

# Wahrscheinlichkeit und Statistik Prüfungsnotizen

Fabian Bösiger

30.06.2020

## Inhalt

Wahrscheinlichkeiten . . . . .	2
Grundbegriffe . . . . .	2
Diskrete Wahrscheinlichkeitsräume . . . . .	2
Bedingte Wahrscheinlichkeiten . . . . .	2
Unabhängigkeit . . . . .	3
Diskrete Zufallsvariablen und Verteilungen . . . . .	3
Grundbegriffe . . . . .	3
Erwartungswerte . . . . .	3
Gemeinsame Verteilungen und unabhängige Zufallsvariablen . . . . .	4
Funktionen von mehreren Zufallsvariablen . . . . .	4
Bedingte Verteilungen . . . . .	4
Wichtige diskrete Verteilungen . . . . .	4
Allgemeine Zufallsvariablen . . . . .	5
Grundbegriffe . . . . .	5
Übersicht . . . . .	5
Wichtige stetige Verteilungen . . . . .	6
Gemeinsame Verteilungen und unabhängige Zufallsvariablen . . . . .	6
Funktionen und Transformationen von Zufallsvariablen . . . . .	6
Ungleichungen und Grenzwertsätze . . . . .	6
Ungleichungen . . . . .	6
Das Gesetz der grossen Zahlen . . . . .	7
Der Zentrale Grenzwertsatz . . . . .	7
Grosse Abweichungen und Chernoff-Schranken . . . . .	7
Schätzer . . . . .	7
Grundbegriffe . . . . .	7
Die Maximum-Likelihood-Methode . . . . .	8
Verteilungsaussagen . . . . .	9
Tests . . . . .	9
Grundbegriffe . . . . .	9
Konstruktion von Tests . . . . .	9
Konfidenzbereiche . . . . .	10
Konfidenzintervall . . . . .	10
Kombinatorik kurz und knapp . . . . .	10
Anhang . . . . .	11
Integration . . . . .	11

# Wahrscheinlichkeiten

## Grundbegriffe

**Ereignisraum**  $\Omega$ : Menge aller möglichen elementaren Ereignissen.

*Beispiel:* Bei einem Würfelwurf sind die Elementarereignisse  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

**Potenzmenge**  $\mathcal{P}(\Omega)$  oder  $2^\Omega$ : Menge aller Teilmengen von  $\Omega$ .

**Klasse aller beobachtbaren Ereignisse**  $\mathcal{F}$ :  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  und  $\mathcal{F}$  ist eine  $\sigma$ -Algebra. Bei diskreten, d.h. endlichen bzw. abzählbaren Wahrscheinlichkeitsräumen wird  $\mathcal{F} = \mathcal{P}(\Omega)$  gewählt.

$\sigma$ -Algebra:  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  ist eine  $\sigma$ -Algebra, wenn gilt:

1.  $\Omega \in \mathcal{F}$
2.  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3.  $(A_n)_{n \in \mathbb{N}}, A_n \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

*Beispiel:* Jemand wirft einen Würfel und teilt uns mit, ob die gewürfelte Zahl gerade oder ungerade ist.

Wir könnten den Grundraum  $\Omega_1 = \{G, U\}$  wählen mit  $G$  für gerade und  $U$  für ungerade. In diesem Fall wäre  $\mathcal{F} = \{\emptyset, \Omega_1, \{G\}, \{U\}\}$ .

Jedoch könnten wir auch den Grundraum  $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$  wählen. Dann wäre  $\mathcal{F} = \{\emptyset, \Omega_2, \{2, 4, 6\}, \{1, 3, 5\}\} \neq \mathcal{P}(\Omega_2)$ , da beispielsweise das prinzipielle Ereignis  $\{1\}$  nie beobachtbar ist.

**Wahrscheinlichkeitsmass**  $P : \mathcal{F} \rightarrow [0, 1]$ :  $P[A] \in \mathcal{F} \in [0, 1]$  ist die Wahrscheinlichkeit, dass  $A$  eintritt. Dabei muss gelten:

1.  $\forall A \in \mathcal{F} : P[A] \geq 0$
2.  $P[\Omega] = 1$
3.  $P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$ , sofern die  $A_i \in \mathcal{F}$  paarweise disjunkt sind.

Folgende Rechenregeln lassen sich herleiten:

1.  $P[A^c] = 1 - P[A]$
2.  $P[\emptyset] = 0$
3. Für  $A \subseteq B$  gilt  $P[A] \leq P[B]$
4. Additionsregel:  $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

## Diskrete Wahrscheinlichkeitsräume

Bei diskreten, d.h. endlichen bzw. abzählbaren Wahrscheinlichkeitsräumen gilt  $\mathcal{F} = \mathcal{P}(\Omega)$  und  $P[A] = \sum_{w_i \in A} P[\{w_i\}]$ .

