Probability & Statistics

ETH Zurich

Janik Schuettler
Marcel Graetz

FS18

Inhaltsverzeichnis

0 Formeln

$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$, falls $A_i \cap A_j = \emptyset$ für $i \neq j$

$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{k+1} \sum_{i \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$

$\mathbb{P}(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$, falls $(A_i)_{i \in J}$ unabhängig

$\mathbb{P}(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i | \bigcap_{j=1}^{i-1} A_j) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\ldots$

$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

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

$\mathbb{P}(A) = \sum_{i:\mathbb{P}(B_i)>0} \mathbb{P}(A|B_i)\mathbb{P}(B_i) = \sum_{i:\mathbb{P}(B_i)>0} \mathbb{P}(A \cap B_i)$

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

$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$

1 Der Wahrscheinlichkeitsbegriff

Definition 1.1 (Wahrscheinlichkeitsraum) *ist das Triplet $(\Omega, \mathcal{A}, \mathbb{P})$, wobei Ω der **Grundraum** aller möglichen Ergebnisse oder Fälle ω, $\mathcal{A}$ die Klasse der beobachtbaren **Ereignisse** und $\mathbb{P} : \mathcal{A} \to [0,1]$ das **Wahrscheinlichkeitsmass** bezeichnet.*

2 Diskrete Wahrscheinlichkeitsräume

2.1 Grundbegriffe

Wahrscheinlichkeit für Elemente $\omega \in \Omega$ und Ereignisse $A \in \mathcal{A}$

$$p(\omega) := \mathbb{P}(\{\omega\}), \quad \mathbb{P}(A) = \sum_{\omega \in A} p(\omega),\ A \in \mathcal{A},$$

sodass $\sum_{\omega \in \Omega} p(\omega) = 1$ (Normierung).

Axiome *der Wahrscheinlichkeitsrechnung*

Axiom 1: Für jedes Ereignis A ist $\mathbb{P}(A) \geq 0$,

Axiom 2: $\mathbb{P}(\Omega) = 1$,

Axiom 3: $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$ für paarweise disjunkte $A_1, A_2, \ldots$

Rechenregeln als Folge der Axiome

$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

$\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n (-1)^{k+1} \sum_{i \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1}) \cap \ldots \cap \mathbb{P}(A_{i_k})$

$A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Zufallsvariable ist eine Abbildung $X : \Omega \to \mathbb{R}^1$. X ist eine diskrete Zufallsvariable, falls das Bild $X(\Omega)$ abzählbar ist.

Verteilungsfunktion ist die Abbildung $\nu : x \in X(\Omega) \to \mathbb{P}(X = x) \equiv \mathbb{P}(\{\omega \in \Omega | X(\omega) = x\})$

Erwartungswert direkt und über die Verteilungsfunktion

$$\begin{aligned} \mathbb{E}(X) &= \sum_{\omega \in \Omega} X(\omega)p(\omega) = \sum_{x \in X(\Omega)} \sum_{\omega : X(\omega) = x} X(\omega)p(\omega) \\ &= \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x\mathbb{P}(X^{-1}(\{x\})). \end{aligned}$$

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

Lemma 2.1 *Wenn X nur die Werte $0, 1, 2, \ldots$ annimmt, so ist*

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n).$$

2.2 Laplace-Modelle

Laplace-Modell besagt $p(\omega) = \frac{1}{|\Omega|} = const.$

2.3 Irrfahrten

2.3.1 Definition

Modell Ω die Menge aller binären Folgen der Länge N, $\Omega = \{\omega = (x_1, \ldots, x_N) | x_i \in \{1, -1\}\}$. $X_k(\omega) = k$-te Komponente von $\omega = (x_1, \ldots, x_N) \in \Omega$ und $S_n(\omega) = \sum_{k=1}^n X_k(\omega)$, wobei $S_0(\omega) = 0$.

Definition 2.2 (Irrfahrt) *Sei $(\Omega, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit Ω wie oben und $\mathbb{P}$ die Gleichverteilung auf Ω. Die Folge der Zufallsvariablen $S_n (n = 0, \ldots, N)$ auf $(\Omega, \mathbb{P})$ heisst **Irrfahrt** (mit Start in 0).*

Folgerungen

$\mathbb{P}(X_k = +1) = \frac{2^{N-1}}{2^N} = \frac{1}{2}$

$\mathbb{P}(X_{k_1} = x_{k_1}, \ldots, X_{k_l} = x_{k_l}) = \frac{2^{N-l}}{2^N} = 2^{-l}$

$\mathbb{E}(X_k) = (+1)\mathbb{P}(X_k = +1) + (-1)\mathbb{P}(X_k = -1) = 0$

$\mathbb{E}(S_n) = \sum_{k=1}^n \mathbb{E}(X_k) = 0$

Theorem 2.3 *Für festes n nimmt die Zufallsvariable S_n Werte $x \in \{-n, -n+2, \ldots, n-2, n\}$ an mit Wahrscheinlichkeiten*

$$\mathbb{P}(S_n = 2k - n) = \binom{n}{k} 2^{-n}, \quad k = 0, 1, \ldots, n.$$

Für ein anderes x ist $\mathbb{P}(S_n = x) = 0$.

Korollar 2.4

$$\mathbb{P}(S_{2n} = 0) = \mathbb{P}(S_{2n-1} = 1) = \binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}}$$

$$\binom{n}{k} = \binom{n}{k-1} \frac{n-k+1}{k} \geq \binom{n}{k-1} \iff k \leq \frac{n+1}{2}$$

Also ist für n gerade $\mathbb{P}(S_{2n} = x)$ maximal für $x = 0$, für n ungerade ist $\mathbb{P}(S_n = x)$ maximal für $x = \pm 1$.

2.3.2 Reflektionsprinzip

Definition 2.5 *Wir definieren die Zufallsvariable*

$$T_a(\omega) = \min\{n > 0 | S_n(\omega) = a\},$$

d.h. erstes Erreichen des Niveaus $a \neq 0$ bzw. für $a = 0$ erste Rückkehr nach 0. Hierfür setzen wir $\min \emptyset = N + 1$.

Lemma 2.6 (Reflektionsprinzip) *Für $a > 0$ und $b \geq -a$ ist*

$$\mathbb{P}(T_{-a} \leq n, S_n = b) = \mathbb{P}(S_n = -2a - b).$$

Theorem 2.7 *Für $a \neq 0$ gilt*

$$\mathbb{P}(T_{-a} \leq n) = 2\mathbb{P}(S_n < -a) + \mathbb{P}(S_n = -a) = \mathbb{P}(S_n \notin (-a, a]).$$

Korollar 2.8 *Für jedes $a \neq 0$ gilt*

$$\mathbb{P}(T_a > N) \overset{N \to \infty}{\to} 0,$$

$$\mathbb{E}(T_a) = \sum_{k=1}^{N+1} k\mathbb{P}(T_a = k) \overset{N \to \infty}{\to} \infty.$$

Theorem 2.9

$$\mathbb{P}(T_0 > 2n) = \mathbb{P}(S_{2n} = 0)$$

2.3.3 Das Arkussinus-Gesetz für den letzten Besuch in Null

Theorem 2.10 (Arkussinus-Gesetz) *Die Verteilung von*

$$L(\omega) = \max\{0 \leq n \leq 2N | S_n(\omega) = 0\}$$

ist die sogenannte diskrete Arkussinus-Verteilung

$$\mathbb{P}(L = 2n) = \mathbb{P}(S_{2n} = 0)\mathbb{P}(S_{2N-2n} = 0) = 2^{-2N} \binom{2n}{n} \binom{2N - 2n}{N - n}$$

.

2.3.4 Spielsysteme

Definition 2.11 (Beobachtbarkeit) *Ein Ereignis $A \subseteq \Omega$ heisst **beobachtbar bis zum Zeitpunkt** n, wenn es von der Form $\{\omega | (X_1(\omega), \ldots, X_n(\omega)) \in C\}$ ist für ein $C \subseteq \{-1, 1\}^n$. Die Menge aller bis zum Zeitpunkt n beobachteten Ereignisse bezeichnen wir mit $\mathcal{A}_n$. Für $n = 0$ definieren wir $\mathcal{A}_0 = \{\emptyset, \Omega\}$.*

Definition 2.12 (Stoppzeit) *ist eine Abbildung $T : \Omega \to \{0, \dots, N\}$, sodass gilt*

$$\{T = n\} = \{\omega | T(\omega) = n\} \in \mathcal{A}_n \quad (n = 0, \dots, N).$$

Theorem 2.13 *Für jede Stoppzeit T ist*

$$\mathbb{E}(S_T) = 0,$$

wobei $S_T(\omega) \equiv S_{T(\omega)}(\omega)$ den bei Benutzung der Stoppzeit T erzielten Betrag bezeichnet.

