

Kapitel 1.

Grundlagen der Wahrscheinlichkeitstheorie und Statistik

Statistik ist die Wissenschaft, die Regeln und Verfahren für die Erhebung, Beschreibung, Analyse und Interpretation von numerischen Daten entwickelt.

Der Schwerpunkt dieses Buches liegt auf der Entwicklung und Darstellung von statistischen Analyseverfahren. Dazu werden *stochastische Modelle* vorgestellt, die von unbekannten Parametern abhängen. Um diese Parameter mit Hilfe von erhobenen Daten bestimmen zu können, werden Verfahren zur *Schätzung* von Parametern konstruiert und verglichen. Unter gewissen Annahmen über die zugrundeliegenden stochastischen Modelle werden hieran anschließend Verfahren zum *Testen von Hypothesen* entwickelt.

Die in den späteren Kapiteln behandelten Schätz- und Testverfahren benötigen einen wahrscheinlichkeitstheoretischen Rahmen. Dieses Kapitel gibt eine kurze Einführung in die dafür notwendigen Hilfsmittel aus der Wahrscheinlichkeitstheorie. Hierbei werden viele verschiedene Verteilungen vorgestellt und in den Beispielen vertieft, was für die erfolgreiche Anpassung an verschiedene Datensätze wichtig ist. Für eine ausgiebige Darstellung sei auf Georgii (2004), Resnick (2003) und Chung (2001) verwiesen.

1.1 Grundbegriffe der Wahrscheinlichkeitstheorie

Dieser Abschnitt beschreibt kurz den Kolmogorovschen Zugang zur Wahrscheinlichkeitstheorie. Jedem zufälligen Ereignis wird hierbei eine Wahrscheinlichkeit zugeordnet. Ein Ereignis ist beschrieben durch eine Menge. Das gleichzeitige Eintreten zweier Ereignisse ist der Schnitt zweier Mengen, welches wieder ein Ereignis sein sollte. Dies erfordert eine Axiomatik, welche im Folgenden vorgestellt wird. Grundlage bildet ein *Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, \mathbb{P})$, wobei Ω den *Grundraum*, \mathcal{A} die zugehörige σ -Algebra und \mathbb{P} ein *Wahrscheinlichkeitsmaß* bezeichnet. Die Elemente von \mathcal{A} beschreiben die Ereignisse, welche in einem Zufallsexperiment auftreten können. Mit zwei

Ereignissen A und B aus \mathcal{A} möchte man auch das Ereignis "A und B" betrachten können, weswegen man von \mathcal{A} gewisse Eigenschaften fordert. Eine Menge \mathcal{A} , dessen Elemente Teilmengen von Ω sind, heißt σ -Algebra, falls:

- (i) $\Omega \in \mathcal{A}$.
- (ii) Für jedes $A \in \mathcal{A}$ gilt $\bar{A} := \Omega \setminus A \in \mathcal{A}$.
- (iii) Für Elemente A_1, A_2, \dots von \mathcal{A} gilt $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Weiterhin wird verlangt, dass das Wahrscheinlichkeitsmaß \mathbb{P} die klassischen Kolmogorovschen Axiome erfüllt. Demnach ist die Abbildung $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ ein *Wahrscheinlichkeitsmaß*, falls die folgenden drei Eigenschaften erfüllt sind:

- (i) $\mathbb{P}(\Omega) = 1$.
- (ii) $0 \leq \mathbb{P}(A) \leq 1$ für alle $A \in \mathcal{A}$.
- (iii) Für Elemente A_1, A_2, \dots von \mathcal{A} mit $A_i \cap A_j = \emptyset$ für jedes $i \neq j$ gilt:

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

Hat der Grundraum Ω die Form $\Omega = \{\omega_1, \omega_2, \dots\}$, so nennen wir den zugehörigen Wahrscheinlichkeitsraum *diskret*. In diesem Fall zerfällt der Grundraum in höchstens abzählbar viele disjunkte Ereignisse, und jedes Ereignis $\{\omega_i\}$ heißt *Elementarereignis*.

Bedingte Wahrscheinlichkeiten und Unabhängigkeit. Beobachtet man ein Ereignis, so hat dies möglicherweise einen Einfluß auf die Einschätzung von anderen Ereignissen. Dies wird durch die Verwendung von bedingten Wahrscheinlichkeiten formalisiert.

Seien $A, B \in \mathcal{A}$ zwei Ereignisse mit $\mathbb{P}(B) > 0$. Die *bedingte Wahrscheinlichkeit* von A gegeben B ist definiert durch

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

Darüber hinaus definiert $\mathbb{P}(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$ das bedingte Wahrscheinlichkeitsmaß gegeben B . Dieses Maß ist in der Tat ein Wahrscheinlichkeitsmaß, was in Aufgabe 1.18 bewiesen werden soll.

Ist $\Omega = \bigcup_{i=1}^n B_i$ und sind die B_i paarweise disjunkt, so schreiben wir $\Omega = \sum_{i=1}^n B_i$. In manchen Situationen sind die bedingten Wahrscheinlichkeiten $\mathbb{P}(A|B_i)$ bekannt und man möchte $\mathbb{P}(B_i|A)$ bestimmen. Als Beispiel betrachten wir einen medizinischen Diagnosetest. Die Wahrscheinlichkeiten, dass ein getesteter Patient ein positives (bzw. negatives) Testergebnis erhält,

wenn er tatsächlich die Krankheit hat, seien bekannt. Als Patient mit positivem Testergebnis ist man an der Wahrscheinlichkeit, ob die Krankheit wirklich vorliegt, interessiert. Diese kann man mit dem Satz von Bayes bestimmen.

Satz 1.1 (Satz von Bayes). Sei $\Omega = \sum_{i=1}^n B_i$ mit $\mathbb{P}(B_i) > 0$ für $i = 1, \dots, n$. Dann gilt für $A \in \mathcal{A}$ mit $\mathbb{P}(A) > 0$, dass

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.$$

Diese Formel wird oft als *Bayes-Formel* bezeichnet. Die Erweiterung auf Zufallsvariablen mit einer Dichte ist Gegenstand von Aufgabe 1.27. Zwei Ereignisse A und B heißen *unabhängig*, falls

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

Dann gilt auch $\mathbb{P}(A|B) = \mathbb{P}(A)$. Für n Ereignisse muss man die (schwächere) paarweise Unabhängigkeit von der folgenden Eigenschaft unterscheiden: Die Ereignisse A_1, \dots, A_n heißen *unabhängig*, falls

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k \mathbb{P}(A_{i_j}) \quad \forall \{i_1, \dots, i_k\} \subset \{1, \dots, n\}.$$

Zufallsvariablen. Ein Zufallsexperiment wird durch eine Zufallsvariable modelliert. Eine (k -dimensionale) Zufallsvariable \mathbf{X} ist intuitiv gesprochen eine Abbildung, welche die Grundereignisse $\omega \in \Omega$ auf Vektoren im \mathbb{R}^k abbildet. Um die Wahrscheinlichkeit etwa für das Ereignis $A := \{\mathbf{X} \leq \mathbf{0}\}$ berechnen zu können, ist $A \in \mathcal{A}$ zu fordern. Das führt zu folgendem Begriff der *Meßbarkeit*: Sei \mathcal{B}^k die Borel- σ -Algebra¹. Eine k -dimensionale *Zufallsvariable* ist eine \mathcal{A} - \mathcal{B}^k meßbare Abbildung $\mathbf{X}: \Omega \rightarrow \mathbb{R}^k$, d.h. für jedes $B \in \mathcal{B}^k$ ist

$$\mathbf{X}^{-1}(B) := \{\omega \in \Omega : \mathbf{X}(\omega) \in B\} \in \mathcal{A}.$$

Wir setzen in diesem Buch die Meßbarkeit der verwendeten Funktionen stets voraus und geben nur an wenigen Stellen Hinweise auf die zugrundeliegenden maßtheoretischen Fragen.

Eine Zufallsvariable \mathbf{X} heißt *diskret*, falls sie höchstens abzählbar viele Werte $\mathbf{x}_1, \mathbf{x}_2, \dots$ annimmt. Dann heißt die Funktion $p_{\mathbf{X}}: \{\mathbf{x}_1, \mathbf{x}_2, \dots\} \rightarrow$

¹ Die Borel- σ -Algebra ist die kleinste σ -Algebra, die alle offenen Rechtecke, in diesem Fall $(a_1, b_1) \times \dots \times (a_k, b_k)$, enthält.

$[0, 1]$ gegeben durch

$$p_{\mathbf{X}}(\mathbf{x}_i) = \mathbb{P}(\mathbf{X} = \mathbf{x}_i), \quad i = 1, 2, \dots$$

die *Wahrscheinlichkeitsfunktion* von \mathbf{X} . Durch sie ist \mathbf{X} vollständig beschrieben, denn für jede Wertemenge $B \subset \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ ist $\mathbb{P}(\mathbf{X} \in B) = \sum_{\mathbf{x}_i \in B} p_{\mathbf{X}}(\mathbf{x}_i)$. Um im Folgenden eine einheitliche Schreibweise mit stetigen Zufallsvariablen nutzen zu können, setzen wir stets $p_{\mathbf{X}}(\mathbf{x}) := 0$ für $\mathbf{x} \notin \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$.

Ist eine Zufallsvariable nicht diskret, so kann man sie oft durch ihre Dichte beschreiben. Eine *Dichte* ist eine nichtnegative Funktion p auf \mathbb{R}^k , die Lebesgue-integrierbar ist mit

$$\int_{\mathbb{R}^k} p(\mathbf{x}) d\mathbf{x} = 1.$$

Gilt für eine Zufallsvariable \mathbf{X} , dass für alle $B \in \mathcal{B}^k$

$$\mathbb{P}(\mathbf{X} \in B) = \int_B p(\mathbf{x}) d\mathbf{x}$$

und ist p eine Dichte, so heißt p die *Dichte von \mathbf{X}* . In diesem Fall heißt \mathbf{X} *stetige Zufallsvariable*.

Unabhängig davon, ob eine Zufallsvariable diskret ist oder etwa eine Dichte besitzt, lässt sie sich stets durch ihre Verteilungsfunktion beschreiben. Die *Verteilungsfunktion* einer Zufallsvariable \mathbf{X} ist definiert durch

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \dots, x_k) := \mathbb{P}(X_1 \leq x_1, \dots, X_k \leq x_k).$$

Die Verteilungsfunktion hat, wie man leicht sieht, folgende Eigenschaften. Zur Einfachheit betrachten wir nur den eindimensionalen Fall. Dann gilt: $0 \leq F \leq 1$, F ist monoton wachsend, rechtsseitig stetig, $\lim_{x \rightarrow \infty} F(x) = 1$ und $\lim_{x \rightarrow -\infty} F(x) = 0$. Neben der Verteilungsfunktion spricht man allgemeiner von der Verteilung einer Zufallsvariable. Die *Verteilung* einer Zufallsvariable \mathbf{X} ist ein Wahrscheinlichkeitsmaß $\mathbb{P}_{\mathbf{X}}$, gegeben durch

$$\mathbb{P}_{\mathbf{X}}(B) := \mathbb{P}(\{\omega \in \Omega : \mathbf{X}(\omega) \in B\}) = \mathbb{P}(\mathbf{X} \in B), \quad B \in \mathcal{B}^k.$$

Die Verteilung einer Zufallsvariable ist je nach Typ der Zufallsvariable unterschiedlich darstellbar. Ist \mathbf{X} eine diskrete Zufallsvariable mit Werten $\mathbf{x}_1, \mathbf{x}_2, \dots$ und mit Wahrscheinlichkeitsfunktion p , so ist

$$\mathbb{P}(\mathbf{X} \in B) = \sum_{\mathbf{x}_i \in B} p(\mathbf{x}_i), \quad B \in \mathcal{B}^k.$$

Hat \mathbf{X} hingegen die Dichte p , so ist

$$\mathbb{P}(\mathbf{X} \in B) = \int_B p(\mathbf{x}) d\mathbf{x}, \quad B \in \mathcal{B}^k.$$

Transformationssatz. Eine Transformation einer k -dimensionalen Zufallsvariable \mathbf{X} ist eine meßbare Abbildung $\mathbf{h}: \mathbb{R}^k \rightarrow \mathbb{R}^m$, d.h. $\mathbf{h}^{-1}(B) \in \mathcal{B}^k$ für alle Mengen B aus der Borel- σ -Algebra \mathcal{B}^m . Die Verteilung der transformierten Zufallsvariable $\mathbf{h}(\mathbf{X})$ ist bestimmt durch

$$\mathbb{P}(\mathbf{h}(\mathbf{X}) \in B) = \mathbb{P}(\mathbf{X} \in \mathbf{h}^{-1}(B))$$

für alle $B \in \mathcal{B}^m$. Als Anwendung betrachten wir folgendes Beispiel.

