(x)} (1 - \theta)^{n - T(x)} \quad \text{mit } T(x) = \sum_{i=1}^n x_i.$$

Wir betrachten zunächst die Familie $\mathcal{P} = \{P_{\theta_0}, P_{\theta_1}\}$ mit $0 < \theta_0 < \theta_1 < 1$ fest. Als Verhältnis der Massenfunktionen ergibt sich

$$\frac{p_1(x)}{p_0(x)} = \left(\frac{1 - \theta_1}{1 - \theta_0}\right)^{n - T(x)} \left(\frac{\theta_1}{\theta_0}\right)^{T(x)}.$$

Wegen $\theta_1 > \theta_0$ ist dies eine streng monoton wachsende Funktion von $T(x)$, d.h. zu jedem c existiert ein \tilde{c} mit der Eigenschaft, dass

$$\begin{array}{ccc} > & & > \\ p_1(x) = cp_0(x) & \iff & T(x) = \tilde{c} \\ < & & < \end{array}$$

für alle $x \in \mathcal{X}$ gilt. Nach dem Neyman-Pearson-Lemma ist also der beste Test für θ_0 gegen θ_1 von der Form

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i > \tilde{c}, \\ \gamma, & \sum_{i=1}^n x_i = \tilde{c}, \\ 0, & \sum_{i=1}^n x_i < \tilde{c}, \end{cases}$$

wobei \tilde{c} und $\gamma \in [0,1]$ bestimmt werden aus

$$P_{\theta_0}\left(\sum_{i=1}^n X_i > \tilde{c}\right) + \gamma P_{\theta_0}\left(\sum_{i=1}^n X_i = \tilde{c}\right) = \alpha.$$

(Die Überlegung, dass streng monoton wachsende Transformationen der Testgröße bei entsprechender Transformation des kritischen Werts den Test unverändert lassen, kann bei Rechnungen sehr hilfreich sein.) Man beachte nun, dass in der Beschreibung des Tests θ_1 nicht mehr auftritt, nur $\theta_1 > \theta_0$ wurde in der Herleitung verwendet. Die Hypothese $H : \theta = \theta_0$ gegen $K : \theta = \tilde{\theta}_1$ würde auf denselben Test führen, wenn nur $\tilde{\theta}_1 > \theta_0$ gilt. Dies zeigt, dass ϕ unter allen Tests zum Niveau α für $H : \theta = \theta_0$ gegen $K : \theta > \theta_0$ gleichmäßig die

Fehlerwahrscheinlichkeiten 2. Art minimiert, ϕ ist also ein **gleichmäßig bester Test** zum Niveau α für $\theta = \theta_0$ gegen $\theta > \theta_0$. Es kommt sogar noch besser: Jeder Test zum Niveau α für $H : \theta = \theta_0$ gegen $K : \theta > \theta_0$. Da $E_\theta \phi$ eine monoton wachsende Funktion von θ ist, hält ϕ auch in dieser größeren Hypothese das Niveau α ein, minimiert also auch in dieser Klasse gleichmäßig die Fehlerwahrscheinlichkeiten zweiter Art. Gelegentlich lassen sich also mit Hilfe des Neyman-Pearson-Lemmas optimale Tests sogar bei zusammengesetzten Hypothesen und Alternativen bestimmen.

Beispiel 7.9

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig und exponentialverteilt mit unbekanntem Parameter $\theta > 0$. Anhand der Realisierungen soll

$$H : \theta = \theta_0 \text{ gegen } K : \theta = \theta_1$$

getestet werden. Wir betrachten den Fall $\theta_1 > \theta_0$. Die Dichtefunktion zu $X = (X_1, \dots, X_n)$ ist

$$f(x | \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta s_n} \text{ mit } s_n := \sum_{i=1}^n x_i.$$

Wie in Beispiel 7.8 ist für den optimalen Test nur die Realisierung s_n der Summe $S_n = \sum_{i=1}^n X_i$ der Zufallsvariablen relevant. Satz 7.7 führt mit $f_i(x) = f(x | \theta_i)$, $i = 0, 1$, auf die Testgröße

$$\frac{f_1(x)}{f_0(x)} = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-(\theta_1 - \theta_0)s_n}.$$

Wegen $\theta_1 > \theta_0$ ist dies eine streng monoton fallende Funktion von s_n , der Neyman-Pearson-Test also von der Form