**Laplace-Raum**  $\Omega$  ist endlich und alle Elementarereignisse  $\Omega = \{w_1, \dots, w_N\}$  sind gleich wahrscheinlich mit  $p_1 = \dots = p_N = \frac{1}{N}$ .

**Diskrete Gleichverteilung** In einem Laplace-Raum gilt für beliebige  $A \subseteq \Omega$ :  $P[A] = \frac{|A|}{|\Omega|}$ .

*Beispiel:* Beim zweimaligen Münzwurf ist  $\Omega = \{KK, KZ, ZK, ZZ\}$ , also  $|\Omega| = 4$  und damit  $p_i = \frac{1}{4}$ . Dann ist  $P[\text{Mindestens einmal Kopf}] = P[\{KK, KZ, ZK\}] = \frac{3}{4}$ .

## Bedingte Wahrscheinlichkeiten

**Bedingte Wahrscheinlichkeit** Wahrscheinlichkeit, dass  $B$  eintritt, unter der Bedingung, dass  $A$  eintritt:

$$P[B \mid A] := \frac{P[B \cap A]}{P[A]} = \frac{P[A \mid B]P[B]}{P[A]}$$

**Satz von der totalen Wahrscheinlichkeit (Satz 1.1)** Sei  $A_1, \dots, A_n$  eine Zerlegung von  $\Omega$  in paarweise disjunkte Ereignisse, dann gilt für beliebige Ereignisse  $B$ :

$$P[B] = \sum_{i=1}^n P[B \mid A_i] P[A_i]$$

**Formel von Bayes (Satz 1.2)** Ist  $A_1, \dots, A_n$  eine Zerlegung von  $\Omega$  mit  $P[A_i] > 0$  und  $B$  ein Ereignis mit  $P[B] > 0$ , so gilt für jedes  $k$ :

$$P[A_k \mid B] = \frac{P[B \mid A_k] P[A_k]}{\sum_{i=1}^n P[B \mid A_i] P[A_i]}$$

## Unabhängigkeit

**Stochastische Unabhängigkeit** Zwei Ereignisse  $A, B$  heissen stochastisch unabhängig, falls  $P[A \cap B] = P[A]P[B]$ .

Allgemeiner: Zwei Ereignisse  $A, B$  heissen stochastisch unabhängig, wenn für jede endliche Teilfamilie  $\{k_1, \dots, k_m\} \subseteq \{1, \dots, n\}$  gilt, dass  $P\left[\bigcap_{i=1}^m A_{k_i}\right] = \prod_{i=1}^m P[A_{k_i}]$ .

Ist  $P[A] = 0$  oder  $P[B] = 0$ , so sind  $A$  und  $B$  immer unabhängig.

Für  $P[A] \neq 0$  gilt:  $A, B$  unabhängig  $\iff P[B \mid A] = P[B]$ .

## Diskrete Zufallsvariablen und Verteilungen

### Grundbegriffe

**Diskrete Zufallsvariable** Funktion  $X : \Omega \rightarrow \mathbb{R}$ ,  $W(X)$ : Wertebereich von  $X$ .

**Indikatorfunktion** Für jede Teilmenge  $A \subset \Omega$  gilt:  $I_A(w) := \begin{cases} 1 & w \in A \\ 0 & w \in A^c \end{cases}$

**Verteilungsfunktion**  $F_x : \mathbb{R} \rightarrow [0, 1]$ ,  $F_X(t) := P[X \leq t] = P[\{w \mid X(w) \leq t\}]$ .

**Gewichtsfunktion**  $p_x : W(X) \rightarrow [0, 1]$ ,  $p_X(x_k) := P[X = x_k] = P[\{w \mid X(w) = x_k\}]$ .

Es gilt  $F_X(t) = P[X \leq t] = \sum_{x_k \leq t} p_X(x_k)$

### Erwartungswerte

**Erwartungswert**  $E[X] := \sum_{x_k \in W(X)} x_k p_X(x_k)$ . Es gilt:

1. Monotonie: Ist  $X \leq Y$  (d.h.  $\forall w : X(w) \leq Y(w)$ ), so gilt auch  $E[X] \leq E[Y]$
2. Linearität:  $E[aX + b] = aE[X] + b$
3. Falls  $X \geq 0$ , so gilt  $E[X] = \sum_{j=1}^{\infty} P[X \geq j]$