B 1.1 Mittelwert und Stichprobenvarianz: Betrachtet man eine Stichprobe gegeben durch k reellwertige Zufallsvariablen $\mathbf{X} = (X_1, \dots, X_k)^\top$ mit $k \geq 2$, so ist der Vektor gegeben durch den arithmetischen Mittelwert und die Stichprobenvarianz eine Transformation von \mathbf{X} : In diesem Fall ist $\mathbf{h}(\mathbf{X}) = (h_1(\mathbf{X}), h_2(\mathbf{X}))$; der *arithmetische Mittelwert* ist $h_1(\mathbf{X})$ und die *Stichprobenvarianz* ist $h_2(\mathbf{X})$ mit

$$h_1(\mathbf{X}) := \frac{1}{k} \sum_{i=1}^k X_i =: \bar{X},$$

$$h_2(\mathbf{X}) := \frac{1}{k-1} \sum_{i=1}^k (X_i - \bar{X})^2 =: s^2(\mathbf{X}).$$

Die besondere Normierung mit $(k-1)$ sorgt dafür, dass die Stichprobenvarianz erwartungstreu ist, eine Eigenschaft welche man verliert, wenn man stattdessen mit k normiert. Dies werden wir in Aufgabe 1.3 diskutieren.

Für stetige, reellwertige Zufallsvariablen hat man folgenden wichtigen Satz:

Satz 1.2 (Transformationssatz). *Sei X eine reellwertige, stetige Zufallsvariable mit Dichte p_X . Die Transformation $h: \mathbb{R} \rightarrow \mathbb{R}$ sei bijektiv auf einer offenen Menge B mit $\mathbb{P}(X \in B) = 1$. Ferner sei h differenzierbar und $h'(x) \neq 0 \forall x \in B$. Dann ist $Y := h(X)$ eine stetige Zufallsvariable und die Dichte von Y ist gegeben durch*

$$p_{h(X)}(y) = \frac{p_X(h^{-1}(y))}{|h'(h^{-1}(y))|} \mathbf{1}_{\{h^{-1}(y) \in B\}}, \quad y \in \mathbb{R}.$$

Diese Behauptung lässt sich leicht durch Differenzieren der Verteilungsfunktion von Y und Anwenden der Kettenregel zeigen.

Im mehrdimensionalen Fall gilt ein analoges Resultat: Sei $\mathbf{h}: \mathbb{R}^k \rightarrow \mathbb{R}^k$, $\mathbf{h} = (h_1, \dots, h_k)$, $h_i: \mathbb{R}^k \rightarrow \mathbb{R}$ und die Jacobi-Determinante gegeben durch

$$J_{\mathbf{h}}(\mathbf{x}) := \begin{vmatrix} \frac{\partial}{\partial x_1} h_1(\mathbf{x}) & \dots & \frac{\partial}{\partial x_1} h_k(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_k} h_1(\mathbf{x}) & \dots & \frac{\partial}{\partial x_k} h_k(\mathbf{x}) \end{vmatrix}.$$

Satz 1.3 (Transformationssatz für Zufallsvektoren). *Sei $\mathbf{h}: \mathbb{R}^k \rightarrow \mathbb{R}^k$ und $B \subset \mathbb{R}^k$ eine offene Menge, so dass gilt:*

- (i) \mathbf{h} hat stetige erste partielle Ableitungen auf B ,
- (ii) \mathbf{h} ist bijektiv auf B ,
- (iii) $J_{\mathbf{h}}(\mathbf{x}) \neq 0$, $\forall \mathbf{x} \in B$

und sei \mathbf{X} eine stetige Zufallsvariable mit $\mathbb{P}(\mathbf{X} \in B) = 1$. Dann ist die Dichte von $\mathbf{Y} := \mathbf{h}(\mathbf{X})$ gegeben durch

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{X}}(\mathbf{h}^{-1}(\mathbf{y})) \cdot |J_{\mathbf{h}^{-1}}(\mathbf{y})| \mathbb{1}_{\{\mathbf{h}^{-1}(\mathbf{y}) \in B\}}, \quad \mathbf{y} \in \mathbb{R}^k.$$

Unabhängigkeit. Die Unabhängigkeit von Zufallsvariablen geht maßgeblich auf die Unabhängigkeit von Ereignissen zurück. Zwei Zufallsvariablen $\mathbf{X}_1 \in \mathbb{R}^k$ und $\mathbf{X}_2 \in \mathbb{R}^m$ heißen *unabhängig*, falls die Ereignisse $\{\mathbf{X}_1 \in A\}$ und $\{\mathbf{X}_2 \in B\}$ unabhängig für *alle* $A \in \mathcal{B}^k$ und $B \in \mathcal{B}^m$ sind.

Unabhängigkeit kann man dadurch charakterisieren, dass die Dichte, die Wahrscheinlichkeitsfunktion oder die Verteilungsfunktion in Produktgestalt zerfällt:

Satz 1.4. *Ist die Zufallsvariable $\mathbf{X} = (X_1, \dots, X_k)^\top$ stetig mit Dichte $p_{\mathbf{X}}$ oder diskret mit Wahrscheinlichkeitsfunktion $p_{\mathbf{X}}$, so sind die folgenden drei Aussagen äquivalent:*

- (i) X_1, \dots, X_k sind unabhängig.
- (ii) $F_{\mathbf{X}}(x_1, \dots, x_k) = F_{X_1}(x_1) \cdots F_{X_k}(x_k)$ für alle $x_1, \dots, x_k \in \mathbb{R}$.
- (iii) $p_{\mathbf{X}}(x_1, \dots, x_k) = p_{X_1}(x_1) \cdots p_{X_k}(x_k)$ für alle $x_1, \dots, x_k \in \mathbb{R}$.

Wir bezeichnen Zufallsvariablen X_1, \dots, X_k oder auch etwa eine ganze Folge X_1, X_2, \dots als unabhängig, falls für jede beliebige Kombination (i_1, \dots, i_{k_1}) und (j_1, \dots, j_{k_2}) , welche sich nicht überschneiden, die Vektoren $(X_{i_1}, \dots, X_{i_{k_1}})^\top$ und $(X_{j_1}, \dots, X_{j_{k_2}})^\top$ unabhängig sind. Im Allgemeinen ist dies stärker als die Annahme der paarweisen Unabhängigkeit, unter welcher jedes X_i und X_j mit $i \neq j$ unabhängig sind.

Zufallsvariablen, welche unabhängig und identisch verteilt sind, bezeichnen wir kurz als *i.i.d.* (independent and identically distributed). Dies ist eine in der Statistik häufig gemachte Annahme.

Momente. Wichtige Charakteristika von Zufallsvariablen können oftmals durch einfachere Funktionale als die Verteilungsfunktion beschrieben werden. Die Normalverteilung beispielsweise ist vollständig durch ihr erstes und zweites Moment beschrieben. Dieser Abschnitt führt zentrale Größen wie Erwartungswert und Varianz und darüber hinausgehend die Momente einer Zufallsvariable ein. Für $\mathbf{x} \in \mathbb{R}^k$ erhält man durch $|\mathbf{x}| := |x_1| + \dots + |x_d|$ eine Norm auf dem Vektorraum \mathbb{R}^k .

Der *Erwartungswert* einer Zufallsvariable \mathbf{X} ist wie folgt definiert: Ist \mathbf{X} diskret mit Werten $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$, so ist der Erwartungswert definiert durch

$$\mathbb{E}(\mathbf{X}) := \sum_{i=1}^{\infty} \mathbf{x}_i \mathbb{P}(\mathbf{X} = \mathbf{x}_i),$$

falls die Summe absolut konvergiert, wofür wir $\mathbb{E}(|\mathbf{X}|) < \infty$ schreiben. Ist \mathbf{X} eine stetige Zufallsvariable mit Dichte $p_{\mathbf{X}}$, so ist

$$\mathbb{E}(\mathbf{X}) := \int_{\mathbb{R}^k} \mathbf{x} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x},$$

falls $\int_{\mathbb{R}^k} |\mathbf{x}| p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \infty$. Gilt $\mathbb{E}(|\mathbf{X}|) < \infty$, so nennen wir \mathbf{X} *integrierbar*. Der Erwartungswert einer Zufallsvariable gibt den Wert an, welchen die Zufallsvariable im Mittel annimmt. Man verifiziert leicht, dass der Erwartungswert ein linearer Operator ist, d.h. für $a_1, \dots, a_n \in \mathbb{R}$ ist

$$\mathbb{E}\left(\sum_{i=1}^n a_i \mathbf{X}_i\right) = \sum_{i=1}^n a_i \mathbb{E}(\mathbf{X}_i).$$

Darüber hinaus ist der Erwartungswert monoton, d.h. aus $\mathbb{P}(\mathbf{X} \geq \mathbf{Y}) = 1$ folgt Hierbei ist für zwei Vektoren der komponentenweise Vergleich gemeint: $\mathbf{a} \geq \mathbf{b} \Leftrightarrow a_i \geq b_i$ für alle $1 \leq i \leq d$.

$$\mathbb{E}(\mathbf{X}) \geq \mathbb{E}(\mathbf{Y}). \quad (1.1)$$

Folgende Ungleichung wird sich als nützlicher Begleiter erweisen. Eine Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ heißt *konvex*, falls $g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y)$ für alle $\lambda \in (0, 1)$ und alle $x, y \in \mathbb{R}$.

Satz 1.5 (Jensensche Ungleichung). *Sei $g: \mathbb{R} \rightarrow \mathbb{R}$ konvex und X eine reellwertige Zufallsvariable mit $\mathbb{E}(|X|) < \infty$. Dann gilt*

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X)). \quad (1.2)$$

Gleichheit in (1.2) gilt genau dann, wenn für jede Gerade $a + bx$ tangential zu g an $x = \mathbb{E}(X)$ gilt, dass $\mathbb{P}(g(X) = a + bX) = 1$.

Ein typisches Beispiel ist $g(x) = x^2$: Für eine Zufallsvariable X mit verschwindenden Erwartungswert folgt bereits aus $x^2 \geq 0$, dass $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2 = 0$.

