Einführung in die Statistik
Gliederung zur Vorlesung
im Sommersemester 2005

Markus Reiß
Universität Heidelberg

9. Januar 2006

Inhaltsverzeichnis

Kapitel 1

Diskrete Wahrscheinlichkeitsräume

1.1 Grundbegriffe

1.1 Definition. Ein diskreter Wahrscheinlichkeitsraum ist ein Tupel $(\Omega, \mathbb{P})$, bestehend aus einer abzählbaren (d.h. endlichen oder abzählbar unendlichen) Menge Ω und einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ mit den Eigenschaften

(a) $\mathbb{P}(\Omega) = 1$ (Normierung),

(b) Für $A_n \subseteq \Omega$, $n \geqslant 1$, paarweise disjunkt, gilt

$$\mathbb{P}\Big(\bigcup_{n\geqslant 1} A_n\Big) = \sum_{n\geqslant 1} \mathbb{P}(A_n) \text{ (σ-Additivität).}$$

Die Mengenfunktion $\mathbb{P}$ heißt Wahrscheinlichkeitsmaß.

1.2 Definition. Es sei Ω eine abzählbare Menge. Dann heißt jede Abbildung $p : \Omega \to [0,1]$ mit $\sum_{\omega\in\Omega} p(\omega) = 1$ Zähldichte auf Ω.

1.3 Lemma. Auf jedem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ wird durch

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

eine Zähldichte definiert.
Andererseits definiert eine Zähldichte p auf Ω ein Wahrscheinlichkeitsmaß $\mathbb{P}$ mittels

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega), \quad A \subseteq \Omega.$$

1.4 Lemma. Es sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Dann gilt:

(a) $\mathbb{P}(\varnothing) = 0$

(b) $A \subseteq B \subseteq \Omega \Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$

(c) $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$ für $A, B \subseteq \Omega$

(d) $\mathbb{P}(\bigcup_{n\geqslant 1} A_n) \leqslant \sum_{n\geqslant 1} \mathbb{P}(A_n)$ für beliebige $A_n \subseteq \Omega$, $n \geqslant 1$

(e) Gilt für $A_n \subseteq \Omega$, $n \geqslant 1$, und $A \subseteq \Omega$, dass $A_n \uparrow A$ (d.h. $A_n \subseteq A_{n+1}$, $n \geqslant 1$, und $\bigcup_n A_n = A$), so folgt $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ (σ-Stetigkeit).

1.2 Diskrete Verteilungen

Laplaceverteilung

Urnenmodelle

hypergeometrische Verteilung

Bernoulli-Verteilung

Binomialverteilung

geometrische Verteilung

Multinomialverteilung

Poissonverteilung

1.5 Satz (Poissonscher Grenzwertsatz)**.** Es seien $p_n \in [0,1]$ Erfolgswahrscheinlichkeiten mit $\lim_{n\to\infty} np_n = \lambda > 0$. Dann gilt für alle $k \geqslant 0$:

$$\lim_{n\to\infty} \mathrm{Bin}_{n,p_n}(k) = \mathrm{Poiss}_\lambda(k).$$

Kapitel 2

Bedingte Wahrscheinlichkeiten und Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeiten

2.1 Definition. Es seien A und B zwei Ereignisse mit $\mathbb{P}(B) > 0$. Mit

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

wird die bedingte Wahrscheinlichkeit von A unter/gegeben B bezeichnet.

2.2 Satz. Es sei B ein Ereignis mit $\mathbb{P}(B) > 0$. Dann gilt:

(a) $\mathbb{Q}(A) := \mathbb{P}(A \,|\, B)$, $A \subseteq \Omega$, definiert ein Wahrscheinlichkeitsmaß mit $\mathbb{Q}(B) = 1$.

(b) Ist B die Vereinigung paarweise disjunkter Ereignisse B_n mit $\mathbb{P}(B_n) > 0$, so folgt für jedes Ereignis A

$$\mathbb{P}(A \cap B) = \sum_n \mathbb{P}(B_n)\, \mathbb{P}(A \,|\, B_n).$$

(Formel von der totalen Wahrscheinlichkeit)

(c) Ist A ein Ereignis mit $\mathbb{P}(A) > 0$ und $\Omega = \bigcup_n B_n$ eine Zerlegung mit paarweise disjunkten Ereignissen $(B_n)_{n \geqslant 1}$, für die $\mathbb{P}(B_n) > 0$ gilt, so folgt für jedes n

$$\mathbb{P}(B_n \,|\, A) = \frac{\mathbb{P}(B_n)\, \mathbb{P}(A \,|\, B_n)}{\sum_m \mathbb{P}(B_m)\, \mathbb{P}(A \,|\, B_m)}.$$

(Formel von Bayes)

2.3 Lemma. Für Ereignisse $A_1, \ldots, A_n$ mit $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$ gilt die Multiplikationsformel

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

2.2 Unabhängigkeit

2.4 Definition.

(a) Zwei Ereignisse A und B heißen unabhängig, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ gilt.

(b) Eine Familie $(A_i)_{i \in I}$ von Ereignissen heißt unabhängig, falls für jede endliche Teilmenge $J \subseteq I$ gilt

$$\mathbb{P}\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} \mathbb{P}(A_j).$$

2.5 Satz. Die Ereignisse $A_1, \dots, A_n$ seien unabhängig. Dann gilt für alle $B_k \in \{A_k, A_k^\complement\}$, $k = 1, \dots, n$,

$$\mathbb{P}\Big(\bigcap_k B_k\Big) = \prod_{k=1}^n \mathbb{P}(B_k),$$

und die Ereignisse $B_1, \dots, B_n$ sind unabhängig.

2.3 Produkträume

2.6 Definition. Es seien $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$ diskrete Wahrscheinlichkeitsräume mit entsprechenden Zähldichten $p_1, \dots, p_n$. Durch die Produktmenge $\Omega = \Omega_1 \times \cdots \times \Omega_n$ und die Zähldichte (!)

$$p(\omega) = p(\omega_1, \dots, \omega_n) := \prod_{k=1}^n p_k(\omega_k), \quad \omega \in \Omega,$$

wird das Produkt $\otimes_{k=1}^n (\Omega_k, \mathbb{P}_k)$ der Wahrscheinlichkeitsräume definiert. Für das durch p induzierte Wahrscheinlichkeitsmaß P schreibt man auch $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n$ und nennt es das Produktmaß.

2.7 Satz. Es seien $(\Omega_1, \mathbb{P}_1), \dots, (\Omega_n, \mathbb{P}_n)$ diskrete Wahrscheinlichkeitsräume mit jeweiligen Ereignissen $A_k \subseteq \Omega_k$, $k = 1, \dots, n$. Setzt man $(\Omega, \mathbb{P}) := \otimes_{k=1}^n (\Omega_k, \mathbb{P}_k)$ und

$$A_k^{(k)} := \{\omega \in \Omega \,:\, \omega_k \in A_k\} \subseteq \Omega,$$

so sind die Ereignisse $A_1^{(1)}, \dots, A_n^{(n)}$ unabhängig in $(\Omega, \mathbb{P})$, und es gilt $\mathbb{P}(A_k^{(k)}) = \mathbb{P}_k(A_k)$ für alle $k = 1, \dots, n$.