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i < \tilde{c} \\ \gamma, & \sum_{i=1}^n x_i = \tilde{c} \\ 0, & \sum_{i=1}^n x_i > \tilde{c} \end{cases}$$

wobei wieder \tilde{c} und $\gamma \in [0, 1]$ bestimmt werden aus

$$P_0(S_n < \tilde{c}) + \gamma P_0(S_n = \tilde{c}) = \alpha.$$

Unter P_0 ist S_n $\Gamma(n, \theta_0)$ -verteilt, insbesondere gilt also $P_0(S_n = c) = 0$ für alle $c \in \mathbb{R}$ und eine Randomisierung wird nicht benötigt. Der zweite Parameter der Gamma-Verteilung präsentiert nur eine Umskalierung (siehe auch der Beginn des Beweises zu Satz 6.10), insbesondere ist $\theta_0 S_n$ unter der Hypothese $\Gamma(n, 1)$ -verteilt. Einer Tafel für die unvollständige Gammafunktion entnimmt man den Wert c mit

$$\int_0^{\theta_0 c} x^{n-1} e^{-x} dx = \alpha \Gamma(n)$$

(alternativ kann beispielsweise im Computeralgebra-Programm Maple die linke Seite mit `GAMMA(n, c)` berechnet werden), dieses c ist der kritische Wert bei Signifikanzniveau α . Wie im letzten Beispiel ergibt sich auch hier für alle Alternativwerte $\theta_1 > \theta_0$ derselbe Test und die Wahrscheinlichkeit für eine Ablehnung wird mit fallendem θ kleiner, d.h. der Neyman-Pearson-Test ist sogar der gleichmäßig beste Test zum Niveau α für $H : \theta \leq \theta_0$ gegen $K : \theta > \theta_0$.

Hat man ganz allgemein eine parametrische Familie $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ von für die Beobachtungen in Frage kommenden Verteilungen (durchaus mit mehrdimensionalem Parameterraum Θ), so lassen sich Hypothese und Alternative durch Teilmengen von Θ beschreiben, d.h. man möchte

$$H : \theta \in \Theta_0 \text{ gegen } K : \theta \in \Theta_1 := \Theta - \Theta_0$$

testen. Sind die Verteilungen P_θ , $\theta \in \Theta$, alle diskret oder alle stetig, so machen die bisher behandelten Ideen das folgende Vorgehen plausibel: Schätze θ durch die Werte, die die Likelihood-Funktion $\theta \mapsto l(\theta | x)$ (wobei wieder $l(\theta | x) = p(x | \theta)$ im diskreten und $l(\theta | x) = f(x | \theta)$ im stetigen Fall) auf Θ_0 bzw. Θ_1 maximieren und verwende den dann erhaltenen Dichtequotienten als Testgröße. Dies führt auf den **Likelihood-Quotienten-Test** (oder kurz LQ-Test), der ablehnt, wenn die Testgröße

$$T_{\text{LQ}(x)} = \frac{\sup_{\theta \in \Theta_1} l(\theta | x)}{\sup_{\theta \in \Theta_0} l(\theta | x)}$$

einen durch die Forderung

$$\sup_{\theta \in \Theta_0} P_\theta(T \geq c) = \alpha$$

festgelegten kritischen Wert c übersteigt (man kann auch hier wieder randomisieren, wenn beispielsweise im diskreten Fall ein solches c nicht existiert).

Beispiel 7.10

Wir gehen aus von einer Stichprobe X_1, \dots, X_n aus einer Normalverteilung $N(\mu, \sigma^2)$ mit unbekanntem $\mu \in \mathbb{R}$, $\sigma^2 > 0$ und wollen

$$H : \mu = \mu_0 \text{ gegen } K : \mu \neq \mu_0$$

zum Niveau α testen (μ_0 und α sind vorgegeben). Dies passt in den oben beschriebenen Rahmen, mit $\theta = (\mu, \sigma^2)$,

$$\Theta = \mathbb{R} \times (0, \infty), \Theta_0 = \{\mu_0\} \times (0, \infty), \Theta_1 = (\mathbb{R} \setminus \{\mu_0\}) \times (0, \infty).$$