Das k -te *Moment* von X ist $\mathbb{E}(X^k)$ und das k -te *zentrierte (zentrale) Moment* von X ist definiert durch

$$\mu_k := \mathbb{E} \left((X - \mathbb{E}(X))^k \right).$$

Das zweite zentrierte Moment spielt eine besondere Rolle: Die *Varianz* von X ist definiert durch

$$\sigma^2 := \text{Var}(X) = \mathbb{E} \left((X - \mathbb{E}(X))^2 \right) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Die letzte Gleichheit lässt sich durch Ausmultiplizieren und Verwendung der Linearität des Erwartungswertes leicht zeigen. Gilt $\mathbb{E}(X^2) < \infty$, so nennen wir X *quadrat-integrierbar*. Die Varianz ist ein Maß für die Streuung einer Zufallsvariable. Um die Abweichung einer Zufallsvariable von einer Normalverteilung zu messen, nutzt man typischerweise noch ein geeignetes drittes und viertes Moment, die *Schiefte* (skewness): $\gamma_1 = \frac{\mu_3}{\sigma^3}$ und die *Kurtosis*: $\gamma_2 := \frac{\mu_4}{\sigma^4} - 3$. Betrachtet man zwei reellwertige Zufallsvariablen X_1 und X_2 , so kann man deren *lineare* Abhängigkeit durch die Kovarianz erfassen. Dieses Maß zeigt allerdings außerhalb der Normalverteilungsfamilien prekäre Eigenschaften und sollte dort nur mit Vorsicht angewendet werden, siehe Aufgabe 1.2 und Schmidt (2007). Für zwei quadrat-integrierbare Zufallsvariablen X_1 und X_2 definiert man die *Kovarianz* von X_1 und X_2 durch

$$\text{Cov}(X_1, X_2) := \mathbb{E} \left((X_1 - \mathbb{E}(X_1)) \cdot (X_2 - \mathbb{E}(X_2)) \right) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2).$$

Die Kovarianz ist dabei abhängig von den Varianzen der einzelnen Zufallsvariablen. Ein skalenunabhängiges Maß für die lineare Abhängigkeit ist die *Korrelation* zwischen X_1 und X_2 . Sie ist definiert durch

$$\text{Corr}(X_1, X_2) := \frac{\text{Cov}(X_1, X_2)}{(\text{Var}(X_1) \text{Var}(X_2))^{1/2}};$$

es gilt $\text{Corr}(X_1, X_2) \in [-1, 1]$. Zwei Zufallsvariablen X_1, X_2 mit $\text{Cov}(X_1, X_2) = 0$ nennt man *unkorreliert*. Sind die quadrat-integrierbaren Zufallsvariablen X_1 und X_2 unabhängig, so folgt aus $\mathbb{E}(X_1 X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$, dass

$$\text{Cov}(X_1, X_2) = \text{Corr}(X_1, X_2) = 0.$$

Die Umkehrung trifft typischerweise nicht zu, siehe Aufgabe 1.2. Weiterhin gilt die so genannte *Cauchy-Schwarz Ungleichung*

$$(\text{Cov}(X, Y))^2 \leq \text{Var}(X) \cdot \text{Var}(Y). \quad (1.3)$$

Für quadrat-integrierbare Zufallsvariablen X_1, \dots, X_n gilt

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i,j=1, i < j}^n \text{Cov}(X_i, X_j).$$

Sind X_1, \dots, X_n darüber hinaus paarweise unkorreliert (dies folgt aus deren Unabhängigkeit), so gilt die wichtige Regel von *Bienaymé*

$$\text{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \text{Var}(X_i). \quad (1.4)$$

Momentenerzeugende Funktion. Mitunter ist es günstig, zur Beschreibung der Verteilung einer Zufallsvariable ein weiteres Hilfsmittel zur Verfügung zu haben. Ein solches ist die so genannte *momentenerzeugende Funktion* Ψ_X . Ist X eine reellwertige Zufallsvariable, so ist $\Psi_X: \mathbb{R} \rightarrow [0, \infty]$ definiert durch

$$\Psi_X(s) := \mathbb{E}(e^{sX}).$$

Offensichtlich ist $\Psi_X(0) = 1$. Ist Ψ_X endlich in einer Umgebung der Null, so bestimmt $\Psi_X(s)$ eindeutig die Verteilung von X . Darüber hinaus gilt dann auch, dass

$$\left. \frac{d^k}{ds^k} \Psi_X(s) \right|_{s=0} = \mathbb{E}(X^k).$$

Ψ_X wird sich für die Beschreibung der Verteilung von Summen unabhängiger Zufallsvariablen als extrem nützlich erweisen. Denn, sind X_1, \dots, X_n unabhängig, so folgt

$$\Psi_{X_1 + \cdots + X_n}(s) = \prod_{i=1}^n \Psi_{X_i}(s).$$

In Satz 2.12 wird die momentenerzeugende Funktion für exponentielle Familien bestimmt. Weitergehende Informationen über die momentenerzeugende Funktion finden sich etwa in: Gut (2005), Kapitel 4.8 auf Seite 189 – 191.

Anders als die momentenerzeugende Funktion existiert die *charakteristische Funktion* $\varphi_X(s) := \mathbb{E}(\exp(isX))$ stets für alle $s \in \mathbb{R}$. Auch sie charakterisiert die Verteilung eindeutig (siehe Shao (2008), Seite 35) und die Inversion ist ein klassisches Resultat (siehe dazu Gut (2005), Kapitel 4.1, Seite 157 – 165 oder Billingsley (1986), Seite 395).

1.2 Klassische Verteilungen der Statistik

In diesem Abschnitt werden die klassischen Verteilungen kurz vorgestellt. Sie bilden eine wesentliche Grundlage für die späteren Aussagen. Oft ist es in der Statistik notwendig, sich auf eine bestimmte Verteilung oder eine Verteilungsklasse festzulegen, weswegen den angeführten Beispielen eine wichtige Funktion zukommt. Diese bieten jedoch nur einen kleinen Ausschnitt der be-

kannten Verteilungen, wie ein Blick in die Standardwerke: Johnson, Kotz und Balakrishnan (1994a), Johnson, Kotz und Balakrishnan (1994b), Johnson, Kotz und Kemp (1992) zeigt.

Diskrete Verteilungen. Wir betrachten eine diskrete Zufallsvariable X mit Wahrscheinlichkeitsfunktion p .

- *Binomialverteilung:* Wir schreiben $X \sim \text{Bin}(n, p)$, falls $p \in (0, 1)$ und für jedes $k \in \{0, \dots, n\}$

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

Als Spezialfall erhält man die *Bernoulli-Verteilung* $\text{Bin}(1, p)$. Dies ist eine Zufallsvariable, welche nur die Werte 0 oder 1 annimmt. Jede binomialverteilte Zufallsvariable lässt sich als Summe von Bernoulli-Zufallsvariablen schreiben (siehe Beispiel 1.3 und Aufgabe 1.4).

- *Poisson-Verteilung:* Wir schreiben $X \sim \text{Poiss}(\lambda)$, falls $\lambda > 0$ und für $k \in \{0, 1, 2, \dots\}$

$$\mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}. \quad (1.5)$$

- *Multinomialverteilung:* Wir schreiben $\mathbf{X} \sim M(n, p_1, \dots, p_k)$, falls $n \in \mathbb{N}$, $p_1, \dots, p_k \in (0, 1)$ mit $\sum_{i=1}^k p_i = 1$, $\mathbf{X} \in \mathbb{N}^k$ und für beliebige Zahlen $i_1, \dots, i_k \in \{0, \dots, n\}$ mit $\sum_{j=1}^k i_j = n$ gilt, dass

$$\mathbb{P}(\mathbf{X} = (i_1, \dots, i_k)^\top) = \frac{n!}{i_1! \dots i_k!} p_1^{i_1} \dots p_k^{i_k}.$$

Diese Verteilung entsteht durch die Klassifizierung von n Objekten in k Klassen und i_j repräsentiert die Anzahl der Objekte in Klasse j .

Laplacesche Modelle. Betrachtet man einen endlichen Grundraum $\Omega = \{\omega_1, \dots, \omega_n\}$, so erhält man die wichtige Klasse der *Laplaceschen Modelle*, falls $\mathbb{P}(\{\omega_i\}) = n^{-1}$ für alle $1 \leq i \leq n$. Alle Elementarereignisse haben demzufolge die gleiche Wahrscheinlichkeit. Notiert man die Anzahl der Elemente in A durch $|A|$, so ergibt sich für $A \subset \Omega$

$$\mathbb{P}(A) = \sum_{w_i \in A} \mathbb{P}(\{w_i\}) = \sum_{w_i \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|},$$

wonach die Wahrscheinlichkeit eines Ereignisses durch die Formel „Günstige durch Mögliche“ berechnet werden kann. Dies gilt allerdings nur unter der Annahme, dass alle Elementarereignisse die gleiche Wahrscheinlichkeit haben. Das folgende Beispiel werden wir in Kapitel 2 auf der Seite 37 wieder aufgreifen.

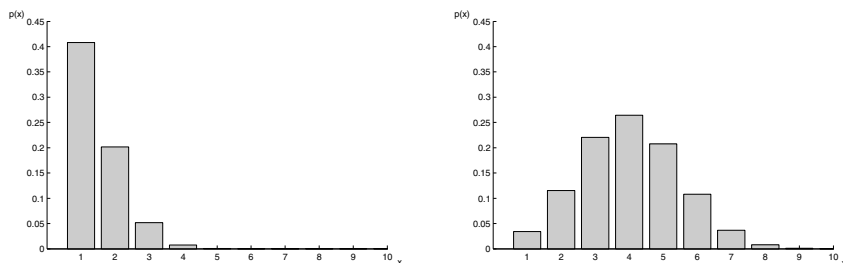


Abb. 1.1 Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung aus Beispiel 1.2 mit $N = 100$, $n = 10$ und $\theta = 0.1$ (links) bzw. $\theta = 0.4$ (rechts).

B 1.2 Hypergeometrische Verteilung: Man betrachtet eine Menge mit N Elementen, wobei jedes Element den Wert 0 oder 1 annehmen kann. Der Anteil der Elemente mit Wert 0 sei $\theta \in (0, 1)$, so dass $N\theta$ Elemente den Wert 0 haben. Es werde eine Teilmenge mit n Elementen ausgewählt und die Zufallsvariable X bezeichne die Anzahl der Elemente in der Teilmenge, welche den Wert 0 haben. Jede Kombination habe die gleiche Wahrscheinlichkeit, es handelt sich folglich um ein Laplacesches Modell. Dann erhält man die *hypergeometrische Verteilung*

$$\mathbb{P}(X = k) = \frac{\binom{N\theta}{k} \binom{N-N\theta}{n-k}}{\binom{N}{n}}, \quad 0 \leq k \leq n$$

oder kurz $X \sim \text{Hypergeo}(N, n, \theta)$ durch Abzählen der möglichen Kombinationen: Insgesamt gibt es $\binom{N}{n}$ Möglichkeiten aus N Teilen eine Stichprobe des Umfangs n zu ziehen. Sollen davon $k \in \{0, \dots, n\}$ Teile den Wert 0 haben, so gibt es zum einen $\binom{N\theta}{k}$ Möglichkeiten, k Teile mit dem Wert 0 aus den $N\theta$ Teilen mit dem Wert 0 zu ziehen. Zum anderen gibt es $\binom{N-N\theta}{n-k}$ Möglichkeiten $n-k$ Teile mit dem Wert 1 aus insgesamt $N - N\theta$ Teilen mit dem Wert 1 auszuwählen. Die zugehörige Wahrscheinlichkeitsfunktion ist in Abbildung 1.1 dargestellt.

Stetige Verteilungen. Wenn die beobachteten Daten keiner diskreten Wertemenge unterliegen, arbeitet man mit stetigen Verteilungen. Zu Beginn seien einige wichtige Beispiele von reellwertigen Zufallsvariablen mit Dichte p vorgestellt.

- *Exponentialverteilung:* Wir schreiben $X \sim \text{Exp}(\lambda)$, falls $\lambda > 0$ und

$$p(x) = \mathbb{1}_{\{x>0\}} \lambda e^{-\lambda x}.$$

- *Gleichverteilung:* Wir schreiben $X \sim U(a, b)$, falls $a < b$ und

$$p(x) = \mathbb{1}_{\{x \in [a, b]\}} \frac{1}{b - a}.$$

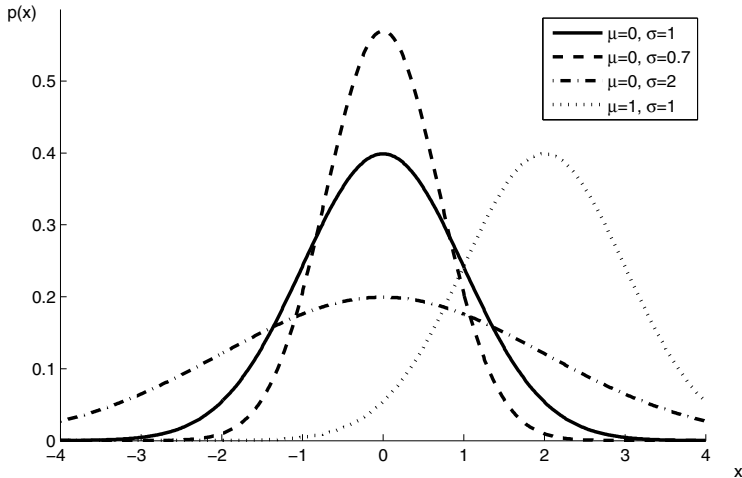


Abb. 1.2 Dichte der Normalverteilung für verschiedene Parameterkonstellationen.

- *Normalverteilung:* Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$, falls $\mu \in \mathbb{R}$, $\sigma > 0$ und

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (1.6)$$

Dann gilt, dass $\mathbb{E}(X) = \mu$ und $\text{Var}(X) = \sigma^2$. Die Dichte ist in Abbildung 1.2 dargestellt. Ist $\mu = 0$ und $\sigma = 1$, so spricht man von einer *Standardnormalverteilung*.