**Varianz**  $\text{Var}[X] := E[(X - E[X])^2]$ . Es gilt:

1.  $\text{Var}[X] = E[X^2] - (E[X])^2$
2.  $\text{Var}[aX + b] = a^2 \text{Var}[X]$
3. Für unabhängige  $X, Y$  gilt:  $\text{Var}[aX + bY] = a^2 \text{Var}[X] + b^2 \text{Var}[Y]$
4. Für abhängige  $X, Y$  gilt:  $\text{Var}[aX + bY] = a^2 \text{Var}[X] + b^2 \text{Var}[Y] + 2ab \text{Cov}[X, Y]$

**Standardabweichung**  $\sigma(X) = \sqrt{\text{Var}[X]}$ .

## Gemeinsame Verteilungen und unabhängige Zufallsvariablen

**Gemeinsame Verteilungsfunktion**  $F: \mathbb{R}^n \rightarrow [0, 1]$ ,  $F(x_1, \dots, x_n) := P[X_1 \leq x_1, \dots, X_n \leq x_n]$

**Gemeinsame Gewichtsfunktion**  $p: \mathbb{R}^n \rightarrow [0, 1]$ ,  $p(x_1, \dots, x_n) := P[X_1 = x_1, \dots, X_n = x_n]$

$$F(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n] = \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} P[X_1 = y_1, \dots, X_n = y_n] = \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} p(y_1, \dots, y_n)$$

**Randverteilung** Verteilungsfunktion der Randverteilung von  $X$ :  $F_X(x) := P[X \leq x] = P[X \leq x, Y < \infty] = \lim_{y \rightarrow \infty} F(x, y)$

$$\text{Gewichtsfunktion der Randverteilung von } X: p_X(x) := P[X = x] = \sum_{y_j \in W(Y)} P[X = x, Y = y_j] = \sum_{y_j \in W(Y)} p(x, y_j)$$

**Unabhängigkeit**  $X_1, \dots, X_n$  heißen unabhängig, falls  $F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$  beziehungsweise  $p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$

$X$  und  $Y$  sind unabhängig, wenn gilt  $E[XY] = E[X]E[Y]$ .

## Funktionen von mehreren Zufallsvariablen

**Linearität (Satz 2.4)** Seien  $X_1, \dots, X_n$  diskrete Zufallsvariablen mit endlichen Erwartungswerten. Sei  $Y = a + \sum_{l=1}^n b_l X_l$  mit Konstanten  $a, b_1, \dots, b_n$ . Dann gilt  $E[Y] = a + \sum_{l=1}^n b_l E[X_l]$ .

**Kovarianz**  $\text{Cov}(X_1, X_2) := E[X_1 X_2] - E[X_1]E[X_2]$

$$\text{Cov}(X, X) = \text{Var}[X]$$

**Unkorreliertheit**  $X_1$  und  $X_2$  sind unkorreliert, wenn gilt  $\text{Cov}(X_1, X_2) = 0$ .

Unabhängigkeit impliziert Unkorreliertheit, die andere Richtung folgt jedoch nicht.

**Produkte von Zufallsvariablen (Satz 2.5)** Seien  $X_1, \dots, X_n$  diskrete Zufallsvariablen mit endlichen Erwartungswerten. Falls  $X_1, \dots, X_n$  unabhängig sind, so gilt:  $E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E[X_i]$ . Ausserdem sind dann  $X_1, \dots, X_n$  unkorreliert und es gilt:  $\text{Var}\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n \text{Var}[X_i]$ .

## Bedingte Verteilungen

**Bedingte Gewichtsfunktion** Seien  $X$  und  $Y$  diskrete Zufallsvariablen mit gemeinsamer Gewichtsfunktion  $p(x, y)$ . Die bedingte Gewichtsfunktion von  $X$ , gegeben dass  $Y = y$ , ist definiert durch  $p_{X|Y}(x | y) := P[X = x | Y = y] = \frac{P[X=x, Y=y]}{P[Y=y]} = \frac{p(x, y)}{p_Y(y)}$  für  $p_Y(y) > 0$  und 0 sonst.

## Wichtige diskrete Verteilungen

Verteilung	Gewichtsfunktion $p_X(k) = P[X = k]$	Verteilungsfunktion $F_X(t) = P[X \leq t]$	Erwartungswert $E[X]$	Varianz $\text{Var}[X]$
Diskrete Gleichverteilung	$\frac{1}{n}$	$\frac{\lfloor t \rfloor}{n}$	$\frac{a+b}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Bernoulli-Verteilung $X \sim \text{Be}(p)$	$p^k(1-p)^{1-k}$	$1-p$	$p$	$p(1-p)$
Binomialverteilung $X \sim \text{Bin}(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$		$np$	$np(1-p)$