Zur Bestimmung des LQ-Tests müssen wir die Funktion

$$l(\theta | x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

auf Θ_1 bzw. Θ_0 maximieren. Da diese Funktion stetig ist und Θ_1 dicht liegt in Θ , gilt

$$\sup_{\theta \in \Theta_1} l(x | \theta) = \sup_{\theta \in \Theta} l(x | \theta)$$

und mit den Rechnungen aus Beispiel 7.3 (die ML-Schätzer sind $\hat{\mu} = \bar{x}_n$ und $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$) folgt

$$\sup_{\theta \in \Theta_1} l(x | \theta) = (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}.$$

Zur Bestimmung des Nenners der Testgröße muss l auf Θ_0 maximiert werden, wodurch $\mu = \mu_0$ festgelegt ist. Das Maximum der Funktion

$$\sigma^2 \mapsto (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2\right)$$

wird in $\tilde{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$ angenommen, also gilt

$$\sup_{\theta \in \Theta_0} l(x | \theta) = (2\pi\tilde{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}$$

und man erhält insgesamt die Testgröße

$$T_{LQ}(x) = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{\sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x}_n)^2}\right)^{\frac{n}{2}} = \left(1 + \frac{(\bar{x}_n - \mu_0)^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}}.$$

Dies ist offensichtlich eine streng monoton wachsende Funktion von

$$T(x) = \frac{|\bar{x}_n - \mu_0|}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}},$$

man erhält also denselben Test, wenn man als Testgröße T verwendet. Dies ergibt den **zweiseitigen t -Test** zur Hypothese $\mu = \mu_0$ bei Stichproben aus der Normalverteilung mit unbekannter Varianz.

Zur praktischen Ausführbarkeit muss allerdings noch die Verteilung der Testgröße unter der Hypothese bestimmt werden. Da die Hypothese nun aus mehr als einem Wert besteht, ist zunächst nicht einmal klar, ob nicht sogar mehrere Verteilungen, abhängig von dem unbekanntem σ^2 , erscheinen. Zumindest diese Frage können wir bereits jetzt beantworten: Sind X_1, \dots, X_n unabhängig und $N(\mu_0, \sigma^2)$ -verteilt, so sind die Zufallsvariablen Y_1, \dots, Y_n mit $Y_i := \frac{X_i - \mu_0}{\sigma}$ unabhängig (Satz 5.28) und $N(0, 1)$ -verteilt (Lemma 5.23 (iii)). Man überprüft leicht, dass mit $\bar{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i$

$$T(X_1, \dots, X_n) = \frac{|\bar{Y}_n|}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$$

gilt. Auf der rechten Seite sind μ_0 und σ^2 verschwunden, $T(X)$ hat also unter allen Verteilungen, für die die Hypothese richtig ist, eine feste Verteilung. Diese hängt nicht von μ_0 ab. Es stellt sich heraus, dass diese Größe, nach Beseitigung der Betragsstriche, die **t -Verteilung mit $n - 1$ Freiheitsgraden** hat. Dies ist die Verteilung mit der Dichte

$$x \mapsto \frac{1}{\sqrt{\pi(n-1)}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \quad -\infty < x < \infty,$$

(genauer in der Vorlesung Stochastik II).

Bemerkung 7.11

- (i) Klassische Tests laufen in den folgenden Schritten ab: Zunächst wird die Hypothese festgelegt, dann eine geeignete Testgröße T gewählt. (Grob gilt, dass große Werte von T gegen die Hypothese sprechen sollen. Die Testgröße bestimmt letztlich, welche Abweichungen von der Hypothese der Test bevorzugt entdeckt. Die Wahl sollte daher von der Alternative abhängen.) Bei nichtrandomisierten Tests mit einem