Definition 2.14 (Spielsystem) *ist eine Folge $V = (V_k)_{k=1,\dots,N}$ von Zufallsvariablen $V_k : \Omega \to \mathbb{R}$ derart, dass $V_1 = const$ und für $k = 2, 3, \dots, N$ existierten Funktionen $\varphi_k : \{-1, 1\} \to \mathbb{R}$ mit*

$$V_k(\omega) = \phi_k(X_1(\omega), \dots, X_{k-1}(\omega)).$$

Theorem 2.15 *Für Spielsystem $V = (V_k)_{k=1,\dots,N}$ ist der erwartete Ertrag*

$$\mathbb{E}((V \cdot S)_N) = 0.$$

Korollar 2.16 (Waldsche Identität) *Für jede Stoppzeit T ist*

$$\mathbb{E}(S_T) = 0 \quad \text{und} \quad \mathbb{V}(S_T) = \mathbb{E}(S_T^2) = \mathbb{E}(T).$$

2.4 Bedingte Wahrscheinlichkeiten

Definition 2.17 (Bedingte Wahrscheinlichkeit) *von A gegeben B mit $\mathbb{P}(B) > 0$ ist*

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

Theorem 2.18 (Totale Wahrscheinlichkeit) *Sei $(B_i)_{i \in I}$ eine disjunkte Zerlegung von Ω (d.h. $\Omega = \bigcup_{i \in I} B_i$, $B_i \cap B_j = \emptyset$ für $i \neq j$). Dann gilt für beliebiges A*

$$\mathbb{P}(A) = \sum_{i:\mathbb{P}(B_i)>0} \mathbb{P}(A|B_i)\mathbb{P}(B_i) = \sum_{i:\mathbb{P}(B_i)>0} \mathbb{P}(A \cap B_i).$$

Theorem 2.19 *Für beliebige Ereignisse $A_1, \dots, A_n$ gilt*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

Bayes-Regel

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

Theorem 2.20 (Bayes) *Ist $(B_i)_{i \in I}$ eine Zerlegung von Ω in disjunkte Ereignisse, $\mathbb{P}(B_i) > 0$ und $\mathbb{P}(A) \neq 0$, so ist*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_j \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.$$

2.5 Unabhängigkeit

Definition 2.21 (Unabhängigkeit) *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Eine Kollektion von Ereignissen $(A_i)_{i \in I}$ heisst (stochastisch) **unabhängig**, wenn gilt*

$$J \subseteq I \text{ endlich} \implies \mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i).$$

Definition 2.22 (Unabhängigkeit von Zufallsvariablen) *Eine Kollektion von diskreten Zufallsvariablen $(X_i)_{i \in I}$ heisst **unabhängig**, falls die Ereignisse $(\{X_i = x_i\})_{i \in I}$ unabhängig sind für jede Wahl von x_i aus dem Wertebereich von X_i.*

Lemma 2.23 *Wenn die diskreten Zufallsvariablen $X_1, \dots, X_n$ unabhängig sind, dann gilt*

$$\mathbb{E}\left(\prod_{i=1}^n g_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}(g_i(X_i))$$

für beliebige Funktionen $g_i : \mathbb{R}^n \to \mathbb{R}$, sofern die Erwartungswerte existieren.

Theorem 2.24 (de Moivre-Laplace) *Es gilt*

$$p_n(k) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)(1 + r_n(k)),$$
$$\sup\{|r_n(k)| : |k - np| \leq A\sqrt{n}\} \to 0 \text{ für } n \to \infty, \quad \forall A > 0.$$

3 Stetige Modelle

3.1 Allgemeine Wahrscheinlichkeitsräume

3.1.1 Die Axiome von Kolmogorov

Definition 3.1 *Das Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ heisst **Wahrscheinlichkeitsraum**, wenn*

- *$\mathcal{A}$ eine **σ-Algebra**,*

$$\Omega \in \mathcal{A},$$
$$A \in \mathcal{A} \implies A^c \in \mathcal{A},$$
$$A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_i A_i \in \mathcal{A},$$

- *P eine **Wahrscheinlichkeitsverteilung** ist,*

$$\mathbb{P}(\Omega) = 1,$$
$$A_1, A_2, \dots \in \mathcal{A}, A_i \cap A_j = \emptyset, i \neq j \implies \mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i).$$

3.1.2 Folgerungen

Theorem 3.2 (Erzeugte σ-Algebra) *Die von $\mathcal{A}_0$ erzeugte σ-Algebra $\mathcal{A} = \sigma(\mathcal{A}_0) = \bigcap_{\mathcal{B} \supseteq \mathcal{A}_0} \mathcal{B}$, wobei $\mathcal{B}$ σ-Algebren sind, ist die kleinste $\mathcal{A}_0$ enthaltende σ-Algebra.*

Theorem 3.3 *Ist $\mathbb{P} : \mathcal{A} \to [0, 1]$ additiv, so sind die folgenden Aussagen äquivalent*

(i) $\mathbb{P}$ *ist σ-additiv,*

(ii) $A_1 \subseteq A_2 \subseteq \dots \in \mathcal{A} \implies \mathbb{P}(\bigcup_n A_n) = \lim_n \mathbb{P}(A_n)$,

(iii) $A_1 \supseteq A_2 \supseteq \dots \in \mathcal{A} \implies \mathbb{P}(\bigcap_n A_n) = \lim_n \mathbb{P}(A_n)$,

Korollar 3.4 *Für beliebige $A_1, A_2, \dots \in \mathcal{A}$ gilt*

$$\mathbb{P}\left(\bigcup_k A_k\right) \leq \sum_k \mathbb{P}(A_k)$$

Lemma 3.5 (Borell-Cantelli) *Sei $A_1, A_2, \dots \in \mathcal{A}$ eine Folge von Ereignissen und sei*

$$A_\infty = \bigcap_n \left(\bigcup_{k \geq n} A_k\right) = \text{'unendlich viele der } A_k \text{ treten ein'}.$$

(i) Aus $\sum_k \mathbb{P}(A_k) < \infty$ folgt $\mathbb{P}(A_\infty) = 0$

(ii) Sind die Ereignisse $A_1, A_2, \dots$ unabhängig, so folgt aus $\sum_k \mathbb{P}(A_k) = \infty$, dass $\mathbb{P}(A_\infty) = 1$.

3.1.3 Sukzessive unabhängige 0-1-Experimente

Theorem 3.6 *Es gibt genau eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\Omega, \mathcal{A})$ derart, dass gilt*

$$\mathbb{P}(X_i = 1) = p, \quad i = 1, 2, \dots$$

Die Ereignisse $\{X_i = 1\}$, $i = 1, 2, \dots$ sind unabhängig bezüglich $\mathbb{P}$.

Theorem 3.7 *Sei $[x_1, \dots, x_N]$ ein "binärer Text"mit $x_i \in \{0, 1\}$. Dann ist die Wahrscheinlichkeit, dass irgendwann dieser Text erscheint (nicht nur einmal, sondern unendlich oft) gleich eins.*

3.1.4 Transformation von Wahrscheinlichkeitsräumen

Definition 3.8 (Messbare Abbildung) *Eine Abbildung $\phi : \Omega \to \hat{\Omega}$ heisst **messbar** (bezüglich $\mathcal{A}$ und $\hat{\mathcal{A}}$), wenn*

$$A \in \hat{\mathcal{A}} \implies \phi^{-1}(A) = \{\omega | \phi(\omega) \in \mathcal{A}\} \in \mathcal{A}.$$

Theorem 3.9 *Ist $\phi : \Omega \to \tilde{\Omega}$ messbar, so ist durch*

$$\tilde{\mathbb{P}}(A) \equiv \mathbb{P}(\phi^{-1}(A)) \quad (A \in \tilde{\mathcal{A}})$$

*eine Wahrscheinlichkeitsverteilung $\tilde{\mathbb{P}}$ auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ definiert. $\tilde{\mathbb{P}}$ heisst das **Bild** von $\mathbb{P}$ unter ϕ bzw. die **Verteilung** von $\mathbb{P}$ unter ϕ.*

3.2 Zufallsvariable und ihre Verteilung

Theorem 3.10 (Verteilung) *Ist $\phi : \Omega \to \hat{\Omega}$ messbar, so ist durch*

$$\hat{\mathbb{P}}(A) \equiv \mathbb{P}(\phi^{-1}(A)), \quad A \in \hat{\mathcal{A}}$$

eine Wahrscheinlichkeitsverteilung $\hat{p}$ auf $(\hat{\Omega}, \hat{\mathcal{A}})$ definiert. $\hat{p}$ heisst das Bild von $\mathbb{P}$ unter ϕ bzw. die Verteilung von ϕ unter $\mathbb{P}$.