Oft verwendet man die Bezeichnung

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

für die Dichte der Standardnormalverteilung und

$$\Phi(x) := \int_{-\infty}^x \phi(y) dy$$

für die Verteilungsfunktion der Standardnormalverteilung. Die Normalverteilung ist mit Abstand die wichtigste Verteilung in der Statistik, da sie durch den zentralen Grenzwertsatz (Satz 1.31) zur Approximation der Verteilung von einer hinreichend großen Zahl unabhängiger und identisch verteilter Zufallsvariablen mit existierendem zweiten Moment benutzt werden kann. Die Normalverteilung ist stabil unter Summenbildung und Skalierung (siehe Aufgabe 1.31).

Die Exponentialverteilung ist ein Spezialfall der Gamma-Verteilung währenddessen die Gleichverteilung ein Spezialfall der Beta-Verteilung ist, welche ab Seite 16 eingeführt werden.

Rund um die Normalverteilung und die Schätzung von μ und σ^2 gibt es eine Familie von unerlässlichen Verteilungen, welche nun kurz vorgestellt werden.

Die χ^2 , F und t-Verteilung. Die χ^2 -Verteilung entsteht als Summe von quadrierten, normalverteilten Zufallsvariablen.

Lemma 1.6. (und Definition) Sind X_1, \dots, X_n unabhängig und standard-normalverteilt, heißt

$$V := \sum_{i=1}^n X_i^2$$

χ^2 -verteilt mit n Freiheitsgraden, kurz χ_n^2 -verteilt. Die Dichte von V ist gegeben durch

$$p_{\chi_n^2}(x) = \mathbb{1}_{\{x>0\}} \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}. \quad (1.7)$$

Hierbei verwenden wir die *Gamma-Funktion*, definiert durch

$$\Gamma(a) := \int_0^\infty t^{a-1} e^{-t} dt, \quad a > 0.$$

Dann ist $\Gamma(n) = (n-1)!$, $n \in \mathbb{N}$ und $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Weiterhin gilt $\mathbb{E}(V) = n$ und $\text{Var}(V) = 2n$. Die Herleitung der Dichte ist Gegenstand von Aufgabe 1.32.

Bemerkung 1.7. Die Darstellung der Dichte in (1.7) zeigt, dass die χ_n^2 -verteilte Zufallsvariable V für $n = 2$ exponentialverteilt ist mit Parameter $\frac{1}{2}$. Aus dem zentralen Grenzwertsatz (Satz 1.31) folgt, dass

$$\frac{\chi_n^2 - n}{\sqrt{2n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Möchte man ein Konfidenzintervall für den Mittelwert einer Normalverteilung mit unbekannter Varianz bilden, so muss man diese schätzen. Dabei taucht die Wurzel einer Summe von Normalverteilungsquadraten (mit Faktor $\frac{1}{n}$) im Nenner auf. Hierüber gelangt man zur t -Verteilung, welche oft auch als Student-Verteilung oder Studentsche t -Verteilung bezeichnet wird.

Definition 1.8. Ist X standardnormalverteilt und V χ_n^2 -verteilt und unabhängig von X , so heißt die Verteilung von

$$T := \frac{X}{\sqrt{\frac{1}{n}V}} \quad (1.8)$$

die t -Verteilung mit n Freiheitsgraden, kurz t_n -Verteilung.

Lemma 1.9. Die Dichte der t_n -Verteilung ist gegeben durch

$$p_{t_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(n/2)\Gamma(1/2)\sqrt{n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

für alle $x \in \mathbb{R}$.

Für Vergleiche von Varianzen werden wir Quotienten der Schätzer betrachten und gelangen so zur F -Verteilung.

Definition 1.10. Sind V und W unabhängig und χ_n^2 bzw. χ_m^2 -verteilt, so heißt die Verteilung von

$$F := \frac{V/n}{W/m}$$

die F -Verteilung mit (n, m) Freiheitsgraden, kurz $F_{n,m}$ -Verteilung.

Für die Dichte sei an die Formel für die *Beta-Funktion* $B(a, b)$ erinnert: Für $a, b > 0$ ist

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt. \quad (1.9)$$

Dann ist $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Damit erhalten wir folgende Darstellung.

Lemma 1.11. Die Dichte der $F_{n,m}$ -Verteilung ist

$$p_{F_{n,m}}(x) = \mathbb{1}_{\{x>0\}} \frac{n^{n/2} m^{m/2}}{B(n/2, m/2)} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{(n+m)/2}}.$$

Beweis. Für die Verteilungsfunktion an der Stelle $t > 0$ erhalten wir aufgrund der Unabhängigkeit von V und W

$$\begin{aligned}
\mathbb{P}\left(\frac{V/n}{W/m} \leq t\right) &= \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \mathbb{1}_{\{\frac{x}{y} \frac{m}{n} \leq t\}} p_{\chi_n^2}(x) p_{\chi_m^2}(y) dx dy \\
&= \int_0^\infty p_{\chi_m^2}(y) \left[\int_0^{t y n/m} p_{\chi_n^2}(x) dx \right] dy.
\end{aligned}$$

Da wir die Dichte bestimmen wollen, transformieren wir das zweite Integral mittels $w = mx/(ny)$ und erhalten, dass

$$\begin{aligned}
\mathbb{P}\left(\frac{V/n}{W/m} \leq t\right) &= \int_0^\infty p_{\chi_m^2}(y) \int_0^t p_{\chi_n^2}(w \cdot ny/m) \frac{ny}{m} dw dy \\
&= \int_0^t \left[\int_0^\infty p_{\chi_m^2}(y) p_{\chi_n^2}(w \cdot ny/m) \frac{ny}{m} dy \right] dw.
\end{aligned}$$

Der Ausdruck in der Klammer gibt die Dichte an. Unter Verwendung von (1.7) ergibt sich die Behauptung. \square

Bemerkung 1.12. Eine *Rayleigh*-verteilte Zufallsvariable X ist nicht negativ und hat zu dem Parameter $\sigma > 0$ die Dichte

$$p(x) = \mathbb{1}_{\{x>0\}} \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Die Rayleigh-Verteilung entsteht als Norm einer zweidimensionalen, zentrierten Normalverteilung: Die Zufallsvariablen Y und Z seien unabhängig und jeweils $\mathcal{N}(0, \sigma^2)$ -verteilt. Dann ist $\sqrt{Y^2 + Z^2}$ Rayleigh-verteilt (siehe Aufgabe 1.36). Aufgrund dessen ist X^2 gerade χ_2^2 -verteilt falls $\sigma = 1$.

Nichtzentrale t -, F - und χ^2 -Verteilung. In diesem Abschnitt stellen wir nichtzentrale Verteilungen vor, die im Zusammenhang mit Hypothesentests in linearen Modellen im Abschnitt 7.3 benötigt werden. Im Unterschied zu den zentrierten Verteilungen können hier die zugrundeliegenden normalverteilten Zufallsvariablen einen nicht verschwindenden Erwartungswert haben.

Definition 1.13. Seien $X \sim \mathcal{N}(\theta, 1)$, $V \sim \chi_n^2$ und X und V unabhängig. Dann heißt

$$T := \frac{X}{\sqrt{\frac{1}{n}V}}$$

nichtzentral t -verteilt mit n Freiheitsgraden und Nichtzentralitätsparameter θ , kurz $t_n(\theta)$ -verteilt.

Analog definiert man die nichtzentrale χ^2 -Verteilung:

Definition 1.14. Seien $X_i \sim \mathcal{N}(\mu_i, 1)$, $i = 1, \dots, n$ und unabhängig. Dann heißt

$$V := \sum_{i=1}^k X_i^2$$

nichtzentral χ^2 -verteilt mit Nichtzentralitätsparameter $\theta := \sum_{i=1}^k \mu_i^2$, oder kurz $\chi_k^2(\theta)$ -verteilt.

In Aufgabe 1.33 wird gezeigt, dass die nichtzentrale χ^2 -Verteilung wohldefiniert ist und die Verteilung in der Tat nicht von den einzelnen μ_1, \dots, μ_n , sondern nur von θ abhängt. Weitere Informationen findet man in Johnson, Kotz und Balakrishnan (1994b).

Definition 1.15. Sei $V \sim \chi_k^2(\theta)$ und $W \sim \chi_m^2$ sowie V und W unabhängig. Dann heißt die Zufallsvariable

$$Z := \frac{V/k}{W/m}$$

nichtzentral F -verteilt mit Nichtzentralitätsparameter θ , kurz $F_{k,m}(\theta)$ -verteilt.

Es gibt noch zahlreiche andere Erweiterungen von Verteilungen auf ihre nichtzentralen Analoga (siehe dazu die nichtzentrale Exponentialverteilung im Beispiel 3.12).

Die Beta- und die Gamma-Verteilung. In diesem Abschnitt führen wir die Beta- und Gamma-Verteilungen ein. Diese beiden Verteilungsklassen beschreiben relativ allgemeine Verteilungen, welche einige bereits bekannte Verteilungen als Spezialfälle enthalten. Die Gamma-Verteilung tritt als Verallgemeinerung der Exponentialverteilung auf und beschreibt deswegen stets positive Zufallsvariablen. Die Beta-Verteilung ist eine Verallgemeinerung der Gleichverteilung auf dem Einheitsintervall und beschreibt demzufolge nur Zufallsvariablen mit Werten in $[0, 1]$.

Definition 1.16. Eine Zufallsvariable X heißt *Gamma-verteilt* zu den Parametern $a, \lambda > 0$, falls sie folgende Dichte besitzt:

$$p_{a,\lambda}(x) = \mathbb{1}_{\{x>0\}} \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}. \quad (1.10)$$

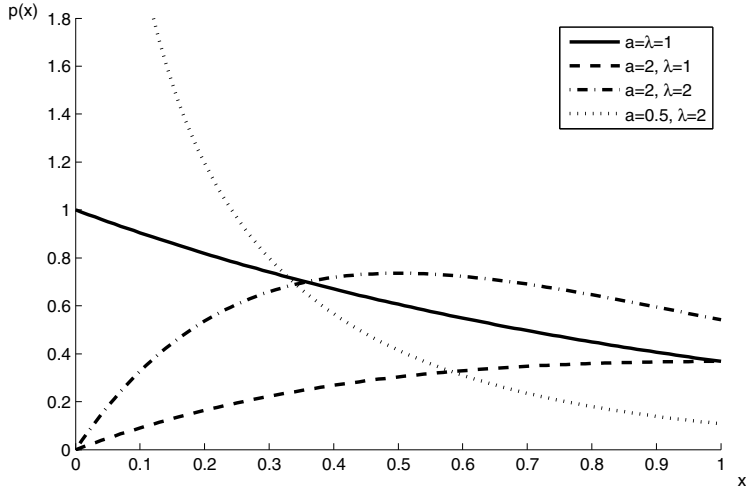


Abb. 1.3 Dichte der $\text{Gamma}(a, \lambda)$ -Verteilung für verschiedene Parameterkonstellationen. Für $a = 1$ erhält man eine Exponentialverteilung.

Ist X Gamma-verteilt, so schreiben wir kurz $X \sim \text{Gamma}(a, \lambda)$. Weiterhin gilt: $cX \sim \text{Gamma}(a, \lambda/c)$ (siehe Aufgabe 1.9 (iii)). Aus diesem Grund nennt man λ^{-1} einen Skalenparameter, während a ein Parameter ist, welcher die Form der Verteilung bestimmt (vgl. Abbildung 1.3). Ist a eine natürliche Zahl, so ist $\Gamma(a) = (a-1)!$. In diesem Fall wird die Verteilung auch eine *Erlang-Verteilung* genannt.