Ablehnungsbereich von der Form $\{c \in \mathcal{X} \mid T(x) \geq c\}$ geht das Signifikanzniveau α nur über den kritischen Wert $c = c(\alpha)$ ein. Dieses Signifikanzniveau wird nun vor Ausführung des Experiments festgelegt und nach Erhebung der Daten x und Berechnung von $T(x)$ die Entscheidung (Ablehnung/keine Ablehnung) festhalten. Bei Ablehnung der Hypothese $H : \mu \leq 0$ beispielsweise in der Form „die Aussage $\mu > 0$ ist statistisch auf dem Niveau α abgesichert“. Hieraus geht nicht hervor, ob nicht vielleicht sogar für ein kleineres α auch eine Ablehnung erzielt worden wäre oder ob nicht ein weniger stringentes α doch eine Ablehnung geliefert hätte. Man gibt daher häufig anstelle eines Signifikanzniveaus den p -Wert der Beobachtung x an: Dies ist der kleinste α -Wert, der noch zu einer Ablehnung der Hypothese geführt hätte. Der p -Wert ist somit die maximale Wahrscheinlichkeit, unter der Hypothese, dass die Testgröße mindestens so groß ist wie der tatsächlich beobachtete Wert. Der Übergang von einem festgelegten Signifikanzniveau zu p -Werten vermeidet einen Informationsverlust und überlässt letztlich dem Anwender die Wahl des Signifikanzniveaus.

- (ii) Wie aus dem Beweis zu Satz 7.7 hervorgeht, dient Randomisierung der Ausschöpfung der zugelassenen Fehlerwahrscheinlichkeit 1. Art. Als konkretes Beispiel betrachten wir die Hypothese, dass „Kopf“ bei einer gegebenen Münze höchstens mit Wahrscheinlichkeit $\frac{1}{2}$ erscheint. Soll dies durch zehnmaligen Wurf überprüft werden, so führt Beispiel 7.8 auf die Anzahl T der „Kopf“-Würfe als Testgröße. Es gilt $P_{0.5}(T \geq 9) = 0.0108\dots$, $P_{0.5}(T \geq 8) = 0.0546\dots$, also ist der beste Test zum Niveau $\alpha = 0.05$ wegen

$$\gamma = \frac{\alpha - P_{0.5}(T \geq 9)}{P_{0.5}(T = 8)} = 0.89\dots$$

von der Form

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i > 8 \\ 0.89\dots, & \sum_{i=1}^n x_i = 8 \\ 0, & \sum_{i=1}^n x_i < 8 \end{cases}$$

Wird nun die Münze zehnmal geworfen, so ist man nur im Falle $T < 8$ oder $T > 8$ fertig: Bei $T = 8$ wird ein weiteres, vom bisherigen Geschehen unabhängiges Zufallsexperiment ausgeführt, in dem mit Wahrscheinlichkeit 0.89... ein bestimmtes Ereignis A eintritt. Erscheint tatsächlich A , so wird die Hypothese abgelehnt, sonst nicht.

Randomisierung wird von vielen Praktikern als mathematische Spielerei angesehen. Im Sinne von Teil (i) würde man beim Erhalt von acht Mal „Kopf“ stattdessen angeben, dass man mit diesem Resultat bei $\alpha \geq 0.0108\dots$ eine Ablehnung erhalten hätte.

- (iii) Tests sind „unsymmetrisch“: Man hat nur für eine Fehlerart eine obere Schranke (nämlich das Signifikanzniveau, für das irrtümliche Ablehnen der Hypothese). Dies ist bei der Wahl der Hypothese zu berücksichtigen, denn das Nichtverwerfen einer Hypothese ist gewissermaßen „nicht interessant“. Wieder einmal lässt sich dies im Englischen besonders kurz und elegant formulieren: „absence of evidence is not evidence of absence“.

7.4 Konfidenzbereiche

Die Daten x seien wieder Realisierungen einer Zufallsgröße X , deren Verteilung ein unbekanntes Element einer vorgegebenen Familie $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ ist. Neben dem diskreten Schätzen des Parameters θ und dem Testen von Aussagen über θ ist die Konstruktion von Konfidenzbereichen das dritte Standardverfahren der Statistik, man spricht hier auch von **Bereichsschätzern**. Jedem $x \in \mathcal{X}$ wird hierbei eine Teilmenge $C(x)$ des Parameterraums Θ zugeordnet. Gilt

$$P_\theta(C(X) \ni \theta) \geq 1 - \alpha \text{ für alle } \theta \in \Theta,$$

so nennt man $C(X)$ ein $100(1 - \alpha)$ **-prozentiges Konfidenzgebiet für θ** . Natürlich muss $\{x \in \mathcal{X} \mid C(x) \ni \theta\}$ für alle $\theta \in \Theta$ eine messbare Teilmenge des Stichprobenraums sein. Ist $C(X)$ ein Intervall, so spricht man naheliegenderweise von einem **Konfidenzintervall**, bei $C(X) = (-\infty, \bar{\theta}(X)]$ nennt man $\bar{\theta}(X)$ eine **obere Konfidenzschranke zum Niveau $1 - \alpha$** etc.. Für α sind wieder die Werte 0.1, 0.05, 0.01, 0.001 gebräuchlich. Wie bei Schätzern ist man auch hier u.U. nicht an dem gesamten Parameter θ , sondern nur an einem Teil $\eta = g(\theta)$ interessiert. Die Ausdehnung dieser Konzepte auf solche Parameterfunktionen dürfte klar sein.