Kapitel 3

Zufallsvariablen und ihre Momente

3.1 Zufallsvariablen

3.1 Definition. Es sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Jede Abbildung $X : \Omega \to S$ in eine beliebige Menge S heißt (diskrete) S-wertige Zufallsvariable. Im Fall $S = \mathbb{R}$ spricht man bloß von einer Zufallsvariablen, im Fall $S = \mathbb{R}^d$ von einem Zufallsvektor.

3.2 Definition. Ist X eine diskrete S-wertige Zufallsvariable, so heißt der diskrete Wahrscheinlichkeitsraum $(X(\Omega), \mathbb{P}^X)$ mit dem Wahrscheinlichkeitsmaß (!)

$$\mathbb{P}^X(A) := \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)), \quad A \subseteq X(\Omega),$$

von X induzierter Wahrscheinlichkeitsraum. $\mathbb{P}^X$ wird als Verteilung von X bezeichnet.

3.2 Unabhängigkeit von Zufallsvariablen

3.3 Definition. Es sei $(X_i)_{i\in I}$ eine Familie von Zufallsvariablen $X_i : \Omega \to S_i$. Die Familie $(X_i)_{i\in I}$ heißt unabhängig, falls für jede Wahl von Teilmengen $A_i \subseteq S_i$ die Ereignisse $(\{X_i \in A_i\})_{i\in I} = (X_i^{-1}(A_i))_{i\in I}$ unabhängig sind.

3.4 Satz. Die diskreten Zufallsvariablen $X_1, \dots, X_n$ sind genau dann unabhängig, wenn für alle $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$ gilt

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k).$$

3.5 Definition. Es seien $X_1 : \Omega \to S_1, \dots, X_n : \Omega \to S_n$ Zufallsvariablen. Dann ist $X : \Omega \to S_1 \times \dots \times S_n$ mit $X = (X_1, \dots, X_n)$ ebenfalls eine Zufallsvariable, und ihre Verteilung $\mathbb{P}^X$ heißt gemeinsame Verteilung von $X_1, \dots, X_n$. Andererseits heißt für jedes $k = 1, \dots, n$ die Verteilung $\mathbb{P}^{X_k}$ der k-ten Komponente die k-te Randverteilung von X.

3.6 Lemma. Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn ihre gemeinsame Verteilung $\mathbb{P}^X$ auf $X(\Omega) = X_1(\Omega) \times \cdots \times X_n(\Omega)$ das Produkt der Randverteilungen $\mathbb{P}^{X_k}$ ist: $\mathbb{P}^X = \mathbb{P}^{X_1} \otimes \cdots \otimes \mathbb{P}^{X_n}$.

3.3 Der Erwartungswert

3.7 Definition. Wir sagen, dass eine (reellwertige, diskrete) Zufallsvariable X auf $(\Omega, \mathbb{P})$ mit Zähldichte p einen Erwartungswert besitze, falls

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

gilt, und schreiben $X \in L^1 = L^1(\Omega, \mathbb{P})$. In diesem Fall definiert

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) p(\omega)$$

den <u>Erwartungswert</u> von X. Falls X nur nicht-negative Werte annimmt und $\sum_{\omega \in \Omega} X(\omega) p(\omega) = +\infty$ gilt, so schreiben wir $\mathbb{E}[X] = +\infty$.

3.8 Satz. Für diskrete Zufallsvariablen X, Y auf $(\Omega, \mathbb{P})$ gilt:

(a) $X \in L^1 \iff \sum_{x \in X(\Omega)} |x| \, \mathbb{P}(X = x) < \infty$. In diesem Fall erhält man den <u>Transformationssatz</u>

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \, \mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x p^X(x).$$

(b) $X, Y \in L^1$ mit $X(\omega) \leqslant Y(\omega)$ für alle $\omega \in \Omega$ impliziert $\mathbb{E}[X] \leqslant \mathbb{E}[Y]$ (Monotonie)

(c) Für $X, Y \in L^1$, $c \in \mathbb{R}$ ist $X + cY$ in L^1 und

$$\mathbb{E}[X + cY] = \mathbb{E}[X] + c \, \mathbb{E}[Y] \text{ (Linearität).}$$

(d) Sind $X, Y \in L^1$ unabhängig, so ist XY in L^1, und es gilt

$$\mathbb{E}[XY] = \mathbb{E}[X] \, \mathbb{E}[Y].$$

3.9 Korollar. Es seien $X_1, \ldots, X_n$ diskrete Zufallsvariablen sowie $g : X_1(\Omega) \times \cdots \times X_n(\Omega) \to \mathbb{R}$ eine beliebige Funktion. Dann gilt

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \sum_{x_1 \in X_1(\Omega)} \cdots \sum_{x_n \in X_n(\Omega)} g(x_1, \ldots, x_n) \, \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n),$$

sofern der Erwartungswert existiert.

3.10 Definition. Wir sagen, dass eine Zufallsvariable X in L^p für $p > 0$ liegt, falls $|X|^p \in L^1$ gilt. Für $X \in L^p$ und $p \in \mathbb{N}$ heißt $\mathbb{E}[X^p]$ das <u>p-te Moment</u> von X.

3.11 Lemma. Für $0 < p < q$ gilt: $X \in L^q \Rightarrow X \in L^p$.

3.12 Lemma. Für eine Zufallsvariable $X \in L^2$ nimmt die Funktion

$$\varphi : \mathbb{R} \to \mathbb{R}, \quad \varphi(z) = \mathbb{E}[(X - z)^2]$$

ihr (endliches) Minimum bei $z = \mathbb{E}[X]$ an.

3.4 Bedingte Erwartung und Vorhersage

3.13 Definition. Es seien X und Y Zufallsvariablen auf $(\Omega, \mathbb{P})$. Für $x \in X(\Omega)$ mit $\mathbb{P}(X = x) > 0$ ist die <u>bedingte Verteilung von Y gegeben $X = x$</u> das Wahrscheinlichkeitsmaß

$$\mathbb{P}^{Y|X=x}(A) := \mathbb{P}(Y \in A \,|\, X = x), \quad A \subseteq Y(\Omega).$$

Gilt $\sum_{y \in Y(\Omega)} |y| \, \mathbb{P}(Y = y \,|\, X = x) < \infty$, so wird der <u>bedingte Erwartungswert von Y gegeben $X = x$</u> definiert als

$$\mathbb{E}[Y \,|\, X = x] := \sum_{y \in Y(\Omega)} y \, \mathbb{P}(Y = y \,|\, X = x).$$

3.14 Lemma. Ist $Y \in L^1$ sowie $\mathbb{P}(X = x) > 0$ so existiert stets der bedingte Erwartungswert $\mathbb{E}[Y \,|\, X = x]$.