Die momentenerzeugende Funktion einer Gamma-Verteilung wird in Aufgabe 1.12 bestimmt. Daraus erhält man die Momente: Ist $X \sim \text{Gamma}(a, \lambda)$, so gilt

$$\mathbb{E}(X) = \frac{a}{\lambda}, \quad \text{Var}(X) = \frac{a}{\lambda^2}.$$

Die Summe von unabhängigen $\text{Gamma}(\cdot, \lambda)$ -verteilten Variablen ist wieder Gamma-verteilt: Seien X_1, \dots, X_n unabhängig mit $X_i \sim \text{Gamma}(a_i, \lambda)$, so ist

$$\sum_{i=1}^n X_i \sim \text{Gamma}\left(\sum_{i=1}^n a_i, \lambda\right). \quad (1.11)$$

Der Beweis kann über die momentenerzeugende Funktion erfolgen (siehe Aufgabe 1.9). Weiterhin ist eine χ_n^2 -verteilte Zufallsvariable $\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$ -verteilt. Als weiteren Spezialfall erhält man die Exponentialverteilung zum Parameter λ für $a = 1$.

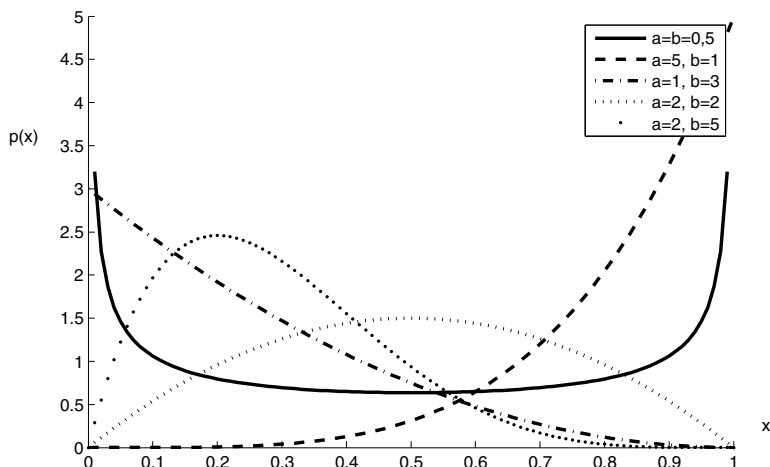


Abb. 1.4 Dichte der Beta-Verteilung für verschiedene Parameterkonstellationen.

Definition 1.17. Eine Zufallsvariable heißt *Beta*-verteilt zu den Parametern $a, b > 0$, falls sie die Dichte

$$p_{a,b}(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{\{x \in [0,1]\}}$$

hat.

Hierbei ist $B(a,b)$ die Beta-Funktion (siehe Gleichung 1.9). Für $a = b = 1$ erhält man die Gleichverteilung auf $[0, 1]$ als Spezialfall. Der Erwartungswert einer $Beta(a,b)$ -Verteilung ist $a/(a+b)$ und die Varianz beträgt

$$\frac{ab}{(1+a+b)(a+b)^2}.$$

Bemerkung 1.18. Sind X, Y unabhängig und $Gamma(a,b)$ bzw. $Gamma(a,c)$ -verteilt, so ist $X/(X+Y)$ gerade $Beta(b,c)$ -verteilt (siehe Aufgabe 1.9).

1.2.1 Die Multivariate Normalverteilung

Dieser Abschnitt widmet sich der mehrdimensionalen Normalverteilung. Für weitergehende Ausführungen sei auf Georgii (2004), Abschnitt 9.1 verwiesen.

Definition 1.19. Ein k -dimensionaler Zufallsvektor \mathbf{X} heißt *k -variater normalverteilt*, falls ein $\boldsymbol{\mu} \in \mathbb{R}^k$ und ein $L \in \mathbb{R}^{k \times m}$ existiert mit $\text{Rang}(L) = m$, so dass

$$\mathbf{X} = L\mathbf{Z} + \boldsymbol{\mu},$$

wobei $\mathbf{Z} = (Z_1, \dots, Z_m)^\top$ und Z_i i.i.d. sind mit $Z_1 \sim \mathcal{N}(0, 1)$.

In diesem Fall schreiben wir $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ mit $\Sigma = LL^\top$. Ist $k = m$, so sagt man, dass \mathbf{Y} eine *nicht singuläre* Normalverteilung besitzt, andernfalls ($k > m$) hat \mathbf{X} eine *singuläre* Normalverteilung.

Für eine quadratintegrierbare, k -dimensionalen Zufallsvariable \mathbf{X} wird die Variabilität durch die *Varianz-Kovarianz Matrix* $\text{Var}(\mathbf{X})$ gemessen. Sie ist gegeben durch die Matrix $D := \text{Var}(\mathbf{X}) \in \mathbb{R}^{k \times k}$ mit den Einträgen

$$d_{ij} = \text{Cov}(X_i, X_j), \quad 1 \leq i, j \leq k.$$

Es gilt, dass für $A \in \mathbb{R}^{k \times m}$

$$\text{Var}(A\mathbf{X}) = A \text{Var}(\mathbf{X}) A^\top. \quad (1.12)$$

Weiterhin ist $\text{Var}(\mathbf{X} - \mathbf{c}) = \text{Var}(\mathbf{X})$ für jedes $\mathbf{c} \in \mathbb{R}^k$.

Lemma 1.20. Ist $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$, so gilt

$$\begin{aligned} \mathbb{E}(\mathbf{X}) &= \boldsymbol{\mu} \\ \text{Var}(\mathbf{X}) &= \Sigma. \end{aligned}$$

Beweis. Nach Definition ist

$$\mathbb{E}(\mathbf{X}) = \mathbb{E}(L\mathbf{Z} + \boldsymbol{\mu}) = L\mathbb{E}(\mathbf{Z}) + \boldsymbol{\mu} = \boldsymbol{\mu}.$$

Für die Varianz-Kovarianz Matrix nutzen wir Gleichung (1.12). Damit folgt, dass

$$\Sigma = \text{Var}(\mathbf{X}) = \text{Var}(\boldsymbol{\mu} + L\mathbf{Z}) = \text{Var}(L\mathbf{Z}) = LL^\top,$$

da die Varianz-Kovarianz Matrix von \mathbf{Z} gerade die Einheitsmatrix ist. \square

Mit $|\Sigma|$ sei die Determinante von Σ bezeichnet. Ist $\text{Rang}(\Sigma) = k$ und $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$, so hat \mathbf{X} die Dichte

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

Der Beweis wird in Aufgabe 1.37 geführt. Wie man sieht, ist die Abhängigkeit von multivariat normalverteilten Zufallsvariablen durch ihre Varianz-Kovarianz Matrix festgelegt. Insbesondere folgt in einer multivariaten Normalverteilung aus einer verschwindenden Kovarianz bereits die Unabhängigkeit, genauer: ist $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ und gilt $\text{Cov}(X_i, X_j) = 0$, so sind X_i und X_j unabhängig. Dieser Sachverhalt soll in Aufgabe 1.39 bewiesen werden.

Bemerkung 1.21. Weiterhin gelten folgende Resultate (vgl. Georgii (2004), Abschnitt 9.1).

- (i) $\Sigma = LL^\top$ ist symmetrisch und *nicht negativ definit*, denn

$$\mathbf{u}^\top \Sigma \mathbf{u} = \mathbf{u}^\top LL^\top \mathbf{u} = \|L^\top \mathbf{u}\|^2 \geq 0 \quad \forall \mathbf{u} \in \mathbb{R}^k.$$

- (ii) $\text{Rang}(\Sigma) = \text{Rang}(LL^\top) = \text{Rang}(L)$. Damit ist für $k = m$ die Matrix Σ nicht singulär, andernfalls singulär.
 (iii) Die Normalverteilung ist stabil unter linearen Transformationen: Falls $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ und $C \in \mathbb{R}^{n \times k}$, so gilt

$$C\mathbf{X} \sim \mathcal{N}_n(C\boldsymbol{\mu}, C\Sigma C^\top).$$

- (iv) Die einzelnen Komponenten einer multivariaten Normalverteilung sind normalverteilt: Falls $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$, so ist $X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$ für $i = 1, \dots, k$. Weiterhin folgt aus $\Sigma = I_k$, dass X_1, \dots, X_k unabhängige Zufallsvariablen sind (vgl. Aufgabe 1.39).

1.3 Bedingte Verteilungen

Die Einführung in die notwendigen Hilfsmittel wird in diesem Kapitel mit bedingten Verteilungen und dem bedeutsamen bedingten Erwartungswert fortgesetzt.

Bedingte Verteilungen. Bedingte Verteilungen verallgemeinern den Begriff der bedingten Wahrscheinlichkeit wesentlich und bilden ein wichtiges Hilfsmittel, zum Beispiel in der Schätztheorie.

Im diskreten Fall geht man eigentlich analog zu dem schon eingeführten Begriff der bedingten Wahrscheinlichkeit vor. Seien X, Y diskrete Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsfunktion $p(x, y)$. Y habe die Wahrscheinlichkeitsfunktion $p_Y(\cdot)$. Die *bedingte Verteilung* von X gegeben $Y = y$ mit $\mathbb{P}(Y = y) > 0$ ist definiert durch die Wahrscheinlichkeitsfunktion

$$p(x|y) := \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p(x, y)}{p_Y(y)}. \quad (1.13)$$

Für stetige Zufallsvariablen X, Y hat man analog folgende Situation: Ist die gemeinsame Dichte $p(x, y)$ und die Dichte von Y gerade $p_Y(\cdot)$, so definiert

man für diejenigen y mit $p_Y(y) > 0$

$$p(x|y) := \frac{p(x, y)}{p_Y(y)}. \quad (1.14)$$

B 1.3 Bernoulli-Verteilung: Die Summe von unabhängigen Bernoulli-Zufallsvariablen ist binomialverteilt: Eine Zufallsvariable X heißt Bernoulli-verteilt, falls $X \in \{0, 1\}$ und $\mathbb{P}(X = 0) \neq 0$. Seien X_1, \dots, X_n i.i.d. und Bernoulli-verteilt mit $\mathbb{P}(X_1 = 1) = p$, dann ist $Y := \sum_{i=1}^n X_i$ gerade $\text{Bin}(n, p)$ -verteilt (siehe Aufgabe 1.4).

B 1.4 Fortsetzung: Setze $\mathbf{X} = (X_1, \dots, X_n)^\top$. Dann ist die Verteilung von \mathbf{X} gegeben Y gerade eine Gleichverteilung: Für $\mathbf{x} \in \{0, 1\}^n$ mit $\sum_{i=1}^n x_i = y$ gilt

$$\mathbb{P}(\mathbf{X} = \mathbf{x} | Y = y) = \frac{\mathbb{P}(\mathbf{X} = \mathbf{x}, Y = y)}{\mathbb{P}(Y = y)} = \frac{p^y (1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \binom{n}{y}^{-1}.$$

So hat $\mathbf{X} | Y = y$ eine Gleichverteilung auf $\{\mathbf{x} \in \{0, 1\}^n : \sum_{i=1}^n x_i = y\}$.

Definition 1.22. Seien X und Y diskrete Zufallsvariablen, X nehme die Werte x_1, x_2, \dots an und es gelte $\mathbb{E}(|X|) < \infty$. Der *bedingte Erwartungswert* von X gegeben $Y = y$ ist für jedes y mit $\mathbb{P}(Y = y) > 0$ definiert durch

$$\mathbb{E}(X | Y = y) := \sum_{i \geq 1} x_i p(x_i | y).$$

Sind X, Y stetige Zufallsvariablen mit $\mathbb{E}(|X|) < \infty$, so ist der *bedingte Erwartungswert* von X gegeben $Y = y$ mit $p_Y(y) > 0$ definiert durch

$$\mathbb{E}(X | Y = y) := \int_{\mathbb{R}} x p(x | y) dx.$$

Sei $g(y) := \mathbb{E}(X | Y = y)$, dann heißt die Zufallsvariable

$$\mathbb{E}(X | Y) := g(Y)$$

bedingter Erwartungswert von X gegeben Y .

Der bedingte Erwartungswert von X gegeben Y bildet im quadratischen Mittel die beste Vorhersage von X , falls man Y beobachtet (siehe Aufgabe 1.20).