Verteilung	Gewichtsfunktion $p_X(k) = P[X = k]$	Verteilungsfunktion $F_X(t) = P(X \leq t)$	Erwartungswert $E[X]$	Varianz $\text{Var}[X]$
Geometrische Verteilung $X \sim \text{Geom}(p)$	$p(1-p)^{k-1}$		$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negativbinomiale Verteilung $X \sim \text{NB}(r, p)$	$\binom{k-1}{r-1} p^r (1-p)^{k-r}$		$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometrische Verteilung	$\frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$		$m \frac{r}{n}$	$m \frac{r}{n} (1 - \frac{r}{n}) \frac{n-m}{n-1}$
Poisson-Verteilung $X \sim \mathcal{P}(\lambda)$	$e^{-\lambda} \frac{\lambda^k}{k!}$		$\lambda$	$\lambda$

## Allgemeine Zufallsvariablen

### Grundbegriffe

Siehe Tabelle *Allgemeine Zufallsvariablen*.

### Übersicht

	Diskrete Zufallsvariablen	Allgemeine Zufallsvariablen
Verteilungsfunktion	$F_X : \mathbb{R} \rightarrow [0, 1]$ $F_X(t) := P[X \leq t] := P[\{w \mid X(w) \leq t\}] = \sum_{x_k \leq t} p_X(x_k)$	$F_X : \mathbb{R} \rightarrow [0, 1]$ $F_X(t) := P[X \leq t] := P[\{w \mid X(w) \leq t\}] = \int_{-\infty}^t f_X(s) ds$
Gemeinsame Verteilungsfunktion	$F : \mathbb{R}^n \rightarrow [0, 1]$ $F(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n] = \sum_{y_1 \leq x_1, \dots, y_n \leq x_n} p(y_1, \dots, y_n)$	$F : \mathbb{R}^n \rightarrow [0, 1]$ $F(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n] = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_n \dots dt_1$
<i>Monoton wachsend</i>	$\forall s \leq t : F_X(s) \leq F_X(t)$	Analog
<i>Rechtsstetig</i>	$\forall u > t, u \rightarrow t : F_X(u) \rightarrow F_X(t)$ $\lim_{t \rightarrow \infty} F_X(t) = 1, \lim_{t \rightarrow -\infty} F_X(t) = 0$	Analog
Verteilung	$\mu_X(B) := P[X \in B] = \sum_{x_k \in B} p_X(x_k)$	$\mu_X(B) := P[X \in B] = \int_B f_X(s) ds$
Randverteilung	$F_X(x) := P[X \leq x] = P[X \leq x, Y \leq \infty] = \lim_{y \rightarrow \infty} F(x, y)$	Analog
Gewichtsfunktion, Dichtefunktion	$p_X(x_k) := P[X = x_k] = P[\{w \mid X(w) = x_k\}]$	$f_X(t) = \frac{d}{dt} F_X(t)$ $f_X \geq 0, f_X = 0$ ausserhalb von $W(X)$ $\int_{-\infty}^{\infty} f_X(s) ds = 1$
Gemeinsame Gewichtsfunktion, Dichtefunktion	$p_X(x_1, \dots, x_n) := P[X_1 = x_1, \dots, X_n = x_n]$	$f(t_1, \dots, t_n)$ $f(x_1, \dots, x_n) \geq 0,$ $f(x_1, \dots, x_n) = 0$ ausserhalb von $W(X_1, \dots, X_n)$ $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_n \dots dx_1 = 1$
Bedingte Gewichtsfunktion	$p_{X Y}(x \mid y) := P[X = x \mid Y = y] = \frac{P[X=x, Y=y]}{P[Y=y]} = \frac{p(x, y)}{p_Y(y)}$	$f_{X Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$

	Diskrete Zufallsvariablen	Allgemeine Zufallsvariablen
Randdichte		$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy,$ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$
Unabhängigkeit	$p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$	$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$
Erwartungswert	$E[X] = \sum_{-\infty}^{\infty} xp(x),$ für nichtnegative ganzzahlige Zufallsvariablen: $E[X] = \sum_{i=1}^{\infty} P[X \geq i]$	$E[X] = \int_{-\infty}^{\infty} xf(x)dx$
Varianz	$\text{Var}(X) = E[X^2] - E[X]^2$	Analog

### Wichtige stetige Verteilungen

Verteilung	Dichtefunktion $f_X(t)$	Verteilungsfunktion $F_X(t)$	Erwartungswert $E[X]$	Varianz $\text{Var}[X]$
Gleichverteilung $X \sim \mathcal{U}(a, b)$	$\begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & \text{sonst} \end{cases}$	$\begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \leq t \leq b \\ 1 & t > b \end{cases}$	$\frac{a+b}{2}$	
Exponentialverteilung $X \sim \text{Exp}(\lambda)$	$\begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normalverteilung $X \sim \mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$		$\mu$	$\sigma^2$

**Standard-Normalverteilung** Für die Standard-Normalverteilung  $\mathcal{N}(0, 1)$  gilt  $F_X(t) = \Phi(t)$ .

Ist  $X \sim \mathcal{N}(\mu, \sigma^2)$ , so ist  $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$ .  $\Phi(t)$  ist tabelliert.