Definition 3.11 (Zufallsvariable, Verteilung) *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine **Zufallsvariable** ist eine messbare Abbildung*

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

*Die **Verteilung** μ von X ist das Bild von $\mathbb{P}$ unter X, d.h. für jedes $A \in \mathcal{B}$*

$$\mu(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega | X(\omega) \in A\}) = \mathbb{P}(X \in A).$$

3.2.1 Verteilungsfunktion

Definition 3.12 (Verteilungsfunktion) *ist die Funktion von X bzw. μ*

$$F(b) = \mathbb{P}(X \leq b) = \mu((-\infty, b]), \quad b \in \mathbb{R}.$$

Definition 3.13 (t-Quantil/ Verallgemeinerte Umkehrabbildung) *ist definiert als*

$$F^{-1}(t) = \inf\{x | F(x) \geq t\}.$$

Theorem 3.14 (Eigenschaften einer Verteilungsfunktion)

(i) Monotonie: $a \leq b \implies F(a) \leq F(b)$,

(ii) Rechtsstetigkeit: $F(a) = \lim_{h \searrow 0} F(a+h)$,

(iii) Normiertheit: $lim_{a \to -\infty} F(a) = 0$ und $lim_{a \to +\infty} F(a) = 1$.

Lemma 3.15 (Eigenschaften von F^{-1}) *Wenn F i) - iii) von 3.14 erfüllt und F^{-1} wie in 3.13 definiert ist, dann ist F^{-1} monoton wachsend, linksstetig und es gilt*

(i) $F^{-1}(F(x)) \leq x, \quad -\infty < x < \infty$,

(ii) $t \leq F(F^{-1}(t)), \quad 0 < t < 1$.

3.2.2 Typen von Verteilungen

Definition 3.16 (Diskrete Zufallsvariable) *Eine Zufallsvariable X heisst **diskret**, wenn es eine abzählbare Menge $A \subset \mathbb{R}$ gibt, sodass $\mathbb{P}(X \in A) = 1$.*

Definition 3.17 (Absolute Stetigkeit, Dichte) *Eine Zufallsvariable heisst **absolut stetig**, falls eine messbare Funktion $f : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ existiert mit $f(x) \geq 0$ und $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$, sodass*

$$F(b) = \int_{-\infty}^{b} f(x)\,\mathrm{d}x.$$

*Die Funktion f heisst die **Dichte** von X.*

3.2.3 Transformation von Zufallsvariablen

Theorem 3.18 (Transformation von Zufallsvariablen) *Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ messbar. Dann ist*

$$Y(\omega) = g(X(\omega))$$

eine Zufallsvariable mit der Verteilungsfunktion

$$F_Y(b) = \mathbb{P}(g(X) \leq b) = \mathbb{P}(X \in g^{-1}((-\infty, b])).$$

Theorem 3.19 (Transformation von Dichten) *Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $Y(\omega) = g(X(\omega))$ mit $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ messbar, differenzierbar, monoton wachsend und $g'(x) > 0$ für alle x. Dann ist $F_Y(b) = F_X(g^{-1}(b))$. Falls die Dichte f_X existiert, dann existiert auch f_Y und ist gegeben durch*

$$f_Y(x) = \frac{1}{g'(g^{-1}(x))} f_X(g^{-1}(x)).$$

3.3 Erwartungswert

Definition 3.20 (Erwartungswert) *Für $X \geq 0$ ist der **Erwartungswert** definiert als*

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} x\mu(\mathrm{d}x) \in [0, \infty],$$

sofern dieser existiert.

Theorem 3.21 (Eigenschaften des Erwartungswertes)

$$\mathbb{E}(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 \mathbb{E}(X_1) + \alpha_2 \mathbb{E}(X_2) \qquad \text{(Linearität)}$$
$$X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y) \qquad \text{(Monotonie)}$$
$$0 \leq X_1 \leq X_2 \leq \ldots \implies \mathbb{E}(\lim_n X_n) = \lim_n \mathbb{E}(X_n) \qquad \text{(Monotone Stetigkeit)}$$

Theorem 3.22 (Konvergenzsatz von Lebesgue) *Sei $X_1, X_2, \ldots$ eine f.s. konvergente Folge von Zufallsvariablen. Wenn $|X_n(\omega)| \leq X(\omega)$ für alle n und $\mathbb{E}(X) < \infty$, dann*

$$\mathbb{E}(\lim_n X_n) = \lim_n \mathbb{E}(X_n).$$

Erwartungswert unter Transformation Sei $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ messbar und $Y(\omega) = g(X(\omega))$. Dann ist

$$\mathbb{E}(Y) = \int g(x)\mu_X(\mathrm{d}x).$$

Definition 3.23 (Moment)

***p-tes Moment**: $\mathbb{E}(X^p)$, $p \in \mathbb{N}$,*

***p-tes absolutes Moment**: $\mathbb{E}(|X|^p)$, $p > 0$,*

***p-tes zentriertes Moment**: $\mathbb{E}((X - \mathbb{E}(X))^p)$, $p \in \mathbb{N}$.*

Definition 3.24 (Varianz, Standardabweichung) *Das zweite zentrierte Moment heisst **Varianz** von X*

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

*Die **Standardabweichung** ist definiert als*

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

Theorem 3.25 (Eigenschaften von Varianz und Standardabw.)

$$\mathbb{V}(aX + b) = a^2 \mathbb{V}(X),$$
$$\sigma(aX + b) = |a|\sigma(X).$$

Theorem 3.26 (Ungleichung von Jensen) *Für eine Zufallsvariable X mit endlichem, existentem Erwartungswert und $g : \mathbb{R} \to \mathbb{R}$ konvex gilt*

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

Theorem 3.27 (Chebychev-Ungleichung) *Sei g eine nichtnegative, monoton wachsende Funktion auf $\mathbb{R}$. Dann gilt für jedes c mit $g(c) > 0$*

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(g(X))}{g(c)}.$$

3.4 Mehrere Zufallsvariablen

Definition 3.28 (Zufallsvektor) *Seien $X_1, X_2, \ldots, X_n$ Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Ein Zufallsvektor ist definiert als*

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n).$$

Definition 3.29 (Unabhängigkeit) *$X_1, \ldots, X_n$ heissen (stochastisch) **unabhängig**, falls für alle $A_1, \ldots, A_n \in \mathcal{B}$ gilt*

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \ldots \mathbb{P}(X_n \in A_n)$$

bzw.

$$\mu(A_1 \times \ldots \times A_n) = \mu_1(A_1) \ldots \mu_n(A_n).$$

Theorem 3.30 *Seien $X_1, \dots, X_n$ unabhängig. Dann ist μ absolut stetig genau dann, wenn jedes μ_i absolut stetig ist. Ferner gilt $f(\boldsymbol{x}) = \prod_{i=1}^n f_i(x_i)$.*

Theorem 3.31 (Transformation von Zufallsvariablen) *Sei $\boldsymbol{X}$ ein n-dimensionaler Zufallsvektor und $g : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^m, \mathcal{B}^m)$ messbar. Dann ist*

$$\boldsymbol{Y}(\omega) = g(\boldsymbol{X}(\omega))$$

ein m-dimensionaler Zufallsvektor. Ferner gilt

$$\mu_{\boldsymbol{Y}}(A) = \mu_{\boldsymbol{X}}(g^{-1}(A)).$$

Theorem 3.32 *Sei $g : \mathbb{R}^n \to \mathbb{R}^n$ linear und umkehrbar, d.h. $g(\boldsymbol{x}) = \boldsymbol{m} + B\boldsymbol{x}$ mit $\det(B) \neq 0$. Wenn $\mu_{\boldsymbol{X}}$ absolut stetig ist, dann ist auch $\mu_{\boldsymbol{Y}}$ absolut stetig und es gilt*

$$f_{\boldsymbol{Y}}(\boldsymbol{x}) = \frac{1}{|\det(B)|} f_{\boldsymbol{X}}(B^{-1}(\boldsymbol{x} - \boldsymbol{m})).$$

Verteilung von $X_1 + X_2$ Sei $\mathbf{X} = (X_1, X_2)$ eine absolut stetige Zufallsvariable. Dann ist $Y = X_1 + X_2$ absolut stetig mit Dichte $f_Y(y) = \int_{\mathbb{R}} f_{\mathbf{X}}(x_1, y - x_1)\,\mathrm{d}x_1$ Wenn X_1, X_2 unabhängig sind, gilt

$$f_Y(y) = \int_{\mathbb{R}} f_1(x_1) f_2(y - x_1)\,\mathrm{d}x_1 = (f_1 * f_2)(y).$$