3.15 Lemma. Es seien X und Y Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann gilt, sofern alles wohldefiniert ist:

(a) Sind X und Y unabhängig, so ist $\mathbb{E}[Y \,|\, X = x] = E[Y]$.

(b) Ist $Y = \alpha X$, so ist $\mathbb{E}[Y \,|\, X = x] = \alpha x$.

(c) $\mathbb{E}[Y \,|\, X = x]$ ist bezüglich Y monoton und linear.

3.16 Definition. Es seien X, Y Zufallsvariablen mit $Y \in L^1$ und $\mathbb{P}(X = x) > 0$ für alle $x \in X(\Omega)$. Setze $\mu^{Y|X}(x) := \mathbb{E}[Y \,|\, X = x]$. Dann heißt die Zufallsvariable (!) $\mathbb{E}[Y \,|\, X](\omega) := \mu^{Y|X}(X(\omega))$ <u>bedingte Erwartung von Y gegeben X</u>.

3.17 Satz. Es seien X, Y Zufallsvariablen mit $Y \in L^1$ und $\mathbb{P}(X = x) > 0$ für alle $x \in X(\Omega)$. Dann gilt $\mathbb{E}[\mathbb{E}[Y \,|\, X]] = \mathbb{E}[Y]$.

3.18 Satz. Es seien X, Y Zufallsvariablen mit $Y \in L^2$ und $\mathbb{P}(X = x) > 0$ für alle $x \in X(\Omega)$. Es bezeichne

$$\mathcal{F}_X := \{Z \in L^2 \,|\, \exists h : X(\Omega) \to \mathbb{R} : \; Z(\omega) = h(X(\omega)) \text{ für alle } \omega \in \Omega\}$$

die Menge aller Zufallsvariablen in L^2, die sich als Funktion von X schreiben lassen. Dann nimmt die Funktion

$$\varphi : \mathcal{F}_X \to \mathbb{R}, \quad \varphi(Z) = \mathbb{E}[(Y - Z)^2]$$

ihr Minimum bei $Z = \mathbb{E}[Y \,|\, X]$ an.

3.5 Varianz und Kovarianz

3.19 Definition. Für eine Zufallsvariable X in L^2 bezeichnet

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

die <u>Varianz</u> von X. $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ heißt <u>Standardabweichung</u> von X.

3.20 Satz. Für $X, Y \in L^2$ gilt:

(a) $\operatorname{Var}(X) = 0 \iff \mathbb{P}(X = \mathbb{E}[X]) = 1$

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

(c) $\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

(d) $\operatorname{Var}(X + Y) \leqslant 2 \operatorname{Var}(X) + 2 \operatorname{Var}(Y)$

(e) Sind X und Y unabhängig, so gilt $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

3.21 Satz. Es seien X, Y Zufallsvariablen in L^2 sowie

$$\mathcal{L}_X := \{aX + b: \ a, b \in \mathbb{R}\}$$

die Menge aller Zufallsvariablen, die sich als affin-lineare Funktion von X schreiben lassen. Dann nimmt die Funktion

$$\varphi : \mathcal{L}_X \to \mathbb{R}, \quad \varphi(Z) = \mathbb{E}[(Y - Z)^2]$$

ihr Minimum bei $Z = a^* X + b^*$ an mit

$$a^* = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\operatorname{Var}(X)}, \quad b^* = \mathbb{E}[Y] - a^* \mathbb{E}[X],$$

falls $\operatorname{Var}(X) > 0$, sowie a^* beliebig, falls $\operatorname{Var}[X] = 0$. Im Fall $\operatorname{Var}(X) > 0$ gilt

$$\min_{Z \in \mathcal{L}_X} \varphi(Z) = \operatorname{Var}(Y) - \frac{(\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])])^2}{\operatorname{Var}[X]}.$$

3.22 Definition. Für X, Y definiert

$$\operatorname{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

die Kovarianz und im Fall $\sigma(X), \sigma(Y) > 0$

$$\rho(X, Y) := \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

die Korrelation von X und Y.

3.23 Satz. Für $X, Y, Z \in L^2$ gilt:

(a) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$, $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$

(b) $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + 2 \operatorname{Cov}(X, Y) + \operatorname{Var}(Y)$

(c) $\forall a, b \in \mathbb{R}: \ \operatorname{Cov}(aX + b, Y) = a \operatorname{Cov}(X, Y)$

(d) $\operatorname{Cov}(X + Y, Z) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(Y, Z)$

(e) Sind X und Y unabhängig, so gilt $\operatorname{Cov}(X, Y) = 0$
(X und Y sind unkorreliert)

(f) $|\operatorname{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$

(g) $\rho(X, Y) \in [-1, 1]$, falls die Korrelation wohldefiniert ist.

Kapitel 4

Testtheorie

4.1 Grundlagen

4.1 Definition. Ein diskretes statistisches Modell ist ein Tupel $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ bestehend aus einer abzählbaren Menge $\mathcal{X}$ (dem Stichprobenraum) und einer Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf $\mathcal{P}(\mathcal{X})$. Die mindestens zweielementige Menge Θ heißt Parametermenge und jedes $\vartheta \in \Theta$ Parameter.

4.2 Definition. Aufbau eines Testverfahrens:

(a) Wahl eines statistischen Modells $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$

(b) Formulierung von Hypothese und Alternative: $\Theta = \Theta_0 \dot\cup \Theta_1$
$\vartheta \in \Theta_0$: ϑ entspricht der Hypothese H_0
$\vartheta \in \Theta_1$: ϑ entspricht der Alternative H_1

(c) Wahl eines Irrtumsniveaus $\alpha \in (0,1)$ für den Fehler erster Art, sich bei Vorliegen der Hypothese für die Alternative zu entscheiden.

(d) Konstruktion eines (randomisierten) Tests $\varphi : \mathcal{X} \to [0,1]$ zum Niveau α:
$\varphi(x) = 0$: Entscheidung für H_0,
$\varphi(x) = 1$: Entscheidung für H_1,
$\varphi(x) \in (0,1)$: Entscheidung mit Wahrscheinlichkeit $\varphi(x)$ für H_1,
$\sup_{\vartheta\in\Theta_0} \mathbb{E}_\vartheta[\varphi] \leqslant \alpha$.

(e) Durchführen des Experiments

4.3 Definition. Weitere Begriffsbildungen:

(a) Jede Zufallsvariable φ auf $\mathcal{X}$ mit Werten in $[0,1]$ heißt Test.

(b) Gilt $\varphi(x) \in \{0,1\}$ für alle $x \in \mathcal{X}$, so heißt der Test φ nicht-randomisiert.

(c) Ist φ ein nicht-randomisierter Test, so heißt $\mathcal{X}_1 := \{x \in \mathcal{X} : \varphi(x) = 1\}$ Ablehnungsbereich oder kritischer Bereich des Tests.