B 1.5 Suffiziente Statistik in der Bernoulli-Verteilung: Wir setzen Beispiel 1.3 fort und betrachten X_1, \dots, X_n i.i.d. $\text{Bin}(1, p)$ sowie $Y := \sum_{i=1}^n X_i$. Dann gilt für $y \in \{0, \dots, n\}$

$$\begin{aligned}
\mathbb{E}(X_1|Y=y) &= \mathbb{P}(X_1=1|Y=y) \\
&= \frac{p \binom{n-1}{p-1} p^{y-1} (1-p)^{(n-1)-(y-1)}}{\binom{n}{y} p^y (1-p)^{n-y}} \\
&= \binom{n-1}{y-1} \cdot \binom{n}{y}^{-1} \\
&= \frac{y}{n}.
\end{aligned}$$

Damit ergibt sich $\mathbb{E}(X_1|Y) = Yn^{-1}$. Man beachte, dass dies eine Zufallsvariable ist. Der Erwartungswert von X_1 gegeben der Statistik Y hängt nicht mehr vom Parameter p ab. Dies steht im Zusammenhang mit dem in Definition 2.5 eingeführten Begriff von Suffizienz.

Bemerkung 1.23. Sind X und Y unabhängig, so gibt Y keine neue Information über X und der bedingte Erwartungswert ist gleich dem unbedingten Erwartungswert: Unter $p_Y(y) > 0$ gilt, dass

$$p(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x)$$

und somit $\mathbb{E}(X|Y=y) = \mathbb{E}(X)$ und auch $\mathbb{E}(X|Y) = \mathbb{E}(X)$.

Bedingte Erwartungswerte lassen sich analog auf mehrdimensionale Zufallsvariablen verallgemeinern. Betrachtet man die zwei Zufallsvariablen $\mathbf{X} = (X_1, \dots, X_n)^\top$ und $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ und beide sind entweder diskret mit gemeinsamer Wahrscheinlichkeitsfunktion $\mathbb{P}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = p(\mathbf{x}, \mathbf{y})$ oder stetig mit gemeinsamer Dichte $p(\mathbf{x}, \mathbf{y})$, so definiert man analog zu (1.13) und (1.14) die bedingte Wahrscheinlichkeitsfunktion bzw. Dichte von \mathbf{X} gegeben $\mathbf{Y} = \mathbf{y}$ für alle \mathbf{y} mit $p_Y(\mathbf{y}) > 0$ durch

$$p(\mathbf{x}|\mathbf{y}) := \frac{p(\mathbf{x}, \mathbf{y})}{p_Y(\mathbf{y})}.$$

Ist $\mathbb{E}(|\mathbf{X}|) < \infty$, so ist der *bedingte Erwartungswert* von \mathbf{X} gegeben $\mathbf{Y} = \mathbf{y}$ definiert durch

$$\mathbb{E}(\mathbf{X} | \mathbf{Y} = \mathbf{y}) = (\mathbb{E}(X_1|\mathbf{Y} = \mathbf{y}), \dots, \mathbb{E}(X_n|\mathbf{Y} = \mathbf{y}))^\top.$$

Mit $g(\mathbf{y}) := \mathbb{E}(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ definieren wir den bedingten Erwartungswert von \mathbf{X} gegeben \mathbf{Y} durch

$$\mathbb{E}(\mathbf{X} | \mathbf{Y}) := g(\mathbf{Y}).$$

Satz 1.24 (Substitutionssatz). *Sei $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ eine messbare Abbildung. Gilt für $\mathbf{y} \in \mathbb{R}^m$, dass $p_Y(\mathbf{y}) > 0$ und $\mathbb{E}(|g(\mathbf{X}, \mathbf{y})|) < \infty$, so ist*

$$\mathbb{E}(g(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = \mathbb{E}(g(\mathbf{X}, \mathbf{y}) \mid \mathbf{Y} = \mathbf{y}).$$

Ein typischer Spezialfall ist $g(\mathbf{X}, \mathbf{y}) = r(\mathbf{X})h(\mathbf{y})$ mit einer beschränkten Funktion h . Hat $r(\mathbf{X})$ eine endliche Erwartung, so ist

$$\mathbb{E}(r(\mathbf{X})h(\mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = \mathbb{E}(r(\mathbf{X})h(\mathbf{y}) \mid \mathbf{Y} = \mathbf{y}) = h(\mathbf{y}) \mathbb{E}(r(\mathbf{X}) \mid \mathbf{Y} = \mathbf{y}).$$

Daraus folgt $\mathbb{E}(r(\mathbf{X})h(\mathbf{Y}) \mid \mathbf{Y}) = h(\mathbf{Y})\mathbb{E}(r(\mathbf{X}) \mid \mathbf{Y})$. Oft hat man die zusätzliche Annahme, dass \mathbf{X} und \mathbf{Y} unabhängig sind. Dann folgt unter den obigen Annahmen sogar, dass

$$\mathbb{E}(g(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = \mathbb{E}(g(\mathbf{X}, \mathbf{y})). \quad (1.15)$$

Der Erwartungswert der bedingten Erwartung ist gleich dem Erwartungswert selbst. Dies ist Inhalt des Satzes vom *iterierten Erwartungswert*.

Satz 1.25. *Gilt $\mathbb{E}(|\mathbf{X}|) < \infty$, so ist*

$$\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbb{E}(\mathbf{X} \mid \mathbf{Y})).$$

Beweis. Wir beweisen den eindimensionalen Fall, der mehrdimensionale Fall folgt analog. Zunächst seien X und Y diskrete Zufallsvariablen, mit Werten $\{x_1, x_2, \dots\}$ bzw. $\{y_1, y_2, \dots\}$, mit $p_Y(y_i) > 0$ für $i = 1, 2, \dots$. Dann gilt

$$\begin{aligned} \mathbb{E}(\mathbb{E}(X \mid Y)) &= \sum_{i \geq 1} p_Y(y_i) \left(\sum_{j \geq 1} x_j p(x_j \mid y_i) \right) \\ &= \sum_{i, j \geq 1} \frac{x_j p(x_j, y_i)}{p_Y(y_i)} p_Y(y_i) = \sum_{i, j \geq 1} x_j p(x_j, y_i) \\ &= \sum_{j \geq 1} x_j p_X(x_j) = \mathbb{E}(X). \end{aligned}$$

Für den Beweis des stetigen Falles sei auf Aufgabe 1.19 verwiesen. \square

Ordnet man eine Stichprobe X_1, \dots, X_n der Größe nach und bezeichnet man mit $X_{(1)}, \dots, X_{(n)}$ die geordneten Größen, so nennt man $X_{(1)}, \dots, X_{(n)}$ *Ordnungsgrößen* oder *Ordnungstatistiken* der Stichprobe. Die kleinste Ordnungsgröße $X_{(1)}$ ist das Minimum der Daten und die größte Ordnungsgröße

$X_{(n)}$ das Maximum. Wie im folgenden Beispiel kann man die Verteilung dieser Größen berechnen, wenn die Daten unabhängig sind.

B 1.6 *Minima und Maxima von gleichverteilten Zufallsvariablen:* Seien X_1, X_2 unabhängig und jeweils $U(0, 1)$ -verteilt. Setze $Y := \min(X_1, X_2)$ und $Z := \max(X_1, X_2)$. Im Folgenden seien x, y, z stets in $(0, 1)$. Die gemeinsame Verteilungsfunktion von Y und Z ist

$$\begin{aligned} F(y, z) &= \mathbb{P}(Y \leq y, Z \leq z) = 2 \mathbb{P}(X_1 < X_2, X_1 \leq y, X_2 \leq z) \\ &= 2 \int_0^z \int_0^{\min(x_2, y)} dx_1 dx_2 = 2 \cdot \begin{cases} \frac{z^2}{2}, & z < y \\ zy - \frac{y^2}{2}, & z \geq y \end{cases}. \end{aligned}$$

Die gemeinsame Dichte erhält man durch Ableiten der Verteilungsfunktion:

$$p(y, z) = \frac{\partial^2 F(y, z)}{\partial y \partial z} = 2 \begin{cases} 0, & z < y \\ 1, & z \geq y \end{cases} = 2 \mathbb{1}_{\{z \geq y\}}.$$

Die Dichte von Y ist

$$p_Y(y) = \int_0^1 p(y, z) dz = \int_y^1 2 dz = 2(1 - y).$$

Damit zeigt sich, dass das Maximum Z gegeben Y auf $(y, 1)$ gleichverteilt ist:

$$p(z|Y = y) = \frac{p(y, z)}{p_Y(y)} = \frac{1}{(1 - y)} \mathbb{1}_{\{z \geq y\}}.$$

1.4 Grenzwertsätze

In diesem Abschnitt stellen wir die fundamentalen Grenzwertsätze für arithmetische Mittel vor. Der erste, das Gesetz der großen Zahl, zeigt die Konvergenz des arithmetischen Mittels gegen den Erwartungswert. Das zweite Gesetz, der zentrale Grenzwertsatz, bestimmt die Grenzverteilung des mit \sqrt{n} skalierten arithmetischen Mittels: Die Normalverteilung. Beide Gesetze sind für asymptotische Aussagen (Konsistenz) und zur Verteilungsapproximation bei hinreichend großer Stichprobenzahl in der Statistik von unerlässlicher Bedeutung. Für Beweise der Aussagen verweisen wir auf Georgii (2004), Kapitel 5.

Das Gesetz der großen Zahl stellen wir in seiner schwachen und starken Form vor. In der schwachen Form konvergiert das arithmetische Mittel stochastisch, in der starken Form sogar mit Wahrscheinlichkeit 1.

Wir betrachten stets einen festen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 1.26. Seien X, X_1, X_2, \dots Zufallsvariablen. Die Folge $(X_n)_{n \geq 1}$ *konvergiert stochastisch* gegen X , falls für jedes $\epsilon > 0$ gilt, dass

$$\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow \infty} 0.$$

Die Folge $(X_n)_{n \geq 1}$ *konvergiert fast sicher* gegen X , falls

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Für die beiden Konvergenzarten verwenden wir folgende kompakte Notation: Konvergiert die Folge (X_n) stochastisch gegen X , so schreiben wir

$$X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X.$$

Konvergiert sie hingegen fast sicher, so schreiben wir

$$X_n \xrightarrow[n \rightarrow \infty]{f.s.} X.$$

Aus der fast sicheren Konvergenz folgt stochastische Konvergenz. Die Umkehrung gilt jedoch nicht.

Für die Konvergenz von Zufallsvariablen unter Transformationen hat man folgendes *Continuous Mapping Theorem*:

Satz 1.27. *Konvergiert die Folge $(X_n)_{n \geq 1}$ stochastisch gegen X und ist die Abbildung g stetig, so gilt*

$$g(X_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} g(X).$$

Sei M die Menge der Stetigkeitspunkte der Abbildung g , dann gilt der Satz auch, falls nur $\mathbb{P}(X \in M) = 1$, wenn g somit F_X -fast sicher stetig ist. Darüber hinaus gilt der Satz auch, wenn man an Stelle von stochastischer Konvergenz fast sichere oder Konvergenz in Verteilung (wie im folgenden zentralen Grenzwertsatz, Satz 1.31) schreibt. Der dazugehörige Beweis findet sich bei Serfling (1980), Abschnitt 1.7 auf S. 24.

Das schwache Gesetz der großen Zahl beweist man mit der Tschebyscheff-Ungleichung, welche sich unmittelbar aus der folgenden Markov-Ungleichung ergibt. Wir setzen $\mathbb{R}^+ := \{x \in \mathbb{R} : x \geq 0\}$.

Satz 1.28 (Markov-Ungleichung). *Sei $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ eine monoton wachsende Funktion und $f(x) > 0$ für $x > 0$. Dann gilt für alle $\epsilon > 0$, dass*

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}(f(|X|))}{f(\epsilon)}.$$

Als Spezialfall erhält man mit $f(x) = x^2$ die *Tschebyscheff-Ungleichung*:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}. \quad (1.16)$$

Satz 1.29 (Schwaches Gesetz der großen Zahl). *Seien X_1, X_2, \dots paarweise unkorreliert mit $\mathbb{E}(X_i) = \mathbb{E}(X_1)$ und $\text{Var}(X_i) < M < \infty$ für alle $i \geq 1$ und ein $M \in \mathbb{R}$. Dann gilt, dass*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbb{E}(X_1).$$

Beweis. Betrachtet man das arithmetische Mittel $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$, so ist $\mathbb{E}(\bar{X}) = \mathbb{E}(X_1)$. Mit der Regel von Bienaymé, (1.4), erhält man

$$\text{Var}(\bar{X}) = \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2} \leq \frac{M}{n}.$$

Damit folgt für jedes $\epsilon > 0$ aus der Tschebyscheff-Ungleichung (1.16), dass

$$\mathbb{P}(|\bar{X} - \mathbb{E}(X_1)| \geq \epsilon) \leq \frac{M}{n\epsilon^2} \xrightarrow[n \rightarrow \infty]{} 0$$

und somit die Behauptung. \square

Die Aussage des schwachen Gesetzes der großen Zahl kann man wesentlich verschärfen. Wir geben eine Version mit den geringsten Integrabilitätsbedingungen an, und setzen lediglich die Existenz der Erwartungswerte der X_i voraus. Im Gegenzug müssen wir verlangen, dass die X_i i.i.d. sind. Die Aussage des folgenden Satzes gilt aber auch unter den Voraussetzungen aus Satz 1.29, allerdings dann mit der Annahme existierender Varianzen.