Beispiel 7.12

Ist X_1, \dots, X_n eine Stichprobe aus der Exponentialverteilung mit unbekanntem Parameter $\theta > 0$, so sind die Zufallsvariablen $\theta X_1, \dots, \theta X_n$ unabhängig und exponentialverteilt mit Parameter 1 und nach Aufgabe 81 ist $Y := \min\{\theta X_1, \dots, \theta X_n\}$ dann exponentialverteilt mit Parameter n . Es gilt also

$$P_\theta\left(\theta \geq \frac{z}{\min\{X_1, \dots, X_n\}}\right) = P_\theta(\theta \min\{X_1, \dots, X_n\} \geq z) = e^{-nz}$$

für alle $\theta \in \Theta = (0, \infty)$ und alle $z > 0$. Wählt man nun z in Abhängigkeit vom Stichprobenumfang n und dem gewählten Konfidenzniveau α so, dass $e^{-nz} = 1 - \alpha$ gilt, so erhält man mit

$$\underline{\theta}(X) = \frac{-\frac{1}{n} \log(1 - \alpha)}{\min\{X_1, \dots, X_n\}}$$

eine $100(1 - \alpha)\%$ -Konfidenzschranke für θ .

Ein Konfidenzbereich $C(X)$ ist eine zufällige Menge, die den unbekanntem Parameter θ mit einer bestimmten Wahrscheinlichkeit, dem Konfidenzniveau, überdeckt (enthält). Setzt man für X die Daten x ein, so erhält man eine Realisierung des Konfidenzbereichs, die den unbekanntem Parameter entweder enthält oder nicht enthält. Ergibt sich beispielsweise das Intervall $[2.5, 3.1]$, so wird häufig, **aber falsch**, formuliert: „das Intervall $[2.5, 3.1]$ enthält den unbekanntem Parameter θ mit Wahrscheinlichkeit 0.95“. Ein ähnliches Missverständnis ist auch bei Anwendern statistischer Tests weit verbreitet: Wird eine Hypothese auf dem Niveau α abgelehnt, so heißt dies nicht, dass sie mit Wahrscheinlichkeit $1 - \alpha$ falsch ist. Zur Verdeutlichung betrachten wir einen analogen Sachverhalt beim Würfelwurf: Die Augenzahl X nimmt mit Wahrscheinlichkeit $\frac{1}{6}$ den Wert 2 an – wurde geworfen und beispielsweise der Wert $x = 5$ erhalten, so heißt dies nicht, dass 5 mit Wahrscheinlichkeit $\frac{1}{6}$ gleich 2 ist! Es bleibt dem Experimentator natürlich unbenommen, Konfidenzintervalle mit subjektiven Wahrscheinlichkeiten im Sinne von Abschnitt 1 dieser Vorlesung zu verbinden und somit zu einer Aussage der Form „die Stärke meines Glaubens daran, dass das Intervall $[2.5, 3.1]$ den unbekanntem Parameter θ enthält, hat den Wert 0.9“ zu kommen.

Zwischen den Ablehnungsbereichen von Tests einfacher Hypothesen und Konfidenzbereichen besteht ein gelegentlich nützlicher Zusammenhang.

Satz 7.13

Für jedes $\theta_0 \in \Theta$ sei $A(\theta_0) \subset \mathcal{X}$ Ablehnungsbereich eines nichtrandomisierten Tests zum Niveau α für $H: \theta = \theta_0$ gegen $K: \theta \neq \theta_0$. Dann ist C ,

$$C(X) := \{\theta \in \Theta \mid X \notin A(\theta)\},$$

ein Konfidenzbereich zum Niveau $1 - \alpha$ für θ .