(d) Die Funktion $G_\varphi : \Theta \to [0,1]$ mit $G_\varphi(\vartheta) = \mathbb{E}_\vartheta[\varphi]$ heißt Gütefunktion des Tests φ. Ist φ ein Test vom Niveau α, so gilt $G_\varphi(\vartheta_0) \leqslant \alpha$ für alle

$\vartheta_0 \in \Theta_0$. Für $\vartheta_1 \in \Theta_1$ heißt $G_\varphi(\vartheta_1)$ die Macht oder Schärfe von φ bei ϑ_1 und $\beta_\varphi(\vartheta_1) = 1 - G_\varphi(\vartheta_1)$ bezeichnet die Wahrscheinlichkeit für den Fehler 2. Art der Entscheidung für H_0, obwohl $\vartheta_1 \in \Theta_1$ vorliegt.

4.4 Definition.

Ein Test φ von $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ heißt gleichmäßig bester Test zum Niveau α, falls φ ein Test zum Niveau α ist und für jeden anderen Test ψ zum Niveau α gilt:

$$\forall\, \vartheta_1 \in \Theta_1 :\ G_\varphi(\vartheta_1) \geqslant G_\psi(\vartheta_1).$$

4.2 Einfache Alternativtests

In diesem Abschnitt liege stets das statistische Modell $(\mathcal{X}, (\mathbb{P}_0, \mathbb{P}_1))$ mit $\Theta_0 = \{0\}$, $\Theta_1 = \{1\}$ zugrunde. Außerdem möge für die zugehörigen Zähldichten $p_0(x) + p_1(x) > 0$ für alle $x \in \mathcal{X}$ gelten. Hypothese und Alternative heißen in diesem Fall einfach, weil sie nur jeweils einen Parameter enthalten.

4.5 Definition. Der Likelihood-Quotient von $\mathbb{P}_1$ bezüglich $\mathbb{P}_0$ ist gegeben durch

$$R(x) = \begin{cases} p_1(x)/p_0(x), & \text{falls } p_0(x) > 0, \\ +\infty, & \text{falls } p_0(x) = 0. \end{cases}$$

Jeder Test φ der Form

$$\varphi(x) = \begin{cases} 1, & \text{falls } R(x) > c, \\ 0, & \text{falls } R(x) < c, \\ \gamma, & \text{falls } R(x) = c \end{cases}$$

mit beliebigem $c \geqslant 0$ und $\gamma \in [0,1]$ heißt Neyman-Pearson-Test.

4.6 Satz. Für das Testen von $H_0 : \vartheta = 0$ gegen $H_1 : \vartheta = 1$ gilt:

(a) Ist φ^* ein Neyman-Pearson-Test, so ist φ^* mindestens so mächtig wie jeder andere Test φ mit $\mathbb{E}_0[\varphi] \leqslant \mathbb{E}_0[\varphi^*]$.

(b) Für jedes Niveau $\alpha \in (0,1)$ existiert ein Neyman-Pearson-Test φ^* mit exakt $\mathbb{E}_0[\varphi^*] = \alpha$.

(c) Ein (gleichmäßig) bester Test zum Niveau α ist gegeben durch einen Neyman-Pearson-Test φ^* mit $\mathbb{E}_0[\varphi^*] = \alpha$.

4.3 Beste einseitige Tests

4.7 Definition. Ein diskretes statistisches Modell $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ mit $\mathcal{X} \subseteq \mathbb{R}$ und $\Theta \subseteq \mathbb{R}$ hat wachsende Likelihood-Quotienten, wenn für alle $\vartheta < \vartheta'$ der Likelihood-Quotient

$$R_{\vartheta':\vartheta}(x) = \begin{cases} p_{\vartheta'}(x)/p_\vartheta(x), & \text{falls } p_\vartheta(x) > 0, \\ +\infty, & \text{falls } p_\vartheta(x) = 0 \end{cases}$$

monoton wachsend in x ist.

4.8 Satz. Es sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell mit wachsenden Likelihood-Quotienten. Für jedes $\vartheta_0 \in \Theta, \alpha \in (0,1)$ existiert dann ein gleichmäßig bester Test zum Niveau α von $H_0 : \vartheta \leqslant \vartheta_0$ gegen $H_1 : \vartheta > \vartheta_0$. Dieser hat die Gestalt

$$\varphi(x) = \begin{cases} 1, & \text{falls } x > c_\alpha, \\ 0, & \text{falls } x < c_\alpha, \\ \gamma_\alpha, & \text{falls } x = c_\alpha, \end{cases}$$

wobei sich $c_\alpha \in \mathbb{R}$ und $\gamma_\alpha \in [0,1]$ aus der Forderung $\mathbb{E}_{\vartheta_0}[\varphi] = \alpha$ ergeben. Ferner ist die Gütefunktion G_φ monoton wachsend.

4.4 Der χ^2-Anpassungstest

Wir betrachten als statistisches Modell $\Theta = \{\vartheta \in (0,1)^r : \vartheta_1 + \cdots + \vartheta_r = 1\}$, $\mathcal{X} = \{x \in \{0,\ldots,n\}^r : x_1 + \cdots + x_r = n\}$ und $\mathrm{Mult}_{n;\vartheta_1,\ldots,\vartheta_r}$ die Multinomialverteilung mit n Versuchen und Wahrscheinlichkeiten $\vartheta_1, \ldots, \vartheta_r$ für die Klassen $1, \ldots, r$. Es soll die Hypothese $H_0 : \vartheta = \bar{\vartheta}$ gegen die Alternative $H_1 : \vartheta \neq \bar{\vartheta}$ getestet werden (sogenannter Signifikanztest).

4.9 Definition. Für ein beliebiges Testproblem heißt jeder Test φ der Form

$$\varphi(x) = \begin{cases} 1, & \text{falls } R(x) > c, \\ 0, & \text{falls } R(x) < c, \\ \gamma, & \text{falls } R(x) = c \end{cases} \qquad \text{mit } R(x) = \frac{\sup_{\vartheta\in\Theta_1} p_\vartheta(x)}{\sup_{\vartheta\in\Theta_0} p_\vartheta(x)}$$

Likelihood-Quotienten-Test.