Satz 1.30 (Starkes Gesetz der großen Zahl). *Seien X_1, X_2, \dots i.i.d. mit $\mathbb{E}(|X_1|) < \infty$. Dann gilt*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{f.s.} \mathbb{E}(X_1).$$

Für den Beweis sei auf Gut (2005), Kapitel 6.6 (Seite 294 – 298) verwiesen.

Schließlich geben wir den zentralen Grenzwertsatz an. Sei Φ die Verteilungsfunktion der Standardnormalverteilung, d.h.

$$\Phi(z) = \int_0^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Satz 1.31 (Zentraler Grenzwertsatz). *Seien X_1, X_2, \dots i.i.d. mit $\mathbb{E}(X_1) := \mu$ und $\text{Var}(X_1) := \sigma^2 < \infty$. Dann gilt*

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \leq z\right) \xrightarrow[n \rightarrow \infty]{} \Phi(z)$$

für alle $z \in \mathbb{R}$.

Die in dem Satz auftretende Konvergenz nennt man auch *Verteilungskonvergenz*, hier gegen die Standardnormalverteilung $\mathcal{N}(0, 1)$. Mit $C(F_X) := \{x \in \mathbb{R} : F_X(x) \text{ ist stetig an } x\}$ bezeichnen wir die Menge der Stetigkeitspunkte der Verteilungsfunktion von X , F_X .

Definition 1.32. Die Folge von Zufallsvariablen $(X_n)_{n \geq 1}$ *konvergiert in Verteilung* gegen X , falls für alle $x \in C(F_X)$ gilt, dass

$$F_{X_n}(x) \rightarrow F_X(x), \quad n \rightarrow \infty.$$

Konvergiert eine Folge $(X_n)_{n \geq 1}$ in Verteilung gegen die Standardnormalverteilung, so schreiben wir kurz

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Das mehrdimensionale Analogon von Satz 1.31 nennt man den multivariaten zentralen Grenzwertsatz. Hier gibt es eine Vielzahl von Varianten und

wir zitieren die Version für eine Folge von unabhängigen, identisch verteilten Zufallsvektoren aus Bauer (1990) (Satz 30.3, Seite 265).

Mit $\Phi_k(\mathbf{z}; \mathbf{0}, \Sigma)$ ist die Verteilungsfunktion einer k -dimensionalen, normalverteilten Zufallsvariablen mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix Σ bezeichnet, siehe auch Abschnitt 1.2.1.

Satz 1.33. *Seien die k -dimensionalen Zufallsvariablen $\mathbf{X}_1, \mathbf{X}_2, \dots$ i.i.d. und $\mathbb{E}(X_{ij}^2) < \infty$ für alle $1 \leq i \leq k$ und $j \geq 1$. Setze $\boldsymbol{\mu} := \mathbb{E}(\mathbf{X}_1)$ und $\Sigma := \text{Var}(\mathbf{X}_1)$. Dann gilt für alle $\mathbf{z} \in \mathbb{R}^k$, dass*

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \leq \mathbf{z}\right) \xrightarrow{n \rightarrow \infty} \Phi_k(\mathbf{z}; \mathbf{0}, \Sigma).$$

Für die aus dem Satz resultierende (multivariate) Verteilungskonvergenz schreibt man auch

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{X}_i - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}_k(\mathbf{0}, \Sigma).$$

Der folgende Satz erlaubt es die Bildung eines Grenzwertes mit dem Erwartungswert unter einer Zusatzbedingung, der Monotonie der zu betrachtenden Folge, zu vertauschen. Eine Alternative zu dieser Zusatzbedingung liefert der Satz der dominierten Konvergenz. Für einen Beweis beider Aussagen siehe Irle (2005), Satz 8.15 auf Seite 114.

Satz 1.34 (Monotone Konvergenz). *Sei X_1, X_2, \dots eine Folge von Zufallsvariablen. Gilt $0 \leq X_1 \leq X_2 \leq \dots$, so folgt*

$$\mathbb{E}\left(\lim_{n \rightarrow \infty} X_n\right) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n).$$

1.4.1 Referenzen

Grenzwertsätze sind ein wichtiges Hilfsmittel in der Statistik und werden in diesem Kapitel nur knapp behandelt. Für eine Vertiefung sei auf die vielfältige Literatur verwiesen: Chung (2001), Kapitel 4 in Gänsler und Stute (1977), Kapitel 9 in Resnick (2003), Billingsley (1986) und Kapitel 15 in Klenke (2008).

1.5 Aufgaben

A 1.1 *Die Potenzmenge ist eine σ -Algebra:* Sei Ω eine Menge (etwa eine endliche Menge). Die Potenzmenge

$$\mathcal{P}(\Omega) := \{A : A \subset \Omega\}$$

ist eine σ -Algebra.

A 1.2 *Unkorreliertheit impliziert nicht Unabhängigkeit:* Sei $X \sim \mathcal{N}(0, 1)$ eine standardnormalverteilte Zufallsvariable und $Y = X^2$. Dann ist $\text{Cov}(X, Y^2) = 0$, aber X und Y sind nicht unabhängig.

A 1.3 *Erwartungstreue der Stichprobenvarianz:* Seien X_1, \dots, X_n i.i.d. mit Varianz σ^2 . Die Stichprobenvarianz ist definiert durch

$$s^2(\mathbf{X}) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Dann gilt $\mathbb{E}(s^2(\mathbf{X})) = \sigma^2$, d.h. die Stichprobenvarianz ist erwartungstreu.

A 1.4 *Darstellung der Binomialverteilung als Summe von unabhängigen Bernoulli-Zufallsvariablen:* Seien X_1, \dots, X_n i.i.d. mit $X_i \in \{0, 1\}$ und $\mathbb{P}(X_i = 1) = p \in (0, 1)$, $1 \leq i \leq n$. Dann ist

$$\sum_{i=1}^n X_i \sim \text{Bin}(n, p).$$

A 1.5 *Erwartungswert und Varianz der Poisson-Verteilung:* Zeigen Sie, dass für eine zum Parameter λ Poisson-verteilte Zufallsvariable X gilt, dass

$$\mathbb{E}(X) = \text{Var}(X) = \lambda.$$

A 1.6 *Gedächtnislosigkeit der Exponentialverteilung:* Sei X exponentialverteilt mit Intensität λ . Dann gilt für $x, h > 0$

$$\mathbb{P}(X > x + h \mid X > x) = \mathbb{P}(X > h).$$

A 1.7 *Gamma-Verteilung: Unabhängigkeit von bestimmten Quotienten:* Seien $X \sim \text{Gamma}(a, \lambda)$ und $Y \sim \text{Gamma}(b, \lambda)$ zwei unabhängige Zufallsvariablen. Zeigen Sie, dass $\frac{X}{X+Y}$ und $X+Y$ unabhängig sind.

A 1.8 *Quotienten von Gamma-verteilten Zufallsvariablen:* Seien X und Y unabhängig mit $X \sim \text{Exp}(\beta)$ und $Y \sim \text{Gamma}(a, \lambda)$ und $a > 1$. Zeigen Sie, dass

$$\mathbb{E}\left(\frac{X}{Y}\right) = \frac{\lambda}{\beta(a-1)}.$$

A 1.9 *Transformationen von Gamma-verteilten Zufallsvariablen:* Seien die Zufallsvariablen $X \sim \text{Gamma}(a, \lambda)$ und $Y \sim \text{Gamma}(b, \lambda)$ unabhängig und $c > 0$. Dann gilt

$$(i) \quad X + Y \sim \text{Gamma}(a + b, \lambda),$$

$$(ii) \quad \frac{X}{X + Y} \sim \text{Beta}(a, b),$$

$$(iii) \quad cX \sim \text{Gamma}(a, \frac{\lambda}{c}).$$

Momente und momentenerzeugende Funktion

A 1.10 *Erwartungswert des Betrages einer Normalverteilung:* Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit einem $\mu \in \mathbb{R}$ und einem $\sigma > 0$. Berechnen Sie den Erwartungswert von $|X|$.

A 1.11 *Momente der Normalverteilung:* Zeigen Sie, dass für eine standardnormalverteilte Zufallsvariable X und $n \in \mathbb{N}$ gilt, dass

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n \cdot n!}.$$

A 1.12 *Momentenerzeugende Funktion einer Gamma-Verteilung:* Es gelte, dass $X \sim \text{Gamma}(a, \lambda)$. Zeigen Sie, dass für $s < \lambda$

$$\Psi_X(s) = \mathbb{E}(e^{sX}) = \frac{\lambda^a}{(\lambda - s)^a}$$

gilt. Bestimmen Sie damit den Erwartungswert und die Varianz von X .

A 1.13 *Momente der Beta-Verteilung:* Bestimmen Sie den Erwartungswert und die Varianz einer $\text{Beta}(a, b)$ -Verteilung.

A 1.14 *Zweiseitige Exponentialverteilung:* Man nehme an, dass die Zufallsvariablen X_1 und X_2 unabhängig und exponentialverteilt sind mit $X_i \sim \text{Exp}(\lambda)$, $i = 1, 2$.

(i) Zeigen Sie, dass $Y := X_1 - X_2$ die Dichte

$$p(y) = \frac{1}{2} \lambda e^{-\lambda|y|}$$

besitzt. Y nennt man dann zweiseitig exponentialverteilt (allerdings mit gleichem Parameter für die linke und rechte Halbachse).

(ii) Berechnen Sie die momentenerzeugende Funktion von Y .

A 1.15 *Existenz von Momenten niedrigerer Ordnung:* Sei X eine (stetige) reellwertige Zufallsvariable. Die so genannte L^p -Norm von X ist definiert durch $\|X\|_p := (\mathbb{E}(|X|^p))^{1/p}$. Zeigen Sie, dass für $n \in \mathbb{N}$

$$(\|X\|_n)^n \leq 1 + (\|X\|_{n+1})^{n+1}.$$

A 1.16 *Lévy-Verteilung:* Sei X_1, \dots, X_n i.i.d. und X_1 sei Lévy verteilt zu den Parametern $\gamma, \delta > 0$, d.h. X_1 hat die Dichte

$$p(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x - \delta)^{3/2}} e^{-\frac{\gamma}{2(x - \delta)}} \mathbb{1}_{\{x > \delta\}}.$$

Der Parameter δ sei bekannt. Bestimmen Sie die Momenterzeugende Funktion von

$$T(\mathbf{X}) := \sum_{i=1}^n \frac{1}{X_i - \delta}$$

und geben Sie explizit deren Definitionsbereich an. Berechnen Sie $\mathbb{E}(T(\mathbf{X}))$ und $\text{Var}(T(\mathbf{X}))$.

A 1.17 *Momentenerzeugende Funktion und Momente der Poisson-Verteilung:* Sei $X \sim \text{Pois}(\lambda)$ mit $\lambda > 0$.

(i) Zeigen Sie, dass die momentenerzeugende Funktion von X gegeben ist durch

$$\Psi_X(s) = \exp(\lambda(e^s - 1)), \quad s \in \mathbb{R}.$$

(ii) Verwenden Sie (i) um zu zeigen, dass

$$\mathbb{E}((X - \lambda)^4) = \lambda + 3\lambda^2.$$

Regeln für bedingten Verteilungen

A 1.18 *Die bedingte Verteilung ist ein Wahrscheinlichkeitsmaß:* Sei $B \in \mathcal{A}$ ein Ereignis mit $\mathbb{P}(B) > 0$. Dann ist durch

$$\mu(A) := \mathbb{P}(A|B) : \mathcal{A} \rightarrow [0, 1]$$

ein Wahrscheinlichkeitsmaß definiert.