Definition 3.33 (Kovarianz) *von X_1 und X_2 ist definiert als*

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))).$$

Theorem 3.34 (Eigenschaften der Kovarianz)

(i) $\mathrm{Cov}(X, X) = \mathbb{V}(X)$

(ii) $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$

(iii) $\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$

(iv) $\mathrm{Cov}(X_1, aX_2 + b) = a\mathrm{Cov}(X_1, X_2)$

(v) $\mathrm{Cov}(X_1, X_2 + X_3) = \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, X_3)$

(vi) $\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + 2\mathrm{Cov}(X_1, X_2)$

(vii) $|cov(X_1, X_2)| \leq \sigma(X_1)\sigma(X_2)$ *(Cauchy-Schwarz)*

Theorem 3.35 (Kovarianz und Unabhängigkeit) *Wenn X_1, X_2 unabhängig sind, ist $\mathrm{Cov}(X_1, X_2) = 0$ und $\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2)$.*

Definition 3.36 (Korrelation) *von X_1 und X_2 ist definiert als*

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}.$$

*Wenn $\rho(X_1, X_2) = 0$ ($\iff \mathrm{Cov}(X_1, X_2) = 0$), dann heissen X_1 und X_2 **unkorreliert**.*

Theorem 3.37 (Eigenschaften der Korrelation)

(i) $\rho(aX_1 + b, cX_2 + d) = \rho(X_1, X_2)$ *für $a > 0$, $c > 0$,*

(ii) $-1 \leq \rho(X_1, X_2) \leq +1$

4 Grenzwertsätze

Seien X_i beliebige Zufallsvariablen, wobei jedoch $\mathbb{E}(X_i) = m$ für alle i. Sei zudem $S_n = \sum_{i=1}^n X_i$.

4.1 Schwaches Gesetz der grossen Zahlen

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \varepsilon\right) \to 0 \text{ für } n \to \infty \qquad \text{(Schwaches LLN)}$$

4.2 Starkes Gesetz der grossen Zahlen (LLN)

$$\lim_n \mathbb{P}\left(\bigcap_{k \geq n} \left\{\left|\frac{S_k}{k} - m\right| \leq \varepsilon\right\}\right) = 1 \iff \qquad \text{(Starkes LLN)}$$

$$\forall \varepsilon > 0 : \mathbb{P}\left(\bigcup_n \bigcap_{k \geq n} \left\{\left|\frac{S_k}{k} - m\right| \leq \varepsilon\right\}\right) = 1 \iff$$

$$\mathbb{P}\left(\bigcap_{\varepsilon > 0} \bigcup_n \bigcap_{k \geq n} \left\{\left|\frac{S_k}{k} - m\right| \leq \varepsilon\right\}\right) = \mathbb{P}\left(\lim_{n \to \infty} \frac{S_n}{n} = m\right) = 1$$

Definition 4.1 (Konvergenzen) *Seien $Z, Z_1, Z_2, \dots$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Wir definieren*

- ***stochastische Konvergenz*** *oder Konvergenz in Wahrscheinlichkeit von Z_n gegen Z als*

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} \mathbb{P}(|Z_n - Z| > \varepsilon) = 0,$$

- ***fast-sichere Konvergenz*** *von Z_n gegen Z als*

$$\mathbb{P}(\{\omega | \lim_n Z_n(\omega) = Z(\omega)\}) = 1.$$

Theorem 4.2 (Beziehung der Konvergenzen) *Fast-sichere Konvergenz impliziert stochastische Konvergenz. Wenn für alle $\varepsilon > 0$ gilt $\sum_n \mathbb{P}(|Z_n - Z| > \varepsilon) < \infty$, dann konvergiert Z_N fast sicher gegen Z.*

Korollar 4.3 *Wenn (Z_n) stochastisch gegen Z konvergiert, dann existiert eine Teilfolge (Z_{n_j}), welche fast sicher gegen Z konvergiert.*

Theorem 4.4 *Sei (X_i) i.i.d. mit $\mathbb{E}(X_i^2) < \infty$. Dann konvergiert $\frac{S_n}{n}$ fast sicher gegen $m = \mathbb{E}(X_i)$.*

Theorem 4.5 (Gesetz vom iterierten Logarithmus) *Sei (X_i) eine i.i.d. Folge von Zufallsvariablen, $\mathbb{E}(X_i) = m$, $\mathbb{V}(X_i) = \sigma^2 < \infty$. Dann gilt mit Wahrscheinlichkeit 1*

$$\mathbb{P}\left(\limsup_{n \to \infty} \frac{S_n - nm}{\sqrt{2\sigma^2 n \log(\log(n))}} = 1\right) = 1$$

$$\mathbb{P}\left(\liminf_{n \to \infty} \frac{S_n - nm}{\sqrt{2\sigma^2 n \log(\log(n))}} = -1\right) = 1$$

4.3 Zentraler Grenzwertsatz (CLT)

Definition 4.6 (Schwache Konvergenz) *Seien μ und μ_n Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathcal{B})$. Wir sagen, dass μ_n **schwach** gegen μ **konvergiert**, falls*

$$\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$$

für alle stetigen und beschränkten f.

Lemma 4.7 *Seien μ und μ_n Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathcal{B})$ mit Verteilungsfunktionen F und F_n. Dann sind folgende Aussagen äquivalent.*

(i) $\mu_n \to \mu$ *schwach,*

(ii) $F_n(x) \to F(x)$ *für jede Stetigkeitsstelle x von F,*

(iii) $\int f d\mu_n \to \int f d\mu$ *für alle $f \in C_b^3(R)$, wobei $C_b^3(R)$ die Menge alle dreimal stetig differenzierbaren Funktionen auf $\mathbb{R}$ bezeichnet, für die f, f', f'', f''' alle beschränkt sind.*

Lemma 4.8 *Wenn X_1, X_2 unabhängig sind und $X \sim \mathcal{N}(m_i, \sigma_i^2)$ ($i = 1,2$), dann ist $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Theorem 4.9 (Zentraler Grenzwertsatz) *Sei (X_i) i.i.d. mit $\mathbb{E}(X_i) = m$ und $\mathbb{V}(X_i) = \sigma^2 < \infty$. Dann konvergiert die Verteilung von S_n^* schwach gegen $\mathcal{N}(0,1)$, d.h.*

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}.$$

Theorem 4.10 (Zentraler Grenzwertsatz nach Lindeberg) *Seien $X_{n,i}$ ($1 \leq i \leq n, n \in \mathbb{N}$) Zufallsvariablen mit*

(i) $X_{n,1}, \dots, X_{n,n}$ *sind unabhängig für alle n,*

(ii) $\mathbb{E}(X_{n,i}) = 0$, $\mathbb{E}(X_{n,i}^2) = \sigma_{n,i}^2$, $\sum_{i=1}^n \sigma_{n,i}^2 = 1$

(iii) $\lim_{n \to \infty} \sum_{i=1}^n \mathbb{E}(X_{n,i}^2 \mathbb{1}_{[|X_{n,i}| > \varepsilon]})\ \forall \varepsilon > 0.$ *(Lindeberg-Bedingung)*

Dann konvergiert die Verteilung von $S_n = X_{n,1} + \dots X_{n,n}$ schwach gegen $\mathcal{N}(0,1)$.

Korollar 4.11 (Monte-Carlo-Integration) *Sei (X_i) eine i.i.d. Folge von d-dimensionalen Zufallsvariablen mit Verteilung μ und sei $f \in L^2(\mathbb{R}^d, \mu)$. Dann gilt*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n f(X_i) = \int_{\mathbb{R}^d} f(x)\mu(\mathrm{d}x) = \mathbb{E}(f(X_i)) = m \qquad \mathbb{P}-f.s.,$$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\frac{1}{n}\sum_{i=1}^n f(X_i) - m}{\sigma_f/\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R},$$

wobei $\sigma_f^2 = \mathbb{V}(f(X_i)) = \int_{\mathbb{R}^d}(f(x)-m)^2\mu(\mathrm{d}x)$. Das heisst der Fehler

$$\frac{1}{n}\sum_{i=1}^n f(X_i) - m$$

ist approximativ normalverteilt mit Mittel 0 *und Standardabweichung $\sigma_f/\sqrt{n}$.*

5 Charakteristische Funktionen

Definition 5.1 (Charakteristische Funktion) *$\phi_X : \mathbb{R}^n \to \mathbb{C}$ eines Zufallverktors $X : \Omega \to \mathbb{R}^n$ ist durch*

$$\phi_X(u) = \mathbb{E}(e^{i\langle u,X\rangle}) = \int_{\mathbb{R}^n} e^{i\langle u,x\rangle}\mu(\mathrm{d}x), \quad u \in \mathbb{R}^n$$

definiert, wobei μ die Verteilung von X bezeichnet.