**Abhängige Zufallsvariablen (Satz 4.1)** Sei  $X$  eine Zufallsvariable und  $Y = g(X)$  eine weitere Zufallsvariable. Ist  $X$  stetig mit Dichte  $f_X(x)$ , so ist  $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$ .

### Gemeinsame Verteilungen und unabhängige Zufallsvariablen

Siehe Tabelle *Allgemeine Zufallsvariablen*.

### Funktionen und Transformationen von Zufallsvariablen

Siehe Tabelle *Allgemeine Zufallsvariablen*.

### Ungleichungen und Grenzwertsätze

#### Ungleichungen

**Markov-Ungleichung** Sei  $X$  eine Zufallsvariable und  $g : W(X) \rightarrow [0, \infty)$  eine wachsende Funktion. Dann gilt für jedes  $c \in \mathbb{R}$  mit  $g(c) > 0$ :

$$P[X \geq c] \leq \frac{E[g(X)]}{g(c)}$$

**Chebyshev-Ungleichung** Sei  $Y$  eine Zufallsvariable mit endlicher Varianz. Für jedes  $b > 0$  gilt dann:

$$P[|Y - E[Y]| \geq b] \leq \frac{\text{Var}[Y]}{b^2}$$

Oder äquivalent dazu:

$$P[|Y - E[Y]| < b] \geq 1 - \frac{\text{Var}[Y]}{b^2}$$

## Das Gesetz der grossen Zahlen

**Schwaches Gesetz der grossen Zahlen** Sei  $X_1, X_2, \dots$  eine Folge von unabhängigen Zufallsvariablen mit dem gleichen Erwartungswert  $E[X_i] = \mu$  und der gleichen Varianz  $\text{Var}[X_i] = \sigma^2$ . Sei  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Dann konvergiert  $\bar{X}_n$  für  $n \rightarrow \infty$  stochastisch gegen  $\mu = E[X_i]$ , d.h.:

$$\boxed{\forall \epsilon > 0 : P[|\bar{X}_n - \mu| > \epsilon] \xrightarrow{n \rightarrow \infty} 0}$$

*Beweis mit Chebyshev-Ungleichung.*

**Starkes Gesetz der grossen Zahlen** Sei  $X_1, X_2, \dots$  eine Folge von unabhängigen Zufallsvariablen mit dem gleichen **endlichen** Erwartungswert  $E[X_i] = \mu$  und der gleichen Verteilung. Sei  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Dann konvergiert  $\bar{X}_n$  für  $n \rightarrow \infty$  fastsicher gegen  $\mu = E[X_i]$ , d.h.:

$$\boxed{P[\{\omega \in \Omega \mid \bar{X}_n(\omega) \xrightarrow{n \rightarrow \infty} \mu\}] = 1}$$

## Der Zentrale Grenzwertsatz

**Zentraler Grenzwertsatz (Satz 5.5)** Sei  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen mit  $E[X_i] = \mu$  und  $\text{Var}[X_i] = \sigma^2$ . Für die Summe  $S_n = \sum_{i=1}^n X_i$  gilt dann:

$$\boxed{\lim_{n \rightarrow \infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] = \Phi(x) \quad \forall x \in \mathbb{R}}$$

$S_n$  hat den Erwartungswert  $E[S_n] = n\mu$  und Varianz  $\text{Var}[S_n] = n\sigma^2$ . Also ist  $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - E[S_n]}{\sqrt{\text{Var}[S_n]}}$  die Standardisierung von  $S_n$  mit  $E[S_n^*] = 0$  und  $\text{Var}[S_n^*] = 1$ . Deshalb gilt für grosse  $n$ :

$$\boxed{P[S_n^* \leq x] \approx \Phi(x) \quad S_n^* \sim \mathcal{N}(0, 1)}$$

## Grosse Abweichungen und Chernoff-Schranken

TODO: Ist dieses Kapitel Prüfungsrelevant?

**Momenterzeugende Funktion** Die momenterzeugende Funktion einer Zufallsvariable  $X$  ist  $M_X(t) := E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx$  für  $t \in \mathbb{R}$ .