4.10 Lemma. Für das vorliegende Testproblem gilt

$$\log(R(x)) = \sum_{i=1}^r x_i \log\left(\frac{x_i}{n\bar{\vartheta}_i}\right) \approx \tfrac{1}{2} V^2(x)$$

mit Pearsons χ^2-Statistik

$$V^2(x) = \sum_{i=1}^r \frac{(x_i - n\bar{\vartheta}_i)^2}{n\bar{\vartheta}_i}.$$

4.11 Satz. Für alle $v > 0$ gilt (mit Kenntlichmachung der Abhängigkeit von n):

$$\lim_{n\to\infty} \mathbb{P}_{\bar{\vartheta}}(V_n^2 \leqslant v) = \int_0^v f_{\chi^2_{r-1}}(x)\,dx,$$

wobei $f_{\chi^2_m}(x) = \frac{1}{\Gamma(n/2)2^{m/2}} x^{m/2-1} e^{-x/2}$ die Dichte der χ^2-Verteilung mit m Freiheitsgraden bezeichnet. Ebenso gilt

$$\lim_{n\to\infty} \mathbb{P}_{\bar{\vartheta}}(2\log(R_n) \leqslant v) = \int_0^v f_{\chi^2_{r-1}}(x)\,dx.$$

Kapitel 5

Allgemeine Wahrscheinlichkeitsräume und Dichten

5.1 σ-Algebren und Wahrscheinlichkeitsräume

5.1 Satz. Sei $\Omega = \{0,1\}^{\mathbb{N}}$ der Raum der 0-1-Folgen. Dann gibt es keine Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ die normiert und σ-additiv ist sowie folgende Invarianz eigenschaft besitzt

$$\forall\, A \subseteq \Omega,\ n \geqslant 1 : \mathbb{P}(T_n(A)) = \mathbb{P}(A) \text{ mit } T_n(\omega) = (\omega_1, \ldots, \omega_{n-1}, 1-\omega_n, \omega_{n+1}, \ldots).$$

5.2 Definition. Für $\Omega \neq \varnothing$ heißt $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ σ-Algebra, falls gilt:

(a) $\Omega \in \mathcal{F}$,

(b) $A \in \mathcal{F} \Rightarrow A^{\complement} \in \mathcal{F}$,

(c) $A_n \in \mathcal{F},\ n \geqslant 1 \Rightarrow \bigcup_{n\geqslant 1} A_n \in \mathcal{F}$.

Das Paar $(\Omega, \mathcal{F})$ heißt messbarer Raum oder Messraum.

5.3 Lemma. Für eine σ-Algebra $\mathcal{F}$ gilt:

(a) $\varnothing \in \mathcal{F}$,

(b) $A_n \in \mathcal{F},\ n \geqslant 1 \Rightarrow \bigcap_{n\geqslant 1} A_n \in \mathcal{F}$,

(c) $A, B \in \mathcal{F} \Rightarrow A \cup B,\ A \cap B \in \mathcal{F}$.

5.4 Lemma. Zu jedem Mengensystem $\mathcal{M} \subseteq \mathcal{P}(\Omega)$ gibt es eine (bzgl. Mengeninklusion) kleinste σ-Algebra die $\mathcal{M}$ enthält. Diese von $\mathcal{M}$ erzeugte σ-Algebra wird mit $\sigma(\mathcal{M})$ bezeichnet.

5.5 Definition. Es sei $\Omega = \mathbb{R}^d$ und

$$\mathcal{M}_d = \left\{ \prod_{i=1}^{d} [a_i, b_i] \ : \ a_i < b_i; a_i, b_i \in \mathbb{Q} \right\}$$

das System aller d-dimensionalen Quader mit rationalen Endpunkten. Dann heißt $\mathfrak{B}_{\mathbb{R}^d} = \sigma(\mathcal{M}_d)$ die Borel-σ-Algebra auf $\mathbb{R}^d$ und jedes $B \in \mathfrak{B}_{\mathbb{R}^d}$ Borelmenge.

5.6 Lemma. Jede offene und jede abgeschlossene Menge $B \subseteq \mathbb{R}^d$ ist eine Borelmenge.

5.7 Definition. Ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ bestehend aus einer Menge Ω, einer σ-Algebra $\mathcal{F}$ auf Ω und einer Abbildung $\mathbb{P} : \mathcal{F} \to [0, 1]$ mit $\mathbb{P}(\Omega) = 1$ und

$$\mathbb{P}\Big(\bigcup_{n \geqslant 1} A_n\Big) = \sum_{n \geqslant 1} \mathbb{P}(A_n) \text{ für alle paarweise disjunkten } A_n \in \mathcal{F},\ n \geqslant 1$$

heißt (allgemeiner) Wahrscheinlichkeitsraum. $\mathbb{P}$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{F}$ und jede Menge $A \in \mathcal{F}$ heißt Ereignis.

5.8 Satz. Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{M}$ ein $\mathcal{F}$ erzeugendes Mengensystem. Ist $\mathcal{M}$ durchschnittstabil (d.h. $A, B \in \mathcal{M} \Rightarrow A \cap B \in \mathcal{M}$), so ist das Wahrscheinlichkeitsmaß $\mathbb{P}$ bereits durch seine Werte auf $\mathcal{M}$ eindeutig bestimmt.

5.2 Exkurs: mehrdimensionales Riemann-Integral

5.3 Modelle mit Dichten

5.9 Definition. Eine integrierbare Funktion $f : \mathbb{R}^d \to [0, \infty)$ mit

$$\int_{\mathbb{R}^d} f(x)\, dx = 1$$

heißt Wahrscheinlichkeitsdichte oder bloß Dichte auf $\mathbb{R}^d$.

5.10 Satz. f sei eine Dichte auf $\mathbb{R}^d$. Für jeden Quader $Q = \prod_{i=1}^d [a_i, b_i]$ setze

$$\mathbb{P}(Q) = \int_Q f(x)\, dx = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(x_1, \ldots, x_d)\, dx_d \cdots dx_1.$$

Dann lässt sich $\mathbb{P}$ in eindeutiger Weise zu einem Maß auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ fortsetzen. Für jede Borelmenge gilt

$$\mathbb{P}(B) = \int_B f(x)\, dx,$$

sofern das Integral wohldefiniert ist.

Gleichmäßige Verteilung auf dem Intervall

Exponentialverteilung

Normalverteilung

Gleichverteilung auf einem Gebiet

Produkte von Dichten

5.11 Definition. Jede Dichte f auf $\mathbb{R}$ induziert die Verteilungsfunktion

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \mathbb{P}((-\infty, x]),$$

wobei $\mathbb{P}$ das von f induzierte Wahrscheinlichkeitsmaß bezeichnet.

5.12 Lemma. Die von einer Dichte induzierte Verteilungsfunktion F hat folgende Eigenschaften:

(a) F ist monoton wachsend.

(b) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$

(c) F ist stetig.

5.4 Zufallsvariablen

5.13 Definition. Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(S, \mathcal{S})$ ein Messraum. Dann heißt eine Funktion $f : \Omega \to S$ messbar (bzgl. $(\mathcal{F}, \mathcal{S})$), falls

$$\forall\, A \in \mathcal{S} :\ f^{-1}(A) \in \mathcal{F}$$

gilt. Jede solche messbare Funktion heißt $(S, \mathcal{S})$-wertige Zufallsvariable. Für $S = \mathbb{R}^d$ wird kanonisch $\mathcal{S} = \mathfrak{B}_{\mathbb{R}^d}$ gewählt, und man spricht bloß von einer Zufallsvariablen ($d = 1$) bzw. einem Zufallsvektor ($d \geqslant 2$).
Die Verteilung einer $(S, \mathcal{S})$-wertigen Zufallsvariablen X ist das Wahrscheinlichkeitsmaß (!)

$$\mathbb{P}^X(A) := \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)), \quad A \in \mathcal{S}.$$

Der Wahrscheinlichkeitsraum $(S, \mathcal{S}, \mathbb{P}^X)$ heißt von X induzierter Wahrscheinlichkeitsraum.