Lemma 5.2 (Eigenschaften der charakteristischen Funktion) *Es sei $X : \Omega \to \mathbb{R}^n$ eine Zufallsvariable.*

(i) Die charakteristische Funktion $\phi_X : \mathbb{R}^n \to \mathbb{C}$ ist stetig mit $\phi_X(0) = 1$ und $|\phi_X(u)| \leq 1$ für alle $u \in \mathbb{R}^n$.

(ii) Es seien $A \in \mathbb{R}^{n\times n}$ und $b \in \mathbb{R}^m$. Dann gilt für die charakteristische Funktion der neuen $\mathbb{R}^m$-wertigen Zufallsvariablen $AX + b$

$$\phi_{AX+b}(u) = \phi_X(A^T u) \cdot e^{i\langle u,b\rangle}, \quad u \in \mathbb{R}^m.$$

(iii) Zwei Zufallsvariablen X, Y sind genau dann unabhängig, wenn

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u), \quad u \in \mathbb{R}^n.$$

(iv) Die charakteristische Funktion ϕ_X ist genau dann reellwertig, wenn die Verteilung von X symmetrisch ist.

Theorem 5.3 (Momente aus der charakteristischen Funktion) *Es sei $X : \Omega \to \mathbb{R}^n$ eine Zufallsvariable mit $\mathbb{E}(\|X\|^m) < \infty$ für ein $m \in \mathbb{N}$. Dann ist $\phi_X : \mathbb{R}^n \to \mathbb{C}$ eine $\mathbb{C}^m$-Funktion und für alle $m_1, \ldots, m_n \in \mathbb{N}_0$ mit $m_1 + \ldots + m_n = m$ gilt*

$$\frac{\partial^m}{\partial x_1^{m_1}\ldots\partial x_n^{m_n}}\phi_X(u) = i^m\mathbb{E}(X_1^{m_1}\ldots X_n^{m_n}e^{i\langle u,X\rangle}), \quad u \in \mathbb{R}^n.$$

Theorem 5.4 (Eindeutigkeitssatz) *Für zwei Zufallsvariablen X_1, X_2 mit Verteilungen μ_1, μ_2 auf $\mathbb{R}^n$ und $\phi_{X_1} = \phi_{X_2}$ gilt $\mu_1 = \mu_2$.*

Korollar 5.5 *Es sei $X = (X_1, \ldots, X_n)$ eine $\mathbb{R}^n$-wertige Zufallsvariable. Dann sind die reellwertigen Zufallsvariablen $X_1, \ldots, X_n$ genau dann unabhängig, wenn für alle $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ gilt*

$$\phi_X(u_1, \ldots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j).$$

Theorem 5.6 (Bochner) *Eine Funktion $\phi : \mathbb{R}^n \to \mathbb{C}$ ist die charakteristische Funktion einer Zufallsvariablen X genau dann, wenn $\phi(0) = 1$ gilt, die Funktion in $u = 0$ stetig ist und nichtnegativ definit ist, d.h. für alle $m \in \mathbb{N}$ und alle $u_1, \ldots, u_m \in \mathbb{R}^n$ und $z_1, \ldots, z_m \in \mathbb{C}$ gilt*

$$\sum_{j=1}^m\sum_{k=1}^m \phi(u_j - u_k)z_j\bar{z}_k \geq 0.$$

Korollar 5.7 (Verteilung unter Summation) *Seien $X_1, ..., X_n$ unabhängig. Dann gilt:*

- *X_i Bernoulli-verteilt mit Parameter $p \Rightarrow \sum_{i=1}^n X_i$ binomialverteilt mit Parametern (n, p)*
- *X_i Poisson-verteilt mit Parametern $\lambda_i > 0 \Rightarrow \sum_{i=1}^n X_i$ Poisson-verteilt mit Parameter $\sum_{i=1}^n \lambda_i$*
- *X_i binomialverteilt mit Parametern $(n_i, p) \Rightarrow \sum_{i=1}^n X_i$ binomialverteilt mit Parametern $(\sum_i n_i, p)$*
- *X_i normalverteilt mit Parametern $(\mu_i, \tau_i) \Rightarrow \sum_{i=1}^n X_i$ normalverteilt mit Parametern $(\sum_i \mu_i, \sum_i \tau_i)$*
- *X_i Gamma-verteilt mit Parametern $(\alpha_i, \lambda) \Rightarrow \sum_{i=1}^n X_i$ Gamma-verteilt mit Parametern $(\sum_i \alpha_i, \lambda)$*
- *X, Y Cauchy-verteilt $\Rightarrow \frac{X+Y}{2}$ Cauchy-verteilt*

Theorem 5.8 (Levys Stetigkeitssatz) *Sei $(\mu)_n$ eine Folge von Verteilungen reellwertiger Zufallsvariablen (X_n) und (ϕ_{X_n}) bezeichne deren charakteristische Funktionen. Dann gilt*

(i) Falls μ_n schwach gegen eine Verteilung μ einer Zufallsvariable X konvergiert, dann konvergieren die charakteristischen Funktionen punktweise, d.h. $\phi_{X_n}(u) \to \phi_X(u)$ punktweise für alle $u \in \mathbb{R}^n$.

(ii) Falls $\phi_{X_n}(u)$ punktweise gegen eine Funktion $g(u)$ konvergiert und diese zusätzlich stetig in $u = 0$ ist, dann ist g die charakteristische Funktion einer Zufallsvariablen $X : g(u) = \phi_X(u)$ und μ_n konvergiert schwach gegen die Verteilung μ von X.

6 Einführung in die Statistik

Idee der Statistik Auswählen eines Modells oder Schätzen der Parameter eines Modells $(\Omega, \mathcal{F}, \mathbb{P})$ mit ZV $X : \Omega \to \mathbb{R}^n$, sodass Beobachtungen $x \in \mathbb{R}^n$ optimal erklärt werden.

6.1 Klassische Statistik

Problemstellungen

- Punktschätzung eines unbekannten Parameters
- statistische Tests, um zu prüfen, ob gegebene Parameter zu den Daten passen
- Angabe von Konfidenzintervallen, um die Lage eines Parameters auf den wahrscheinlichsten Bereich einzugrenzen

Definition 6.1 (Schätzer) *Eine Funktion $T : \mathbb{R}^n \to \Theta$, die Beobachtungen auf Parameter abbildet, heisst Schätzer. Der Parameter von Interesse wird mit $\eta = g(\theta)$ bezeichnet, wobei $g : \Theta \to \mathbb{R}$ meist eine Projektion ist. Der Parameterschätzer ist definiert als $U = g(T) = g(T(\mathbf{X}))$.*

Definition 6.2 (Fehler) *Der **Mittlere Quadratische Fehler** ist definiert als*

$$\mathbb{E}_\theta(|U - g(\theta)|^2),$$

*der **systematische Schätzfehler/ Bias** als*

$$b_U(\theta) = \mathbb{E}_\theta(U) - g(\theta),$$

*der **Standardfehler** als*

$$\sigma_U(\theta) = \sqrt{\mathbb{V}_\theta(U)}.$$

Theorem 6.3 (MSE-Zerlegung)

$$\mathbb{E}_\theta(|U - g(\theta)|^2) = \sigma_U^2(\theta) + b_U^2(\theta).$$

Definition 6.4 (Erwartungstreue) *U heisst erwartungstreu für $g(\theta$, falls der Bias verschwindet, also*

$$\mathbb{E}_\theta(U) = g(\theta), \quad \forall\theta \in \Theta.$$

Definition 6.5 (Konsistenz) *(U_n) heisst konsistent für $g(\theta)$, falls*

$$\mathbb{P}_\theta(|U_n - g(\theta)| > \varepsilon) \overset{n\to\infty}{\to} 0 \quad \forall\varepsilon > 0, \forall\theta.$$

Definition 6.6 (Asymptotische Normalverteiltheit) *(U_n) heisst asymptotisch normalverteilt mit asymptotischer Varianz $\tau^2(\theta)$ und wir schreiben $U_n \approx\sim \mathcal{N}(g(\theta), \frac{1}{n}\tau^2(\theta))$, falls für alle θ*

$$\mathbb{P}_\theta(\sqrt{n}(U_n - g(\theta)) \leq x) \overset{n\to\infty}{\to} \Phi\left(\frac{x}{\tau(\theta)}\right) \quad \forall x.$$

Definition 6.7 (Bruchpunkt) *ist definiert als*

$$\varepsilon^*(x_1, \ldots, x_n) = \frac{1}{n}\max\{k \in \mathbb{N}_0 | \\ \sup\{|U(y_1, \ldots, y_n)| \,;\, \#\{y_i \neq x_i\} = k\} < \infty\}.$$

6.1.1 Maximum-Likelihood-Methode (ML)

Definition 6.8 (Likelihood-Funktion)

$$L : \Theta \to \mathbb{R} : \theta \mapsto L_\theta(x_1, \dots, x_n) = \begin{cases} \mu_\theta(x_1, \dots, x_n) & \textit{diskret} \\ f_\theta(x_1, \dots, x_n) & \textit{absolut stetig} \end{cases}$$

für feste Beobachtungen $(x_1, \dots, x_n)$ und diskrete Verteilung μ_θ bzw. Dichte f_θ.