(Satz 5.6)

## Schätzer

### Grundbegriffe

**Stichprobe** Gesamtheit der Beobachtungen  $x_1, \dots, x_n$  oder Zufallsvariablen  $X_1, \dots, X_n$ . Die Anzahl  $n$  heisst Stichprobenanzahl.

**Stochastischer Mechanismus**  $P_\vartheta$  ist ein konkreter stochastischer Mechanismus, der besagt, wie sich  $X_1, \dots, X_n$  verhalten. Dabei wird der Parameter  $\vartheta$  zu bestimmen versucht.

**Schätzer** Die Schätzer  $T_1, \dots, T_m$  schätzen die Parameter  $\vartheta_1, \dots, \vartheta_m$ . Sie sind Zufallsvariablen mit der Form  $T_j = t_j(X_1, \dots, X_n)$ .

**Schätzwert**  $T_j(\omega) = t_j(X_1(\omega), \dots, X_n(\omega))$  eines konkreten Experiments  $\omega$ .

**Erwartungstreueit** Ein Schätzer  $T$  heisst erwartungstreu für  $\vartheta \in \Theta$ , falls gilt  $E_\vartheta[T] = \vartheta$

**Bias**  $E_\vartheta[T] - \vartheta$

**Mittlerer quadratischer Schätzfehler**  $\text{MSE}_\vartheta[T] := E_\vartheta[(T - \vartheta)^2] = \text{Var}_\vartheta[T] + (E_\vartheta[T] - \vartheta)^2$

**Konsistenz** Eine Folge von Schätzern  $T^{(n)}$  heisst konsistent für  $\vartheta$ , falls für jedes  $\vartheta \in \Theta$  und jedes  $\varepsilon > 0$  gilt:  
 $\lim_{n \rightarrow \infty} P_\vartheta[|T^{(n)} - \vartheta| > \varepsilon] = 0$ . *Beweisen mit Chebyshev-Ungleichung.*

**Die Maximum-Likelihood-Methode**

**Likelihood-Funktion**  $L(x_1, \dots, x_n; \vartheta) := \begin{cases} p(x_1, \dots, x_n; \vartheta) = \prod_{i=1}^n p_X(x_i; \vartheta) & \text{Im diskreten Fall} \\ f(x_1, \dots, x_n; \vartheta) = \prod_{i=1}^n f_X(x_i; \vartheta) & \text{Im stetigen Fall} \end{cases}$

**log-Likelihood-Funktion**  $l(x_1, \dots, x_n; \vartheta) := \log L(x_1, \dots, x_n; \vartheta)$

**ML-Schätzer** Der ML-Schätzer  $T$  für  $\vartheta$  wird definiert als Maximierung von  $L(X_1, \dots, X_n; \vartheta)$  als Funktion von  $\vartheta$ .

Um den ML-Schätzer zu finden:

1. Bilde log-Likelihood-Funktion, da sie meistens einfacher ist zu maximieren
2. Bilde Ableitung  $\frac{\partial}{\partial \vartheta} \log L(x_1, \dots, x_n; \vartheta)$
3. ML-Schätzer  $\vartheta$  kann durch das Nullsetzen von  $\frac{\partial}{\partial \vartheta} \log L(x_1, \dots, x_n; \vartheta)$  gefunden werden

**Momentenschätzer** Der ML-Schätzer für  $\vartheta = (\mu, \sigma^2) = (E_\vartheta[X], \text{Var}_\vartheta[X])$ , genannt Momentenschätzer, ist:

$$T = (T_1, T_2) \quad T_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n \quad T_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - \bar{X}_n^2$$

Dieser Schätzer ist nicht erwartungstreu.

Um einen Erwartungstreuen Schätzer für  $(E_\vartheta[X], \text{Var}_\vartheta[X])$  zu haben, verwendet man meistens:

$$T = (T'_1, T'_2) \quad T'_1 = T_1 = \bar{X}_n \quad T'_2 = \frac{1}{n-1} T_2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{1}{n-1} \bar{X}_n^2$$

$T_1 = T'_1$  wird *Stichprobenmittel*,  $S^2 := T'_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  wird *Stichprobenvarianz* genannt.

Um den Momentenschätzer zu finden:

1. Berechne  $T$
2. Setze  $T = (E[X], \text{Var}[X])$ , wobei  $E[X] = \int_{-\infty}^{\infty} x f(x) dx$
3. Löse auf nach der zu schätzenden Variable  $\vartheta$



## Verteilungsaussagen

**Mehrere Normalverteilte Variablen (Satz 7.1)** Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ . Dann gilt:

1.  $\bar{X}_n$  ist normalverteilt gemäss  $\sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$ , und  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$ .
2.  $\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  ist  $\chi^2$ -verteilt mit  $n - 1$  Freiheitsgraden.
3.  $\bar{X}_n$  und  $S^2$  sind unabhängig.
4. Der Quotient  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  ist  $t$ -verteilt mit  $n - 1$  Freiheitsgraden.