Beweis

Die Aussage ergibt sich sofort aus

$$P_\theta(C(X) \ni \theta) = P_\theta(X \notin A(\theta)) = 1 - P_\theta(X \in A(\theta)) \geq 1 - \alpha. \quad \square$$

Eine weitere im Zusammenhang mit der Konstruktion von Konfidenzbereichen sehr nützliche Idee ist die des **Pivots** (Englisch für „Drehpunkt“): Hat man eine Funktion $h: \mathcal{X} \times \Theta \rightarrow \mathcal{Y}$ mit den Eigenschaften, dass erstens die Verteilung Q von $h(X, \theta)$ bei $\mathcal{L}(X) = P_\theta$ nicht von θ abhängt und dass zweitens Mengen der Form $\{x \in \mathcal{X} \mid h(x, \theta) \in A\}$ nach θ aufgelöst werden können (hier hat man oft eine Art „Drehung“), so erhält man durch $C(X)$ mit $C(x) := \{\theta \in \Theta \mid h(x, \theta) \in A\}$ einen $100(1 - \alpha)\%$ -Konfidenzbereich, wenn man für A eine Menge mit $Q(A) \geq 1 - \alpha$ wählt. In Beispiel 7.12 ist $h(x, \theta) := \theta \min\{x_1, \dots, x_n\}$ ein solcher Pivot, ein anderer (und besserer) ist $h(x, \theta) := \theta \sum_{i=1}^n x_i$.

Der Zusammenhang von Tests und Konfidenzintervallen, die Idee des Pivots und schließlich der Umgang mit Parameterfunktionen werden im folgenden Beispiel illustriert, bei dem es um Konfidenzbereiche für den Mittelwert bei normalverteilten Größen geht.

Beispiel 7.14

Es sei X_1, \dots, X_n eine Stichprobe aus $N(\mu, \sigma^2)$, wobei sowohl μ als auch $\sigma^2 (> 0)$ als unbekannt betrachtet werden. Es seien wieder

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

der Stichprobenmittelwert und die Stichprobenvarianz. Bereits beim t -Test in Beispiel 7.10 wurde verwendet, dass $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$ eine t -Verteilung mit $n - 1$ Freiheitsgraden hat. Bezeichnet wieder $t_{n-1;1-\alpha}$ das $(1 - \alpha)$ -Quantil zu dieser Verteilung, so gilt daher

$$P_{\mu, \sigma^2} \left(\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq t_{n-1;1-\alpha} \right) = 1 - \alpha \text{ für alle } \mu \in \mathbb{R}, \sigma^2 > 0.$$

Unter Verwendung der einfachen Umformung

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq t_{n-1;1-\alpha} \iff \mu \geq \bar{X}_n - \frac{1}{\sqrt{n}} S_n t_{n-1;1-\alpha}$$

(dies entspricht der oben erwähnten Auflösung oder „Drehung“) folgt hieraus, dass

$$\underline{\mu} = \bar{X}_n - \frac{1}{\sqrt{n}} S_n t_{n-1;1-\alpha}$$

eine $100(1 - \alpha)$ %-Konfidenzunterschranke für μ ist. Ganz analog sieht man, dass

$$\left[\bar{X}_n - \frac{1}{\sqrt{n}} S_n t_{n-1;1-\frac{\alpha}{2}}, \bar{X}_n + \frac{1}{\sqrt{n}} S_n t_{n-1;1-\frac{\alpha}{2}} \right]$$

ein $100(1 - \alpha)$ %-Konfidenzintervall für μ ist.

Die obigen Beispiele beziehen sich auf stetige Verteilungen. In der Tat sind Konfidenzintervalle bei diskreten Verteilungen oft ein recht mühsames Geschäft. Wir bringen ein Beispiel, Konfidenzintervalle für Wahrscheinlichkeiten, bei dem asymptotische Überlegungen zu einer Vereinfachung führen.