Definition 6.9 (Maximum-Likelihood-Funktion) *ist gegeben durch die messbare Abbildung T*

$$T : \mathbb{R}^n \to \Theta : (x_1, \dots, x_n) \mapsto \operatorname*{argmax}_\theta L_\theta(x_1, \dots, x_n).$$

6.1.2 Bayesianische Statistik

Definition 6.10 (Bayes-Schätzer) *gegeben Daten $(x_1, \dots, x_n)$, eines Priors π_0, Likelihood $f(x_1, \dots, x_n|\theta)$ mit $\theta \in \Theta$ ist definiert als*

$$\pi_1(\mathrm{d}\theta|x) = \frac{f(x_1, \dots, x_n|\theta)\pi_0(\mathrm{d}\theta)}{\int_\Theta f(x_1, \dots, x_n|\theta)\pi_0(\mathrm{d}\theta)}.$$

6.2 Vertrauensintervalle

Definition 6.11 (Vertrauensintervalle) *Seien $\underline{T}, \overline{T} : \mathbb{R}^n \to \mathbb{R}$ messbare Funktionen mit $\underline{T} < \overline{T}$. Dann heisst $(\underline{T}(\boldsymbol{X}), \overline{T}(\boldsymbol{X}))$ ein **Vertrauensintervall** für $g(\theta)$ zum Niveau $1 - \alpha$, falls*

$$\forall \theta \in \Theta : \mathbb{P}_\theta(\underline{T}(\boldsymbol{X}) < g(\theta) < \overline{T}(\boldsymbol{X})) > 1 - \alpha.$$

Theorem 6.12 (Dualitätssatz) *Sei C eine messbare Teilmenge von $\mathbb{R}^n \times \mathbb{R}$ mit den messbaren Schnitten $A(\gamma) = \{\boldsymbol{x} \in \mathbb{R}^n | (\boldsymbol{x}, \gamma) \in C\}$ und $B(\boldsymbol{x}) = \{\gamma \in \mathbb{R} | (\boldsymbol{x}, \gamma) \in C\}$. Dann sind die beiden folgenden Aussagen äquivalent*

(i) Für jedes γ ist $\varphi(\boldsymbol{x}) = \mathbb{1}_{A(\gamma)^c}(\boldsymbol{x})$ ein Test der Nullhypothese $g(\theta) = \gamma$ zum Niveau α mit Verwerfungsbereich $A(\gamma)^c$.

(ii) $B(\boldsymbol{X})$ bildet einen Verwerfungsbereich für $g(\theta)$ zum Niveau $1 - \alpha$.

6.3 Statistische Tests

6.3.1 Das Neyman-Pearson-Lemma

Definition 6.13 (Randomisierter Test) *Ein randomisierter Test ist die messbare Funktion $\varphi : (\mathbb{R}^n, \mathcal{B}^n) \to [0, 1]$.*

Theorem 6.14 (Neyman-Pearson-Lemma) *Seien μ_0 und μ_1 Wahrscheinlichkeiten mit Dichten p_0 und p_1 bezüglich $\mu_0 + \mu_1$ gegeben durch die Radon-Nikodym-Ableitung*

$$\mu_i(A) = \int_A p_i(\boldsymbol{x})(\mu_0(\mathrm{d}x) + \mu_1(\mathrm{d}x)) = \begin{cases} \frac{\mu_i(\boldsymbol{x})}{\mu_0(\boldsymbol{x}) + \mu_1(\boldsymbol{x})} & \mu_i \textit{ diskret,} \\ \frac{f_i(\boldsymbol{x})}{f_0(\boldsymbol{x}) + f_1(\boldsymbol{x})} & \mu_i \textit{ absolut stetig.} \end{cases}$$

Sei $\alpha \in [0, 1]$ gegeben. Dann

(i) existiert ein randomisierter Test φ und ein $c \in [0, \infty]$ derart, dass

$$\mathbb{E}_0(\varphi) = \alpha, \tag{6.1}$$

$$\phi(x) = \begin{cases} 1 & p_1(x) > cp_0(x) \\ 0 & p_1(x) < cp_0(x) \end{cases}, \tag{6.2}$$

wobei $\infty \cdot 0$ als 0 definiert ist,

(ii) ist jeder Test, der 6.1 und 6.2 erfüllt, ein mächtigster Test zum Niveau α,

(iii) erfüllt jeder mächtigste Test zum Niveau α Bedingung 6.1 $(\mu_0 + \mu_1)$-überall. Er erfüllt auch 6.2, ausser wenn es einen Test ϕ' gibt mit $\mathbb{E}_1(\phi') = 1$ und $\mathbb{E}_0(\phi') < \alpha$.

Remark 6.15 (Likelihoodquotient) *ist $p_1(x)/p_0(x)$. Der Test aus 1. heisst Likelihoodquotiententest. Kurz gesagt ist der Likelihoodquotiententest optimal.*

6.3.2 Testtheorie

Setting Beobachtungen $\mathbf{X} = (X_1, \dots, X_n)$, mögliche Verteilungen $(u_\theta)_{\theta \in \Theta}$, Nullhypothese $\Theta_0 \subset \Theta$, Alternative Θ_0^c.

- **Fehler 1. Art**: Nullhypothese wird abgelehnt (verworfen), obwohl sie richtig ist. Die Wahrscheinlichkeit eines Fehlers 1. Art ist $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\varphi = 1)$.
- **Fehler 2. Art**: Nullhypothese wird akzeptiert (beibehalten), obwohl sie falsch ist. Die Wahrscheinlichkeit eines Fehlers 2. Art ist $\sup_{\theta \in \Theta_0^c} \mathbb{P}_\theta(\varphi = 0)$.

Definition 6.16 (Statistischer Test) *Ein statistischer Test ist die messbare Abbildung $\varphi : (\mathbb{R}^n, \mathcal{B}^n) \to \{0, 1\} : x \mapsto \mathbb{1}_K(x)$, wobei $\phi = 0$ bedeutet, dass die Nulllhypothese angenommen wird, und $\phi = 1$ bedeutet, dass die Nullhypothese abgelehnt wird. $K \subset \mathbb{R}^n$ bezeichnet den **Verwerfungsbereich** oder kritischen Bereich des Tests.*

Ziel eines Tests ist

- kleiner Fehler 1. Art: $\mathbb{P}_\theta(\varphi = 1) = \mathbb{E}_\theta(\phi)$ klein für $\theta \in \Theta_0$.
- kleiner Fehler 2. Art: $\mathbb{P}_\theta(\varphi = 0)$ klein, $\mathbb{E}_\theta(\phi)$ gross für $\theta \in \Theta_0^c$.

Definition 6.17 (Niveau, Macht) *Ein Test φ heisst zum **Niveau** α, wenn $\sup_{\theta \in \Theta_0} \mathbb{E}_\theta(\varphi) \leq \alpha$, d.h. die Wahrscheinlichkeit eines Fehlers 1. Art $\leq \alpha$. Für $\theta \notin \Theta_0$ heisst $\mathbb{E}_\theta(\phi)$ auch die **Macht des Tests** an der Stelle $\theta \notin \Theta_0$. Die Macht ist also Eins minus die Wahrscheinlichkeit eines Fehlers 2. Art.*

Definition 6.18 (Kompatible Tests) *sind Tests, für die gilt $\alpha' < \alpha \implies \varphi_{\alpha'} \leq \varphi_\alpha$.*

Definition 6.19 (P-Wert) *eines kompatiblen Tests φ zum Wert $\boldsymbol{x}$ ist die Zufallsvariable definiert als*

$$\pi(\boldsymbol{x}) = \inf\{\alpha | \varphi_\alpha(\boldsymbol{x}) = 1\}.$$

Lemma 6.20 (Uniforme Verteilung des P-Werts) *Wenn $\theta \in \Theta_0$, dann gilt $\mathbb{P}_\theta(\pi(\boldsymbol{X}) \leq u) \leq u$. Wenn $\mathbb{P}_\theta(\varphi_\alpha(\boldsymbol{X}) = 1) = \alpha$ für alle α, dann gilt $\mathbb{P}_\theta(\pi(\boldsymbol{X}) \leq u) = u$, d.h. der P-Wert ist uniform verteilt.*

Vorgehen bei klassischen Tests Klassische Tests haben eine Teststatistik T als Funktion der betrachteten Zufallsvariablen. Von dieser kennt man die Verteilung. Man verwirft die Nullhypothese, wenn $T < c$ bzw. $T > c$ (einseitiger Test) oder $|T| > c$ (zweiseitiger Test), wobei c meist als das α- (einseitig links), $(1 - \alpha)$- (einseitig rechts) oder $(1 - \alpha/2)$-Quantil (zweiseitig) der Verteilung von T gewählt wird.