## Tests

### Grundbegriffe

**Hypothese**  $H_0 : \vartheta \in \Theta_0$

**Niveau** Je kleiner das Niveau  $\alpha$ , desto kleiner ist die Wahrscheinlichkeit, dass die Nullhypothese  $H_0$  abgelehnt wird, obwohl sie richtig ist.

**Alternative**  $H_A : \vartheta \in \Theta_A$ , wobei  $\Theta_A = \Theta_0^c$ , falls keine explizite Alternative spezifiziert wurde.

**Realisierung**  $\tilde{K} = \{T \in K\} = \{\omega \mid T(\omega) \in K\}$  ist eine Teilmenge von  $\Omega$  mit  $P_\vartheta[\tilde{K}] = P_\vartheta[T \in K]$ .

**Fehler erster Art**  $\vartheta \in \Theta_0$  und  $T \in K$ .

**Fehler zweiter Art**  $\vartheta \in \Theta_A$  und  $T \notin K$ .

### Vorgehen

1. *Minimierung des Fehlers erster Art.* Wähle Signifikanzniveau  $\alpha \in (0, 1)$  so dass  $\sup_{\vartheta \in \Theta_0} P_\vartheta[T \in K] \leq \alpha$ .
2. *Minimierung des Fehlers zweiter Art.* Maximiere die Macht des Tests  $\beta(\vartheta) := P_\vartheta[T \in K]$ .

### Konstruktion von Tests

**Likelihood-Quotient**  $R(x_1, \dots, x_n; \vartheta_0, \vartheta_A) := \frac{L(x_1, \dots, x_n; \vartheta_0)}{L(x_1, \dots, x_n; \vartheta_A)}$

Ist der Likelihood-Quotient klein, sprechen die Daten gegen die Nullhypothese und für die Alternativhypothese.

**Neyman-Pearson-Lemma (Satz 9.1)** Sei  $\Theta_0 = \{\vartheta_0\}$  und  $\Theta_A = \{\vartheta_A\}$ . Sei  $T = R(X_1, \dots, X_n, \vartheta_0, \vartheta_A)$  und  $K = [0, c]$  sowie  $\alpha^* := P_{\vartheta_0}[T \in K] = P_{\vartheta_0}[T < c]$ . Der Likelihood-Quotienten-Test mit  $T$  und  $K$  ist dann optimal im Sinn dass jeder andere Test mit Signifikanzniveau  $\alpha \leq \alpha^*$  eine kleinere Macht hat.

Beispiel zur Konstruktion von Tests:

Ist eine Münze gezinkt? Vermutung, dass für einen Wurf  $X_i$  gilt, dass  $p > 0$  ist. Test mit  $n = 10$ ,  $\alpha = 2.5\%$ . Die Resultate der Würfe sind  $6 \times X_i = 1$  und somit  $4 \times X_i = 0$ .

1. Modell:  $X_i \sim \text{Ber}(p)$
2. Nullhypothese:  $H_0 : p_0 = p = 0.5$
3. Alternativhypothese:  $H_A : p > p_0$
4. Teststatistik:  $T = \sum_{i=1}^n X_i$ , denn der Likelihood-Quotient kleiner als 1 und somit wird  $R(x_1, \dots, x_n; p_0, p_A)$  klein genau dann wenn  $T = \sum_{i=1}^n x_i$  gross ist.
5. Verteilung der Teststatistik unter der Nullhypothese:  $p_0 = p = 0.5 \implies T \sim \text{Bin}(10, 0.5)$
6. Verwerfungsbereich: Da gilt, dass die Summe  $T$  gross wird bei einem kleinen Likelihood-Quotient, hat der Verwerfungsbereich die Form  $K = (k, \infty)$ . Zudem muss  $P_{p_0}[T \leq k] \geq (1 - \alpha)$  gelten, woraus folgt, dass  $k = 8$ . Somit ist  $K = [8, \infty)$ .
7. Beobachteter Wert der Teststatistik:  $T(\omega) = 6$
8. Testentscheid: Da  $T(\omega) = 6 \notin K = [8, \infty)$ , wird die Nullhypothese nicht verworfen.

**z-Test** Seien  $X_1, \dots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$  Normalverteilt mit bekannter Varianz  $\sigma^2$  und unbekanntem Erwartungswert  $\vartheta$ . Wir wollen die Hypothese  $H_0 : \vartheta = \vartheta_0$  testen.