Beispiel 7.15

Es seien wieder einmal X_1, X_2, \dots unabhängig und $\text{Bin}(1, \theta)$ -verteilt mit unbekanntem $\theta \in (0, 1)$. Wir verwenden $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ als Schätzer für θ (siehe auch Beispiel 7.2). Nach dem in Abschnitt 6 diskutierten zentralen Grenzwertsatz gilt mit $S_n = \sum_{i=1}^n X_i = n\bar{X}_n$

$$\lim_{n \rightarrow \infty} P_\theta \left(a \leq \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq b \right) = \Phi(b) - \Phi(a),$$

wobei wieder Φ die Verteilungsfunktion zur Standardnormalverteilung bezeichnet. Ist u_α das zugehörige α -Quantil, also $\Phi(u_\alpha) = \alpha$, so folgt mit $b := u_{1-\frac{\alpha}{2}}$, $a := -b$ bei großem n

$$P_\theta \left(-u_{1-\frac{\alpha}{2}} \leq \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq u_{1-\frac{\alpha}{2}} \right) \approx 1 - \alpha,$$

denn $\Phi(-u_{1-\frac{\alpha}{2}}) = 1 - \Phi(u_{1-\frac{\alpha}{2}}) = 1 - (1 - \frac{\alpha}{2}) = \frac{\alpha}{2}$. Wegen $\theta(1 - \theta) \leq \frac{1}{4}$ gilt

$$-u_{1-\frac{\alpha}{2}} \leq \frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq u_{1-\frac{\alpha}{2}} \implies \bar{X}_n - \frac{u_{1-\frac{\alpha}{2}}}{2\sqrt{n}} \leq \theta \leq \bar{X}_n + \frac{u_{1-\frac{\alpha}{2}}}{2\sqrt{n}},$$

also ergibt sich

$$\left[\bar{X}_n - \frac{1}{2\sqrt{n}} u_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{1}{2\sqrt{n}} u_{1-\frac{\alpha}{2}} \right]$$

als (asymptotisches, konservatives) $100(1 - \alpha)$ %-Konfidenzintervall für θ .

Bemerkenswert ist hier, dass die Länge des Intervalls mit $\frac{1}{\sqrt{n}}$ fällt. Für eine weitere Dezimalstelle müsste man also den Stichprobenumfang verhundertfachen. Numerisches Beispiel: Soll bei einer Wahl ein Konfidenzintervall für die Anzahl der Stimmen einer Partei von der Form „Prozentsatz in Stichprobe $\pm 1\%$ “ auf dem Niveau 0.95 erhalten werden, so muss

$$\frac{1}{2\sqrt{n}} u_{0.975} \leq 0.01$$

gelten. Mit $u_{0.975} = 1.96\dots$ ergibt sich $n \geq 9604$. Bei $\pm 0.1\%$ würde man schon $n \geq 960400$ benötigen. (Bei Umfragen werden in der Regel komplizierte Verfahren verwendet, die von zusätzlicher Information, beispielsweise über das Wahlverfahren bestimmter Personenkreise, Gebrauch machen.)

Literaturverzeichnis

- Feller, W., **An Introduction to Probability Theory and its Applications, Vol. I, II.** Wiley, New York, 1968.
- Krengel, U., **Einführung in die Wahrscheinlichkeitstheorie und Statistik.** Vieweg, Braunschweig, 1988.
- Morgenstern, D., **Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik.** Springer, Berlin, 1968.
- Plachky, D., L. Baringhaus und N. Schmitz. **Stochastik I.** Akademische Verlagsgesellschaft, Wiesbaden, 1978.

Index

| | | | |
|------------------------------------|------|--|----|
| A | | bedingter ~ von Y unter X | 25 |
| Ablehnungsbereich | 58 | erzeugende Funktion | 31 |
| Alpha-Quantil (α -Quantil) | 41 | Erzeugendensystem zu \mathcal{A} | 35 |
| Alternative | 58 | Exponentialverteilung | 44 |
| antiton | 7 | F | |
| B | | Faltung von P und Q | 30 |
| bedingte Wahrscheinlichkeit | 8 | Faltungshalbgruppe | 30 |
| Bereichsschätzer | 62 | fixpunktfreie Permutationen | 16 |
| Bernstein-Polynom | 34 | Formel | |
| Bias | 57 | ~ von Bayes (Satz 2.2) | 8 |
| Boolesche Ungleichung (Satz 1.6) | 6 | ~ von Poincaré (Satz 1.6) | 6 |
| Borel-Mengen von \mathbb{R} | 35 | Einschluss-Ausschluss-~ (Satz 1.6) | 6 |
| C | | Siebf~ (Satz 1.6) | 6 |
| Cauchy-Schwarz-Ungleichung | 27 | G | |
| Chebyshevsche Ungleichung, die | 33 | Gamma-Verteilung | 43 |
| D | | Gaußsche Glockenkurve | 44 |
| Dirac-Maß | 6 | gemeinsame Dichte | 47 |
| durchschnittsstabil | 37 | gemeinsame Verteilung von X und Y | 24 |
| Dynkin-System | 37 | gemeinsame Verteilungsfunktion | 47 |
| E | | Gesetz der seltenen Ereignisse | 21 |
| einfach (H) | 58 | Gesetz von der totalen Wahrscheinlichkeit (Satz 2.2), das | 8 |
| Einpunktmaß | 6 | Gleichheit von Bienaymé | 29 |
| Elementarereignis | 4 | Gleichverteilung | 43 |
| endliche Additivität (Satz 1.6) | 6 | Gleichverteilung auf dem Einheitsintervall | 36 |
| Ereignis | 4 | Gütefunktion | 58 |
| Ereignissystem | 5 | I | |
| Ergebnisraum | 4, 5 | Indikatorfunktion | 24 |
| erwartungstreu | 57 | isoton | 7 |
| Erwartungswert | 22 | K | |
| ~ existiert nicht | 22 | Kolmogorov-Axiome (Definition), die | 5 |