(i) **Annahmen**: Liste Modellannahmen, zu testende Grösse und das Niveau α auf.

(ii) **Test**: Wähle einen geeigneten Test.

(iii) **Hypothesen**: Stelle Nullhypothese H_0 und Alternativhypothese H_1 auf. Der Test macht nur dann eine Aussage, wenn die Nullhypothese abgelehnt wird. Wähle deswegen die Nullhypothese als Negation der in dem Test gestellten Frage.

(iv) **Verteilung**: Identifiziere die Verteilung der Teststatistik unter der Annahme von H_0 mit den jeweiligen Parametern bzw. geeignete Näherungen für diese Verteilung.

(v) **Verwerfungsbereich** K aus den Quantilen der Verteilung berechnen. Meist kann man die Quantile aus Tabellen entnehmen. Alternativ kann man die geeignet genäherte Verteilung zum Berechnen der Quantilen verwenden, insbesondere für grosse Stichprobenzahl n (beispielsweise Nähern durch die Normalverteilung).

(vi) **Teststatistik** aus den gegebenen Daten berechnen.

(vii) **Testentscheid**: Annahme oder Ablehnung der Nullhypothese durch durch Vergleich von Teststatistik mit Verwerfungsbereich meist in der Form $\phi(x) = 1 \iff T \in K$. Alternativ bestimme den p-Wert, für $p < \alpha$ wird H_0 abgelehnt.

(viii) **Interpretation**: Wird die Nullhypothese abgelehnt, kann man aussagen, dass entweder die Nullhypothese falsch ist, oder ein seltenes Ereignis, dessen Wahrscheinlichkeit höchstens α ist, eingetreten ist. Wird die Nullhypothese angenommen, können wir keine Aussage treffen.

6.3.3 Binomialtest

Test auf	Wahrscheinlichkeit p
Daten	$X_1, \dots, X_n$ i.i.d. $\sim \mathrm{Ber}(p)$
Raum	$\Theta = [0,1]$
Nullhypothese	$\Theta_0 = \{p\} = \{p_0\}$ (zweiseitig) $\Theta_0 = [p_0, 1]$ (einseitig links) $\Theta_0 = [0, p_0]$ (einseitig rechts)
Teststatistik	$T = \sum_{i=1}^n \mathbb{1}_{\{1\}}(X_i)$
Verteilung	$F \sim \mathrm{Bin}(n,p)$
Test	$\phi(x) = 1 \Leftrightarrow T \in \begin{cases} (F^{-1}(\frac{\alpha}{2}), F^{-1}(1-\frac{\alpha}{2})) & \text{zweis.} \\ (F^{-1}(1-\alpha), 1] & \text{e. li.} \\ [0, F^{-1}(\alpha)) & \text{e. re.} \end{cases}$

Ab ca. $n \geq 30$ kann als Näherung die Standard-Normalverteilung $\mathcal{N}(0,1)$ verwendet werden.

Test auf	Wahrscheinlichkeit p
Daten	$X_1, \dots, X_n$ i.i.d. $\sim \mathrm{Ber}(p)$
Raum	$\Theta = [0,1]$
Teststatistik	$T = \sqrt{n}\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)}}$ mit $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$
Verteilung	$F \sim \mathcal{N}(0,1)$
Test	$\phi(x) = 1 \Leftrightarrow \lvert T\rvert > \begin{cases} F^{-1}(1-\frac{\alpha}{2}) & \text{zweiseitig} \\ F^{-1}(1-\alpha) & \text{einseitig} \end{cases}$

6.3.4 1-Stichproben-t-Test

Test auf	Mittelwert
Daten	$X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$
Raum	$\Theta = \mathbb{R} \times \mathbb{R}_+$
Nullhypothese	$\Theta_0 = \{\mu_0\} \times \mathbb{R}_+$ (zweiseitig) $\Theta_0 = [\mu_0, \infty) \times \mathbb{R}_+$ (einseitig links) $\Theta_0 = (-\infty, \mu_0] \times \mathbb{R}_+$ (einseitig rechts)
Teststatistik	$T = \frac{\sqrt{n}(\overline{X}-\mu_0)}{S_n}$, wobei $S_n^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$ und $\overline{X} = \frac{1}{n}\sum X_i$
Verteilung	$F \sim$ t-Verteilung mit $n-1$ Freiheitsgraden
Näherungen	Für grosse n kann die t-Verteilung mit der Standard-Normalverteilung $\mathcal{N}(0,1)$ genähert werden.
Test	$\phi(x) = 1 \Leftrightarrow \lvert T\rvert > \begin{cases} F^{-1}(1-\frac{\alpha}{2}) & \text{zweiseitig} \\ F^{-1}(1-\alpha) & \text{einseitig} \end{cases}$
Vertrauensint.	$[\overline{X} - \frac{S_N}{\sqrt{n}}F^{-1}(1-\frac{\alpha}{2}), \overline{X} + \frac{S_N}{\sqrt{n}}F^{-1}(1-\frac{\alpha}{2})]$ (zweis.) $[\overline{X} - \frac{S_N}{\sqrt{n}}F^{-1}(1-\alpha),\ \infty]$ (einseitig links)

Theorem 6.21 (t-Verteilung mit $n-1$ Freiheitsgraden) *Die Verteilung von T hat für $\mu = \mu_0$, $n \geq 2$ und alle $\sigma > 0$ die Dichte*

$$f_{n-1}(t) = \frac{\Gamma(n/2)}{\sqrt{(n-1)\pi}\,\Gamma(\frac{n-1}{2})}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}.$$

6.3.5 1-Stichproben-z-Test

Für bekannte Standardabweichung σ der X_i, ersetze im t-Test S_n durch σ und verwende die Standard-Normalverteilung $\mathcal{N}(0,1)$.

Test auf	Mittelwert
Daten	$X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$
Raum	$\Theta = \mathbb{R}$
Teststatistik	$T = \frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma}$, wobei $\overline{X} = \frac{1}{n}\sum X_i$
Verteilung	$F \sim \mathcal{N}(0,1)$
Test	$\phi(x) = 1 \Leftrightarrow \lvert T\rvert > \begin{cases} F^{-1}(1-\frac{\alpha}{2}) & \text{zweiseitig} \\ F^{-1}(1-\alpha) & \text{einseitig} \end{cases}$

6.3.6 Vorzeichentest

Alternative zum t-Test. Unbekannte Verteilung, dafür Testung auf Median statt Mittelwert.

Test auf	Median
Daten	$X_1, \dots, X_n$ i.i.d. $\sim F(x-\mu)$
Raum	$\Theta = \mathbb{R} \times \{F \mid F(0) = 1/2\}$
Nullhypothese	$\Theta_0 = \mu_0 \times \{F \mid F(0) = 1/2\}$
Teststatistik	$T = \sum_{i=1}^n \mathbb{1}_{[X_i > \mu_0]}$
Verteilung	$F \sim \mathrm{Binomial}(n, p = 1/2)$
Test	$\phi(x) = 1 \iff \lvert T - n/2\rvert > c = F^{-1}(1-\frac{\alpha}{2})$

6.3.7 2-Stichproben-t-Test

Test auf	Mittelwert
Daten	$X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(\mu_1, \sigma^2)$ $Y_1, \dots, Y_m$ i.i.d. $\sim \mathcal{N}(\mu_2, \sigma^2)$
Raum	$\theta = (\mu_1, \mu_2, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$
Nullhypothese	$\Theta_0 = \{(\mu_1, \mu_1, \sigma^2) \mid \mu_1 = \mu_2\} \times \mathbb{R}_+$ (zweiseitig) $\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) \mid \mu_1 \geq \mu_2\} \times \mathbb{R}_+$ (eins. links) $\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) \mid \mu_1 \leq \mu_2\} \times \mathbb{R}_+$ (eins. rechts)
Teststatistik	$T = \frac{\overline{X}-\overline{Y}}{\sqrt{1/n+1/m}}\sqrt{\frac{n+m-2}{\sum(X_i-\overline{X})^2+\sum(Y_i-\overline{Y})^2}}$ mit $\overline{X} = \frac{1}{n}\sum X_i$, $\overline{Y} = \frac{1}{n}\sum Y_i$
Verteilung	$F \sim$ t-Verteilung mit $n+m-2$ Freiheitsgraden
Test	$\phi(x) = 1 \Leftrightarrow \lvert T\rvert > \begin{cases} F^{-1}(1-\frac{\alpha}{2}) & \text{zweiseitig} \\ F^{-1}(1-\alpha) & \text{einseitig} \end{cases}$