5.14 Satz. Jede stetige Funktion $f : \mathbb{R}^m \to \mathbb{R}^n$ und jede stückweise stetige Funktion $g : \mathbb{R} \to \mathbb{R}^n$ ist Borel-messbar.

5.15 Definition. Für eine (reellwertige) Zufallsvariable X heißt

$$F^X(x) = \mathbb{P}(X \leqslant x), \quad x \in \mathbb{R},$$

Verteilungsfunktion von X. Gilt für eine $\mathbb{R}^d$-wertige Zufallsvariable X und eine Dichte f^X auf $\mathbb{R}^d$

$$\mathbb{P}(X \in Q) = \int_Q f^X(x)\,dx \text{ für alle Quader } Q,$$

so heißt f^X Dichte von X.

5.16 Lemma. Ist X eine reellwertige Zufallsvariable mit Dichte f^X, so gilt für die Verteilungsfunktion

$$F^X(x) = \int_{-\infty}^{x} f^X(u)\,du, \quad x \in \mathbb{R}\,.$$

Ist f^X stetig in $x \in \mathbb{R}$, so gilt $(F^X)'(x) = f^X(x)$.

5.17 Satz. Besitzt die Zufallsvariable X die Dichte f^X und ist die Funktion $\varphi : X(\Omega) \subseteq \mathbb{R} \to \mathbb{R}$ differenzierbar mit $\varphi'(x) > 0$ für alle x oder aber $\varphi'(x) < 0$ für alle x, so hat $Y = \varphi(X)$ die Dichte

$$f^Y(y) = f^X(\varphi^{-1}(y))|(\varphi^{-1})'(y)| = \frac{f^X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}.$$

5.18 Lemma. Besitzt der Zufallsvektor $X = (X_1, \ldots, X_n)$ die Dichte f, so besitzt jede Komponente X_i ebenfalls eine Dichte und zwar

$$f_i(x_i) = \int_{\mathbb{R}^{n-1}} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n, \quad x_i \in \mathbb{R}\,.$$

In Worten: die i-te Randverteilung von X besitzt die Dichte f_i, die sich durch Integration über die anderen Koordinaten ergibt.

5.19 Definition. Eine Familie $(X_i)_{i\in I}$ von $(S_i, \mathcal{S}_i)$-wertigen Zufallsvariablen heißt unabhängig, falls für alle $A_i \in \mathcal{S}_i$ die Ereignisse $(\{X_i \in A_i\})_{i\in I}$ unabhängig sind.

5.20 Satz. Reellwertige Zufallsvariablen $(X_i)_{i\in I}$ sind bereits dann unabhängig, falls für alle $x_i \in \mathbb{R}$ die Ereignisse $(\{X_i \leqslant x_i\})_{i\in I}$ unabhängig sind.

5.21 Satz. Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Dichten $f_1, \ldots, f_n$, so besitzt $X = (X_1, \ldots, X_n)$ die Dichte

$$f(x) = f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i), \quad x \in \mathbb{R}^n\,.$$

Besitzt umgekehrt X eine Dichte f, die sich in dieser Produktform schreiben lässt, so sind die Komponenten $X_1, \ldots, X_n$ unabhängig mit Dichten $f_1, \ldots, f_n$.

5.22 Satz. Sind X_1 und X_2 unabhängige Zufallsvariablen mit Dichten f_1 und f_2, so besitzt $X_1 + X_2$ die Dichte

$$f^{X_1+X_2}(x) := f_1 * f_2(x) := \int_{-\infty}^{\infty} f_1(x-y) f_2(y)\, dy, \quad x \in \mathbb{R}\,.$$

5.23 Definition. Die gerade definierte Funktion $f_1 * f_2$ heißt Faltung von f_1 und f_2.

5.5 Erwartungswert und Varianz bei Zufallsvariablen mit Dichten

5.24 Definition. Es sei X eine Zufallsvariable mit Dichte f^X. Falls $\int_{-\infty}^{\infty} |x| f^X(x)\, dx < \infty$ gilt, so bezeichnet

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f^X(x)\, dx$$

den Erwartungswert von X. Ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit Dichte f, so wird $\int_{-\infty}^{\infty} x f(x)\, dx$ als Erwartungswert der Verteilung $\mathbb{P}$ bezeichnet, sofern $\int_{-\infty}^{\infty} |x| f(x)\, dx$ endlich ist.

5.25 Satz. Es seien X, Y Zufallsvariablen mit Dichten f^X, f^Y sowie $Y = g(X)$ mit einer injektiven differenzierbaren Funktion g. Im Fall $\int_{-\infty}^{\infty} |g(x)| f^X(x)\,dx < \infty$ existiert der Erwartungswert von Y, und es gilt

$$E[Y] := \int_{-\infty}^{\infty} y f^Y(y)\,dy = \int_{-\infty}^{\infty} g(x) f^X(x)\,dx.$$

5.26 Definition. Man definiert allgemeiner für beliebige integrierbare Funktionen g mit $\int_{-\infty}^{\infty} |g(x)| f^X(x)\,dx < \infty$:

$$\mathbb{E}[g(X)] := \int_{-\infty}^{\infty} g(x) f^X(x)\,dx.$$

5.6 Die mehrdimensionale Normalverteilung

5.27 Lemma. Es sei X ein p-dimensionaler Zufallsvektor mit Erwartungswertvektor μ und Kovarianzmatrix Σ. Dann besitzt der q-dimensionale Zufallsvektor $Y = AX + \beta$, wobei A eine $p \times q$-Matrix und β ein q-dimensionaler Vektor ist, den Erwartungswertvektor $A\mu + \beta$ und die Kovarianzmatrix $A\Sigma A^T$.

5.28 Lemma. Jede Kovarianzmatrix eines Zufallsvektors ist, falls sie existiert, symmetrisch und positiv semi-definit.

5.29 Satz. (Cramér-Wold) Die Verteilung eines p-dimensionalen Zufallsvektors ist eindeutig bestimmt durch die eindimensionalen Verteilungen von $c^T X$ für alle $c \in \mathbb{R}^p$.

5.30 Korollar. Der Zufallsvektor $X = (X_1, \dots, X_p)$ ist genau dann standardnormalverteilt, wenn $c^T X \sim \mathcal{N}(0, |c|^2)$ für alle $c \in \mathbb{R}^p$ gilt.

5.31 Definition. Es sei X ein p-dimensionaler Zufallsvektor mit Erwartungswertvektor μ und Kovarianzmatrix Σ. X heißt normalverteilt mit Parametern (μ, Σ) oder $\mathcal{N}(\mu, \Sigma)$-verteilt, falls für alle $c \in \mathbb{R}^p$ gilt $c^T X \sim N(c^T\mu, c^T\Sigma c)$.