Die Teststatistik ist  $T := \frac{\bar{X}_n - \vartheta_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ .

Der kritische Bereich ist von der folgenden Form mit Niveau  $\alpha$ :

- $H_A : \vartheta > \vartheta_0$  mit  $K_> = (c_>, \infty)$  und  $c_> = \Phi^{-1}(1 - \alpha) := z_{1-\alpha}$  wegen der Bedingung  $\alpha = P_{\vartheta_0}[T \in K_>] = P_{\vartheta_0}[T > c_>] = 1 - P_{\vartheta_0}[T \leq c_>] = 1 - \Phi(c_>)$ .
- $H_A : \vartheta < \vartheta_0$  mit  $K_< = (-\infty, c_<)$  und  $c_< = -\Phi^{-1}(1 - \alpha) := -z_{1-\alpha}$  wegen der Bedingung  $\alpha = P_{\vartheta_0}[T \in K_<] = P_{\vartheta_0}[T < c_<] = \Phi(c_<)$ .
- $H_A : \vartheta \neq \vartheta_0$  mit  $K_\neq = (-\infty, c_\neq) \cup (c_\neq, \infty)$  und  $c_\neq = \Phi^{-1}(1 - \frac{\alpha}{2}) := z_{1-\frac{\alpha}{2}}$  wegen der Bedingung  $\alpha = P_{\vartheta_0}[T \in K_\neq] = P_{\vartheta_0}[T < -c_\neq] + P_{\vartheta_0}[T > c_\neq] = \Phi(-c_\neq) + 1 - \Phi(c_\neq) = 2(1 - \Phi(c_\neq))$ .

Die Nullhypothese wird verworfen, wenn für die Realisierung der Teststatistik gilt  $t \in H_a$ .

**t-Test** Seien  $X_1, \dots, X_n \sim \mathcal{N}(\vartheta, \sigma^2)$  Normalverteilt mit unbekannter Varianz und unbekanntem Erwartungswert  $\vartheta = (\mu, \sigma^2)$ . Wir wollen die Hypothese  $H_0 : \mu = \mu_0$  testen.

Die Teststatistik ist  $T := \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ .

Wir ersetzen die unbekannte Varianz durch den Schätzer  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ .

Der kritische Bereich ist von der folgenden Form mit Niveau  $\alpha$ :

- $H_A : \mu > \mu_0$  mit  $K_> = (c_>, \infty)$  und  $c_> = t_{n-1, 1-\alpha}$ .
- $H_A : \mu < \mu_0$  mit  $K_< = (-\infty, c_<)$  und  $c_< = t_{n-1, \alpha}$ .
- $H_A : \mu \neq \mu_0$  mit  $K_\neq = (-\infty, c_\neq) \cup (c_\neq, \infty)$  und  $c_\neq = t_{n-1, 1-\frac{\alpha}{2}}$ .

Die Nullhypothese wird verworfen, wenn für die Realisierung der Teststatistik gilt  $t \in H_a$ .

## Gepaarter Zweistichproben-Test

## Ungepaarter Zweistichproben-Test

## Konfidenzbereiche

### Konfidenzintervall

$C(X_1, \dots, X_n)$  heisst der Konfidenzbereich zum Niveau  $1 - \alpha$ , falls gilt  $P_\vartheta[\vartheta \in C(X_1, \dots, X_n)] \geq 1 - \alpha$ .

### Konfidenzintervall des Erwartungswerts

$$C(X_1, \dots, X_n) = [\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$$

### Konfidenzintervall der Varianz

$$C(X_1, \dots, X_n) = \left[ \frac{(n-1)S^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$$

## Kombinatorik kurz und knapp

Name	Fragestellung	Antwort
Anzahl Permutationen ohne Wiederholung von $n$ Elementen.	Auf wieviele Arten kann man $n$ Objekte nebeneinander anordnen?	$n! = n \times (n-1) \times \dots \times 2 \times 1$
Anzahl Kombinationen ohne Wiederholung.	Auf wie viele Arten kann man $k$ aus den $n$ Objekten auswählen ohne Zurücklegen?	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$
Anzahl der Variationen mit Wiederholung.	Wie viele Sequenzen der Länge $m$ kann man mit den $n$ Symbolen bilden?	$n^m$

## Anhang

### Integration

**Integration durch Substitution**  $x = \varphi(u)$

$$\int_{\varphi(a)}^{\varphi(b)} f(x) dx = \int_a^b f(\varphi(u)) \varphi'(u) du$$

**Partielle Integration**  $\int_a^b f'(x) g(x) dx = [f(x) g(x)]_a^b - \int_a^b f(x) g'(x) dx$