Theorem 6.22 *Für alle $\theta \in \Theta_0$ hat T die t-Verteilung mit n+m-2 Freiheitsgraden.*

6.3.8 2-Stichproben-Wilcoxon-Test (Mann-Whitney U-Test)

Test auf	Mittelwerte und Mediane
Daten	$X_1, \dots, X_n$ i.i.d. $\sim F$ $Y_1, \dots, Y_n$ i.i.d. $\sim G$
Raum	$(F, G) \in \Theta = \mathcal{F} \times \mathcal{F}$, wobei $\mathcal{F}$ die Menge aller stetigen Verteilungsfunktionen bezeichnet
Nullhypothese	$\Theta_0 = \{F = G\}$
Teststatistik	$W = \sum_{i=1}^n \sum_{j=1}^n 1_{[X_i < Y_i]}$
Verteilung	$F \sim$ Wilcoxon-Verteilung mit Parameter n, n
Näherungen	Für grosse n kann die t-Verteilung mit der Standard-Normalverteilung $\mathcal{N}(0,1)$ genähert werden (siehe Lemma 6.23).
Test	$\phi(x) = 1 \iff \lvert T - n^2/2 \rvert > c = F^{-1}(1 - \frac{\alpha}{2})$

Lemma 6.23 *Unter der Nullhypothese $\Theta_0 = \{(F, F)\}$ gilt:*

$$\mathbb{E}(W) = \frac{n^2}{2}, \quad \mathbb{V}(W) = \frac{n^2(2n+1)}{12}$$

$$\mathbb{P}\left(\frac{W - \mathbb{E}(W)}{\sqrt{\mathbb{V}(W)}} \le x\right) \to \Phi(x) \quad (x \in \mathbb{R}; n \to \infty)$$

Diskussion Wilcoxon-Test ist besser als t-Test.

6.3.9 Chi-Quadrat-Anpassungstest

Betrachte n Wiederholungen eines Experiments mit k möglichen Ausgängen und bezeichne θ_i die Wahrscheinlichkeit für den i-ten Ausgang. Wir arbeiten mit der Zufallsvariable N_i = Anzahl Wiederholungen mit Ausgang i für $i = 1, \dots, k$. Dann folgt $(N_1, \dots, N_k)$ der Multinomialverteilung

$$\mathbb{P}(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{n_1! \dots n_k!} \theta_1^{n_1} \dots \theta_k^{n_k} \quad \text{(Multinomialvert.)}$$

mit $\mathbb{E}(N_i) = n\theta_i$, $\mathbb{V}(N_i) = n\theta_i(1 - \theta_i)$, $\mathrm{Cov}(N_i, N_j) = -n\theta_i\theta_j$ für $i \neq j$.

Test auf	Verteilung
Daten	$N_1, \dots, N_k$
Raum	$\Theta = \{(\theta_1, \dots, \theta_k) \mid \theta_i \ge 0, \sum_i \theta_i = 1\}$
Nullhypothese	$\Theta_0 = \{\theta_0\} = \{(\theta_{10}, \dots, \theta_{k0})\}$
Verteilung	$F \sim \chi^2$-Verteilung mit $k - 1$ Freiheitsgraden
Teststatistik	$\chi^2 = \sum_{i=1}^k \frac{(N_i - n\theta_{i0})^2}{n\theta_{i0}}$
Näherung	χ^2-Verteilung ist brauchbare Näherung, sobald ca. 80% der $n\theta_{i0} \ge 4$ und der Rest ≥ 1. Sonst Klassen zusammenfassen.
Test	$\phi(x) = 1 \iff \chi^2 > c = F^{-1}(1 - \alpha)$

Theorem 6.24 (χ^2-Verteilung) *Asymptotisch hat χ^2 die gleiche Verteilung wie $\sum_{i=1}^{k-1} Y_i^2$, wobei $Y_1, \dots, Y_{k-1}$ i.i.d. $\mathcal{N}(0,1)$-verteilt sind. Diese Verteilung heisst Chiquadrat-Verteilung mit $k - 1$ Freiheitsgraden mit Dichte*

$$f_{k-1}(x) = 2^{-\frac{k-1}{2}} \Gamma\left(\frac{k-1}{2}\right)^{-1} e^{-x/2} x^{(k-3)/2}.$$

A Verteilungen & Kombinatorik

Name	Parameter	Träger	Zähldichte $f(k)$	Verteilung $F(x) = \mathbb{P}(X \le x)$	$\mathbb{E}$	$\mathbb{V}$	$\phi_X(u)$
Uniform	$n \in \mathbb{N}, k_i \in \mathbb{R}$	$\{k_i \mid i = 1, \dots, n\}$	$\frac{1}{n}$	$\frac{1}{n} \lvert\{i \mid k_i \le n\}\rvert$	$\frac{1}{n} \sum_{i=1}^n k_i$	$\frac{1}{n} (\sum_{i=1}^n k_i^2 - \frac{1}{n} (\sum_{i=1}^n k_i)^2)$	
Bernoulli	$p \in [0,1]$	$\{0,1\}$	$p\mathbb{1}_{\{1\}}(k) + (1-p)\mathbb{1}_{\{0\}}(k)$	$(1-p)\mathbb{1}_{[0,1]}(x) + \mathbb{1}_{[1,\infty)}(x)$	p	$p(1-p)$	$pe^{iu} + 1 - p$
Binomial	$n \in \mathbb{N}_+$, $p \in [0,1]$	$\{1, 2, \dots, n\}$	$\binom{n}{k} p^k (1-p)^{n-k}$	$\sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$	$(pe^{iu} + 1 - p)^n$
Geometrisch	$p \in (0,1)$	$\mathbb{N}_0$	$p(1-p)^k$	$1 - (1-p)^{\lfloor x+1 \rfloor}$	$\frac{1}{p} - 1$	$\frac{1}{p^2} - \frac{1}{p}$	
Hypergeom.	$N, M, n \in \mathbb{N}_+$ mit $n, M \le N$	$\{\max(0, n + M - N), \dots, \min(nM)\}$	$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$	$\sum_{i=\max(0,n-N)}^{\lfloor x \rfloor} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$	$n\frac{M}{N}$	$n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$	
Poisson	$\lambda \in \mathbb{R}_+$	$\mathbb{N}_0$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} e^{-\lambda}$	λ	λ	$e^{\lambda(e^{iu} - 1)}$

Name	Parameter	Träger	Dichte $f(x)$	Verteilung $F(x)$	$\mathbb{E}$	$\mathbb{V}$	
Uniform	$a, b \in \mathbb{R}$ mit $a < b$	$[a, b]$	$\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$	$\frac{x-a}{b-a}\mathbb{1}_{[a,b]}(x) + \mathbb{1}_{(b,\infty)}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Normal	$\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t$	μ	σ^2	$e^{iu\mu - u^2\sigma^2/2}$
Exponential	$\alpha \in \mathbb{R}_+$	$\mathbb{R}_0^+$	$\alpha e^{-\alpha x}$	$(1 - e^{-\alpha x})\mathbb{1}_{(0,\infty)}(x)$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{\alpha}{\alpha - iu}$
Cauchy	$\mu \in \mathbb{R}$	$\mathbb{R}$	$\frac{1}{\pi} \frac{1}{1+(x-\mu)^2}$	$\frac{1}{2} + \frac{1}{\pi} \arctan(x - \mu)$	$''\infty''$	$''\infty''$	$e^{-\lvert u \rvert}$
Beta	$p, q \in \mathbb{R}^+$	$[0,1]$	$\frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}$	$\mathbb{1}_{[0,1]}(x) \int_0^x \frac{t^{p-1}}{B(p,q)}(1-t)^{q-1} \mathrm{d}t + \mathbb{1}_{(1,\infty)}(x)$	$\frac{p}{p+q}$	$\frac{pq}{(p+q+1)(p+q)^2}$	

Kombinatorik	ohne Wiederholung	mit Wiederholung
Permutation (a, b)	$n!$	$\frac{n!}{k_1! \dots k_s!}$
Variation (a, b)	$k!\binom{n}{k}$	n^k
Kombination $\{a, b\}$	$\binom{n}{k}$	$\binom{n+k-1}{k}$