5.32 Satz. Ist X $\mathcal{N}(\mu, \Sigma)$-verteilt mit $\mu = (\mu_1, \dots, \mu_p)^T$ und $\Sigma = (\sigma_{ij}^2)_{i,j=1,\dots,p}$, so gilt:

(a) $X_i \sim \mathcal{N}(\mu_i, \sigma_{ii}^2)$ für alle $i = 1, \dots, p$.

(b) $X_1, \dots, X_p$ sind genau dann unabhängig, wenn sie paarweise unkorreliert sind.

5.33 Satz. Für $X \sim \mathcal{N}(\mu, \Sigma)$, $A \in \mathbb{R}^{p \times q}$, $\beta \in \mathbb{R}^p$ gilt

$$AX + \beta \sim \mathcal{N}(A\mu + \beta, A\Sigma A^T).$$

5.34 Satz. Ist X ein $\mathcal{N}(\mu, \Sigma)$-verteilter Zufallsvektor im $\mathbb{R}^p$ mit strikt positiv-definiter Kovarianzmatrix Σ, so besitzt X die Dichte

$$f^X(x) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\Big(-\tfrac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \Big).$$

χ^2-Verteilung und Anwendung auf Testprobleme

5.7 Übertragung von Ergebnissen auf den allgemeinen Fall

5.35 Satz. Folgende Definitionen und Aussagen gelten auch im allgemeinen Fall, sofern Ereignisse als Elemente der σ-Algebra, Zufallsvariablen als messbare Funktionen und bei Erwartungswerten die Existenz von Dichten angenommen werden:
Lemma 1.4, Definition 2.1, Satz 2.2, Lemma 2.3, Definition 2.4, Satz 2.5, Definition 3.3, Definition 3.5, Satz 3.8(b),(c),(d), Definition 3.10, Lemma 3.11, Lemma 3.12, Definition 3.19, Satz 3.20, Satz 3.21, Definition 3.22, Satz 3.23.

Kapitel 6

Grenzwertsätze

6.1 Gesetze der großen Zahlen

6.1 Satz. (allgemeine Markow-Ungleichung) Es sei X eine Zufallsvariable und $\varphi : [0,\infty) \to [0,\infty)$ monoton wachsend. Dann gilt für jedes $K > 0$ mit $\varphi(K) > 0$:

$$\mathbb{P}(|X| \geqslant K) \leqslant \frac{\mathbb{E}[\varphi(|X|)]}{\varphi(K)},$$

wobei der Fall, dass die rechte Seite unendlich ist, trivial ist.

6.2 Korollar. (Tschebyschew-Ungleichung) Ist X eine Zufallsvariable in L^2, so gilt für jedes $\varepsilon > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant \varepsilon) \leqslant \frac{\operatorname{Var}(X)}{\varepsilon^2}.$$

6.3 Satz. (schwaches Gesetz der großen Zahlen) Es seien $X_1, \dots, X_n$ unkorrelierte Zufallsvariablen mit demselben Erwartungswert μ und $\sup_i \operatorname{Var}(X_i) \leqslant V < \infty$. Dann erfüllt das arithmetische Mittel

$$A_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|A_n - \mu| > \varepsilon) = 0.$$

6.4 Korollar. (Weierstraßscher Approximationssatz) Zur stetigen Funktion $f : [0,1] \to \mathbb{R}$ definiere das zugehörige Bernstein-Polynom n-ten Grades

$$f_n(x) := \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) \binom{n}{k} x^k (1-x)^{n-k}, \quad x \in [0,1].$$

Dann gilt $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$ mit $\|g\|_\infty := \sup_{x\in[0,1]} |g(x)|$.

6.5 Definition. Es seien $(X_n)_{n\geqslant 1}$ und X Zufallsvariablen auf demselben Raum $(\Omega, \mathcal{F}, \mathbb{P})$. Man sagt, dass X_n stochastisch (oder auch in Wahrscheinlichkeit) gegen X konvergiert für $n \to \infty$, falls für alle $\varepsilon > 0$ gilt

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

Man sagt, dass X_n fast sicher gegen X konvergiert, falls

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

6.6 Lemma. Fast sichere Konvergenz impliziert stochastische Konvergenz, aber nicht umgekehrt.

6.7 Satz. (starkes Gesetz der großen Zahlen) Es seien $(X_n)_{n\geqslant 1}$ paarweise unabhängige Zufallsvariablen in L^1, die identisch verteilt sind. Dann konvergiert $A_n := \frac{1}{n}\sum_{i=1}^n X_i$ fast sicher gegen den Erwartungswert $\mu = \mathbb{E}[X_i]$.

6.2 Der zentrale Grenzwertsatz

6.8 Definition. Die Zufallsvariablen $(X_n)_{n\geqslant 1}$ konvergieren in Verteilung gegen die Zufallsvariable X (bzw. genauer: die Verteilungen $(\mathbb{P}^{X_n})_{n\geqslant 1}$ konvergieren schwach gegen die Verteilung $\mathbb{P}^X$), falls für jede stetige beschränkte Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ gilt

$$\lim_{n\to\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

Notation: $X_n \xrightarrow{\mathcal{L}} X$ bzw. $\mathbb{P}^{X_n} \xrightarrow{\mathcal{L}} \mathbb{P}^X$ bzw. $X_n \xrightarrow{\mathcal{L}} \mathbb{P}^X$.

6.9 Satz. Es sind äquivalent:

(a) $X_n \xrightarrow{\mathcal{L}} X$

(b) Die Verteilungsfunktionen erfüllen $F^{X_n}(x) \to F^X(x)$ für alle $x \in \mathbb{R}$, an denen F^X stetig ist (Stetigkeitspunkte).

6.10 Lemma. Stochastische Konvergenz von $(X_n)_{n\geqslant 1}$ gegen X impliziert Konvergenz in Verteilung: $X_n \xrightarrow{\mathcal{L}} X$.

6.11 Satz. (zentraler Grenzwertsatz) Es sei $(X_n)_{n\geqslant 1}$ eine Folge unabhängiger identisch verteilter Zufallsvariablen in L^2. Dann gilt

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mathbb{E}[X_i]}{\sigma(X_i)} \xrightarrow{\ \mathcal{L}\ } \mathcal{N}(0,1).$$

Insbesondere gilt für $a < b$ also $\mathbb{P}(a < S_n^* \leqslant b) \to \Phi(b) - \Phi(a)$ mit der Verteilungsfunktion Φ der Standardnormalverteilung $\mathcal{N}(0,1)$.