| | | | |
|--|-------|---|--------|
| Kombinationen | | Neyman-Pearson-Test | 58 |
| k -- von M_n mit Wiederholung | 14 | Normalapproximation für Binomialverteilungen | |
| k -- von M_n ohne Wiederholung | 14 | ~, lokale Form | 52 |
| Konfidenzgebiet | | Normalapproximation für Gamma-Verteilungen | |
| $100(1 - \alpha)$ -prozentiges ~ für θ | 62 | ~, kumulative Form | 54 |
| Konfidenzintervall | 62 | ~, lokale Form | 53 |
| Konfidenzschranke | | Normalapproximation für Poisson-Verteilungen | |
| obere ~ zum Niveau $1 - \alpha$ | 62 | ~, kumulative Form | 51 |
| Korrelationskoeffizienten von X und Y | 28 | ~, lokale Form | 50 |
| Kovarianz von X und Y | 28 | Normalverteilung | 44 |
| kritischer Wert (c) | 58 | | |
| | | O | |
| L | | Operationscharakteristik | 58 |
| Lage der Verteilung, die | 23 | | |
| Laplace-Experiment | 5, 12 | P | |
| Lebesgue-Maß | 36 | parametrische Familie | 55 |
| Likelihood-Funktion | 56 | Partialsomme | 52 |
| Log-- | 56 | Permutationen | |
| Likelihood-Quotient | 59 | k -- von M_n mit Wiederholung | 13 |
| Likelihood-Quotienten-Test | 60 | k -- von M_n ohne Wiederholung | 13 |
| Limes Superior | 4 | Pivot | 63 |
| Linearität (Satz 4.8) | 23 | | |
| | | Q | |
| M | | Quantilfunktion | 41 |
| Markovsche Ungleichung , die | 33 | | |
| Maß | 36 | R | |
| Massenfunktion | 19 | Rechteckverteilung | 36, 43 |
| Maximum-Likelihood-Methode | 55 | relative Häufigkeit | 5 |
| Maximum-Likelihood-Schätzer | 56 | Riemann-Dichte | 42 |
| Menge M_n ($\{1, \dots, n\}$), die | 13 | | |
| messbar | | S | |
| $(\mathcal{A}, \mathcal{A}')$ -- | 38 | Schätzer | 55 |
| messbarer Raum | 36 | Schätzfunktion | 55 |
| Moivre-Laplace | 52 | schwach (konvertiert ~) | 49 |
| Moment , das k -te | 23 | Sigma-Additivität (σ -Additivität) | 5 |
| Monotonie (Satz 4.8) | 23 | Sigma-Algebra (σ -Algebra) | 5 |
| Monotonie (Satz 1.6) | 6 | die von X erzeugte σ -Algebra | 45 |
| Multiplikationsregel (Satz 2.2), die | 8 | die von \mathcal{E} erzeugte ~ | 35 |
| Multiplikationsregel für Erwartungswerte | 27 | Simpson's paradox | 10 |
| | | Standardabweichung | 23 |
| N | | Standardnormalverteilung | 44 |
| Neyman-Pearson-Lemma | 58 | Standardparameter | 44 |