A 1.19 *Erwartungswert der bedingten Erwartung:* Sei X eine Zufallsvariable mit Dichte p_X und $\mathbb{E}(|X|) < \infty$. Dann gilt für jede Zufallsvariable Y , dass

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)).$$

A 1.20 *Der bedingte Erwartungswert als beste Vorhersage:* Im quadratischen Mittel ist der bedingte Erwartungswert die beste Vorhersage der Zufallsvariablen X , wenn man Y beobachtet. Hierzu seien X und Y Zufallsvariablen mit endlicher Varianz. Zeigen Sie, dass für alle meßbaren Funktionen $g : \mathbb{R} \rightarrow \mathbb{R}$ gilt:

$$\mathbb{E}((X - g(Y))^2) \geq \mathbb{E}((X - \mathbb{E}(X|Y))^2).$$

- A 1.21** *Perfekte Vorhersagen:* Seien X, Y reellwertige Zufallsvariablen mit endlicher Varianz. Finden Sie ein nichttriviales Beispiel für folgenden Sachverhalt: Bei Kenntnis der Realisation von Y kann die Realisation von X perfekt vorhergesagt werden in dem Sinn, dass

$$\mathbb{E}(X|Y) = X \text{ und } \text{Var}(X|Y) = 0.$$

Andererseits bringt die Kenntnis der Realisation von X keine Information über die Realisation von Y , in dem Sinne, dass

$$\text{Var}(Y|X) = \text{Var}(Y).$$

Ein triviales Beispiel ist wie folgt: X ist konstant und Y eine beliebige, reelle Zufallsvariable mit endlicher Varianz.

- A 1.22** *Bedingte Dichte: Beispiele:* Sei (X, Y) ein Zufallsvektor mit der Dichte

$$f(x, y) = \frac{3}{5} y (x + y) \mathbb{1}_{\{0 < x < 2, 0 < y < 1\}}.$$

Bestimmen Sie die bedingte Dichte $f_{Y|X=x}(y)$, $y \in \mathbb{R}$, $x \in (0, 2)$ und zeigen Sie damit, dass $\mathbb{P}(Y \leq \frac{1}{2} | X = 1) = \frac{1}{5}$. Zeigen Sie weiterhin, dass $\text{Cov}(X + Y, X - Y) = \frac{73}{100}$.

- A 1.23** *Poisson-Binomial Mischung:* X sei Poisson(λ)-verteilt. Bedingt auf $\{X = k\}$ sei Y binomialverteilt mit Parameter (k, p) :

$$\mathbb{P}(Y = l | X = k) = \binom{k}{l} p^l (1-p)^{k-l}, \quad 0 \leq l \leq k;$$

mit $p \in (0, 1)$. Zeigen Sie mit Hilfe der momentenerzeugenden Funktion, dass Y Poisson-verteilt zum Parameter λp ist.

- A 1.24** *Exponential-Exponential Mischung:* Die Zufallsvariable Y sei exponentialverteilt zum Parameter λ . Die Dichte der Zufallsvariablen X gegeben $\{Y = y\}$ sei die Dichte einer Exponentialverteilung mit Parameter y , also

$$f(x | y) = y e^{-yx} \mathbb{1}_{\{x > 0\}}.$$

Bestimmen Sie die bedingte Dichte von Y gegeben X .

- A 1.25** *Linearität des bedingten Erwartungswertes:* Seien X_1, X_2 und Y reelle Zufallsvariablen und $\mathbb{E}(|X_i|) < \infty$ für $i = 1, 2$. Dann gilt für alle $a, b \in \mathbb{R}$, dass

$$\mathbb{E}(aX_1 + bX_2 | Y) = a\mathbb{E}(X_1 | Y) + b\mathbb{E}(X_2 | Y).$$

- A 1.26** *Bedingte Varianz:* Seien X, Y reelle Zufallsvariablen mit $\mathbb{E}(X^2) < \infty$. Die bedingte Varianz einer Zufallsvariablen X gegeben Y ist definiert durch

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

Zeigen Sie, dass

$$\text{Var}(X) = \text{Var}(\mathbb{E}(X|Y)) + \mathbb{E}(\text{Var}(X|Y)).$$

A 1.27 *Satz von Bayes:* Seien X und Y Zufallsvariablen mit endlichem Erwartungswert. Bezeichne $q(y|x)$ die bedingte Dichte von Y gegeben X und $p(x|y)$ die bedingte Dichte von X gegeben Y . Weiterhin sei p_X die Dichte von X . Dann gilt

$$p(x|y) = \frac{p_X(x)q(y|x)}{\int_{\mathbb{R}} p_X(z)q(y|z)dz}.$$

Ebenso gilt ein analoges Resultat für k -dimensionale Zufallsvariablen.

A 1.28 *Exponentialverteilung: Diskretisierung:* Z sei exponentialverteilt mit Erwartungswert 1 und $X := [Z]$ die größte natürliche Zahl kleiner gleich Z . Bestimmen Sie die Verteilung von X und berechnen Sie damit $\mathbb{E}(Z|X)$.

A 1.29 *Erwartungswert einer zufälligen Summe:* Seien Y_1, Y_2, \dots i.i.d. mit $Y_i \geq 0$ und $\mathbb{E}(Y_1) < \infty$. Weiterhin sei N eine Zufallsvariable mit Werten in $0, 1, 2, \dots$, unabhängig von allen Y_i . Dann ist

$$\mathbb{E}\left(\sum_{i=0}^N Y_i\right) = \mathbb{E}(N)\mathbb{E}(Y_1). \quad (1.17)$$

Ist N Poisson-verteilt, so gilt $(1.17) = \lambda\mathbb{E}(Y_1)$.

Summen von Zufallsvariablen

Um die Verteilung von Summen unabhängiger Zufallsvariablen zu bestimmen, kann man zum einen mit der momentenerzeugenden Funktion oder der charakteristischen Funktion arbeiten, zum anderen auch mit der so genannten *Faltungsformel*.

A 1.30 *Faltungsformel:* Haben X und Y die Dichten p_X und p_Y und beide sind unabhängig, so ist die Dichte von $Z := X + Y$ gegeben durch

$$p_Z(z) = \int_{\mathbb{R}} p_X(x) p_Y(z-x) dx.$$

A 1.31 *Die Summe von normalverteilten Zufallsvariablen ist wieder normalverteilt:* Sind die Zufallsvariablen X_1, \dots, X_n unabhängig und normalverteilt mit $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, so ist die Summe wieder normalverteilt:

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Allgemeiner erhält man: Ist eine Zufallsvariable multivariat normalverteilt, $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$, so gilt

$$\mathbf{a}^\top \mathbf{X} \sim \mathcal{N}(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \Sigma \mathbf{a}).$$

A 1.32 *Dichte der χ^2 -Verteilung:* Seien X_1, \dots, X_n unabhängige und standardnormalverteilte Zufallsvariablen. Dann folgt $Y := \sum_{i=1}^n X_i^2$ einer χ^2 -Verteilung mit n Freiheitsgraden. Zeigen Sie, dass die Dichte von Y für $x > 0$ durch

$$p(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n-2}{2}}$$

gegeben ist. Verwenden Sie hierfür die Faltungsformel und die Beta-Funktion aus Gleichung (1.9).

A 1.33 *Wohldefiniertheit der nichtzentralen χ^2 -Verteilung:* Zeigen Sie, dass die Verteilung der $\chi_k^2(\theta)$ -Verteilung nur von $\theta = \sum_{i=1}^k \mu_i^2$ abhängt. Hierfür kann man die charakteristische oder die momentenerzeugende Funktion von Z^2 mit $Z \sim \mathcal{N}(\mu, 1)$ verwenden.

A 1.34 *Verteilung der Stichprobenvarianz:* Seien X_1, \dots, X_n i.i.d., normalverteilt und $\text{Var}(X_1) = \sigma^2$. Für das zentrierte empirische zweite Moment $\hat{\sigma}^2(\mathbf{X}) := n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ gilt, dass

$$\frac{n\hat{\sigma}^2(\mathbf{X})}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2.$$

A 1.35 *Mittelwertvergleich bei Gamma-Verteilungen:* Seien X_1, \dots, X_n i.i.d. und Gamma(a, λ_1)-verteilt, d.h. X_1 hat die Dichte

$$p_1(x) = \frac{\lambda_1^a}{\Gamma(a)} x^{a-1} e^{-\lambda_1 x} \mathbf{1}_{\{x>0\}}.$$

Außerdem seien Y_1, \dots, Y_n i.i.d. und Gamma(a, λ_2)-verteilt. Man nehme an, dass die Vektoren (X_1, \dots, X_n) und (Y_1, \dots, Y_n) unabhängig sind. Das arithmetische Mittel wird wie üblich mit \bar{X} bzw. \bar{Y} bezeichnet. Bestimmen Sie die Verteilung der Statistik $\frac{\bar{X}}{\bar{Y}}$.

A 1.36 *Rayleigh-Verteilung: Momente und Zusammenhang mit der Normalverteilung:* Seien X und Y unabhängig und $\mathcal{N}(0, \sigma^2)$ -verteilt. Dann ist

$$Z := \sqrt{X^2 + Y^2}$$

Rayleigh-verteilt, d.h. Z hat die Dichte $x\sigma^{-2} \exp(-x^2/2\sigma^2)$. Es gilt $\mathbb{E}(Z) = \sigma\sqrt{\pi/2}$, $\mathbb{E}(Z^2) = 2\sigma^2$ und $\text{Var}(Z) = \sigma\sqrt{2 - \pi/2}$.

Multivariate Normalverteilung

A 1.37 *Dichte der multivariaten Normalverteilung:* Zeigen Sie, dass $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ folgende Dichte hat, falls $\text{Rang}(\Sigma) = p$:

$$p(\mathbf{x}) = \frac{1}{\det(\Sigma)^{1/2} (2\pi)^{p/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

A 1.38 *Lineare Transformationen der Normalverteilung:* Sei $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ und $C \in \mathbb{R}^{n \times p}$. Dann gilt

$$C\mathbf{X} \sim \mathcal{N}_n(C\boldsymbol{\mu}, C\Sigma C^\top).$$

A 1.39 *Normalverteilung: Cov(X, Y) = 0 impliziert Unabhängigkeit:* Sei $\mathbf{Z} = (X, Y)^\top \in \mathbb{R}^2$ und $\mathbf{Z} \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$. Gilt $\text{Cov}(X, Y) = 0$, so sind X und Y unabhängig.

A 1.40 *Bedingte Verteilungen der multivariaten Normalverteilung:* Seien $\mathbf{X}_i, i = 1, 2$ zwei k_i -dimensionale Zufallsvariablen, so dass

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}_k\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12}^\top \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}\right);$$

hier ist $k = k_1 + k_2$, $\boldsymbol{\mu}_i \in \mathbb{R}^{k_i}$, $\Sigma_{11} \in \mathbb{R}^{k_1 \times k_1}$, $\Sigma_{12} \in \mathbb{R}^{k_2 \times k_1}$ und $\Sigma_{22} \in \mathbb{R}^{k_2 \times k_2}$. Dann ist die bedingte Verteilung von \mathbf{X}_1 gegeben \mathbf{X}_2 wieder eine Normalverteilung:

$$\mathbb{P}(\mathbf{X}_1 \leq \mathbf{x}_1 \mid \mathbf{X}_2 = \mathbf{x}_2) = \Phi_{k_1}(\mathbf{x}_1; \boldsymbol{\mu}(\mathbf{x}_2), \Sigma(\mathbf{x}_2))$$

mit

$$\boldsymbol{\mu}(\mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{11}^\top \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$\Sigma(\mathbf{x}_2) = \Sigma_{11} - \Sigma_{12}^\top \Sigma_{22}^{-1} \Sigma_{12}.$$

$\Phi_{k_1}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ bezeichnet die Verteilungsfunktion der k_1 -dimensionalen Normalverteilung mit Erwartungswert $\boldsymbol{\mu}$ und Kovarianzmatrix Σ an der Stelle \mathbf{x} .

Mathematische Statistik

Czado, C.; Schmidt, T.

2011, VIII, 280 S. 25 Abb., 3 Abb. in Farbe., Softcover

ISBN: 978-3-642-17260-1