Kapitel 7

Schätztheorie

7.1 Lineare Regression und die Methode der kleinsten Quadrate

7.1 Definition. Ein (einfaches eindimensionales) lineares Regressionsmodell ist gegeben durch die Beobachtungen

$$Y_k = ax_k + b + \varepsilon_k, \quad k = 1, \ldots, n,$$

mit deterministischen Werten $x_1, \ldots, x_n \in \mathbb{R}$ (Versuchsplan), unbekannten Koeffizienten a, $b \in \mathbb{R}$ und unkorrelierten Zufallsvariablen $\varepsilon_1, \ldots, \varepsilon_n$, die $\mathbb{E}[\varepsilon_k] = 0$, $\mathrm{Var}[\varepsilon_k] = \sigma^2 > 0$ für alle $k = 1, \ldots, n$ erfüllen.

7.2 Definition. Die Kleinste-Quadrate-Schätzung vom Tupel (a, b) ist definiert als

$$(\hat{a}, \hat{b}) := \mathrm{argmin}_{(a',b') \in \mathbb{R}^2} \sum_{k=1}^{n} (Y_k - a'x_k - b')^2.$$

7.3 Lemma. Es gilt für die Kleinste-Quadrate-Schätzung

$$\hat{b} = \bar{Y} - \hat{a}\bar{x}, \quad \hat{a} = \frac{\bar{\rho}_{xY}}{\bar{\sigma}_x^2}$$

mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ und

$$\bar{\sigma}_x^2 = \tfrac{1}{n-1} \textstyle\sum_{i=1}^n (x_i - \bar{x})^2, \qquad \bar{\rho}_{xY} = \tfrac{1}{n-1} \textstyle\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}).$$

7.4 Definition. Ein Schätzer $\hat{\vartheta}$ für einen Parameter ϑ heißt erwartungstreu, falls $\mathbb{E}_\vartheta[\hat{\vartheta}] = \vartheta$ gilt.

7.5 Lemma. Die Kleinste-Quadrate-Schätzungen $\hat{a}$ und $\hat{b}$ sind erwartungstreue Schätzer von a bzw. b.

7.6 Lemma. Für den mittleren quadratischen Fehler eines Schätzers $\hat{\vartheta}$ für ϑ, der in L^2 liegt, gilt die Bias-Varianz-Zerlegung

$$\mathbb{E}_\vartheta[(\hat{\vartheta} - \vartheta)^2] = \big(\mathbb{E}_\vartheta[\hat{\vartheta} - \vartheta]\big)^2 + \mathrm{Var}_\vartheta(\hat{\vartheta}).$$

7.7 Satz. (Gauß-Markov) Die Kleinste-Quadrate-Schätzungen $\hat{a}$ und $\hat{b}$ haben unter allen linearen und erwartungstreuen Schätzern die kleinste Varianz.

7.8 Satz. Die Varianz σ^2 im linearen Regressionsmodell wird erwartungstreu geschätzt durch

$$\hat{\sigma}^2 := \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{a}x_i - \hat{b})^2,$$

wobei $\hat{a}$ und $\hat{b}$ die Kleinste-Quadrate-Schätzungen bezeichnen.

7.2 Allgemeine Parameterschätzungen

7.9 Definition. Ist $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein allgemeines statistisches Modell mit $\Theta \subseteq \mathbb{R}^d$, so heißt jede Θ-wertige Zufallsvariable auf $(\mathcal{X}, \mathcal{F})$ Schätzer von ϑ.

7.10 Definition. Eine Folge $(\hat{\vartheta}_n)_{n\geqslant 1}$ von Schätzern eines Parameters ϑ heißt konsistent, falls $\lim_{n\to\infty}\hat{\vartheta}_n = \vartheta$ in Wahrscheinlichkeit gilt.

7.11 Definition. Ein diskretes statistisches Modell $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ mit einem offenen Intervall $\Theta \subseteq \mathbb{R}$ heißt regulär, falls gilt:

(a) $\forall x \in \mathcal{X}\ \forall\vartheta \in \Theta : p_\vartheta(x) > 0,\ \dot{p}_\vartheta(x) = \frac{\partial}{\partial\vartheta}p_\vartheta(x)$ existiert,

(b) $\forall\vartheta \in \Theta :\ \dot{p}_\vartheta/p_\vartheta \in L^2,\ \sum_{x\in\mathcal{X}}\dot{p}_\vartheta(x) = \frac{\partial}{\partial\vartheta}\sum_{x\in\mathcal{X}}p_\vartheta(x) = 0.$

Ein statistisches Modell mit Dichten heißt entsprechend regulär, falls obige Eigenschaften für die Dichten f_ϑ erfüllt sind, wobei die Summen durch Integrale zu ersetzen sind.

7.12 Satz. (Cramér-Rao-Ungleichung) Ist $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein reguläres statistisches Modell, so gilt für jeden erwartungstreuen Schätzer $\hat{\vartheta}$ von ϑ folgende untere Schranke für die Varianz:

$$\mathrm{Var}_\vartheta(\hat{\vartheta}) \geqslant \frac{1}{I(\vartheta)}$$

mit der Fisher-Information

$$I(\vartheta) := \mathbb{E}_\vartheta\Big[\Big(\frac{\dot{p}_\vartheta}{p_\vartheta}\Big)^2\Big] \text{ bzw. } I(\vartheta) := \mathbb{E}_\vartheta\Big[\Big(\frac{\dot{f}_\vartheta}{f_\vartheta}\Big)^2\Big].$$

7.3 Konfidenzbereiche

7.13 Definition. Es seien $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell sowie $\alpha \in (0,1)$. Eine Abbildung $C : \mathcal{X} \to \mathcal{P}(\Theta)$, die jedem Versuchsausgang x eine Teilmenge $C(x)$ der Parametermenge zuordnet, heißt Konfidenzbereich zum Konfidenz-Niveau $1-\alpha$, falls

$$\forall\vartheta \in \Theta :\ \mathbb{P}_\vartheta(\{x \in \mathcal{X} :\ \vartheta \in C(x)\}) \geqslant 1-\alpha.$$

7.14 Satz. Es sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell.

(a) Ist für jedes $\bar{\vartheta} \in \Theta$ ein nicht-randomisierter Test $\varphi_{\bar{\vartheta}}$ von $H_0 : \vartheta = \bar{\vartheta}$ gegen $H_1 : \vartheta \neq \bar{\vartheta}$ zum Signifikanzniveau α gegeben, so definiert

$$C(x) := \{\bar{\vartheta} \in \Theta : \ \varphi_{\bar{\vartheta}}(x) = 0\}$$

einen Konfidenzbereich zum Konfidenz-Niveau $1 - \alpha$.

(b) Umgekehrt wird durch einen Konfidenzbereich $C(\bullet)$ zum Konfidenz-Niveau $1 - \alpha$ mittels

$$\varphi_{\bar{\vartheta}}(x) := 1 - \mathbf{1}_{C(x)}(\bar{\vartheta})$$

ein Test von $H_0 : \vartheta = \bar{\vartheta}$ gegen $H_1 : \vartheta \neq \bar{\vartheta}$ zum Niveau α definiert.