

# Statistik - SoSe 2025

## Inhaltsverzeichnis

<b>I</b>	<b>Wahrscheinlichkeitsrechnung</b>	<b>2</b>
<b>1</b>	<b>Grundbegriffe</b>	<b>2</b>
1.1	Ergebnismenge und Ereignisse . . . . .	2
1.1.1	Rechenregeln für Mengen . . . . .	4
1.2	Wahrscheinlichkeitsmaße . . . . .	5
1.2.1	Rechenregeln für Wahrscheinlichkeitsmaße . . . . .	7
<b>2</b>	<b>Bedingte Wahrscheinlichkeiten und Unabhängigkeit</b>	<b>9</b>
2.1	Bedingte Wahrscheinlichkeiten . . . . .	9
2.1.1	Multiplikationssatz für bedingte Wahrscheinlichkeiten . . . . .	10
2.1.2	Satz von der totalen Wahrscheinlichkeit . . . . .	10
2.1.3	Satz von Bayes . . . . .	11
2.2	Unabhängigkeit . . . . .	12
<b>3</b>	<b>Zufallsvariablen</b>	<b>14</b>
3.1	Zufallsvariablen . . . . .	14
3.2	Unabhängigkeit . . . . .	15
3.2.1	Satz 3.1: . . . . .	15
3.3	Verteilungsfunktionen, W'keitsfunktionen und Dichten . . . . .	16
3.3.1	Satz 3.2: Eigenschaften von VF . . . . .	17
3.3.2	Diskrete ZV . . . . .	18
3.3.3	Stetige ZV . . . . .	19
3.4	Beispiele diskreter Verteilungen . . . . .	19
3.4.1	Bernoulli-Verteilung . . . . .	20
3.4.2	Binomialverteilung . . . . .	20
3.4.3	Geometrische Verteilung . . . . .	21
3.4.4	Poisson-Verteilung . . . . .	22
3.4.5	Hypergeometrische Verteilung . . . . .	22
3.5	Beispiele stetiger Verteilungen . . . . .	23
3.5.1	Gleichverteilung . . . . .	23
3.5.2	Exponentialverteilung . . . . .	24
3.5.3	Normalverteilung . . . . .	24

<b>4</b>	<b>Erwartungswert, Varianz und Kovarianz</b>	<b>25</b>
4.1	Erwartungswert einer ZV . . . . .	25
4.1.1	Satz 4.1: . . . . .	26
4.1.2	Satz 4.2: . . . . .	27
4.1.3	Satz 4.3: . . . . .	28
4.2	Varianz einer ZV . . . . .	28
4.2.1	Satz 4.4: . . . . .	29
4.3	Kovarianz und Korrelation . . . . .	29
4.3.1	Satz 4.5: . . . . .	30
4.4	Ungleichungen . . . . .	31
4.4.1	Satz 4.6: Markov-Ungleichung . . . . .	31
4.4.2	Satz 4.7: Tschebyscheff-Ungleichung . . . . .	32
<b>5</b>	<b>Grenzwertsätze</b>	<b>33</b>
5.1	Gesetz der großen Zahlen . . . . .	33
5.1.1	Satz 5.1: Schwaches Gesetz der großen Zahlen . . . . .	34
5.1.2	Satz 5.2: Starkes Gesetz der großen Zahlen . . . . .	35
5.2	Zentraler Grenzwertsatz . . . . .	35
5.2.1	Satz 5.2 Zentraler Grenzwertsatz ZGS . . . . .	35
<b>II</b>	<b>Schließende Statistik</b>	<b>36</b>
<b>6</b>	<b>Parameterschätzung</b>	<b>36</b>
6.1	Erwartungstreue und Konsistenz . . . . .	36
6.1.1	Konsistenz . . . . .	38
6.2	Die Maximum-Likelihood-Methode . . . . .	39
<b>7</b>	<b>Konfidenzintervalle</b>	<b>42</b>
<b>8</b>	<b>Testverfahren</b>	<b>46</b>
8.1	Einleitung, Definitionen . . . . .	46
8.1.1	Satz 8.1 . . . . .	47
8.2	Testen von Hypothesen über den Mittelwert einer $N(\mu, \sigma^2)$ -Verteilung . . .	47
8.3	Der $\chi^2$ -Anpassungstest . . . . .	50

## Teil I

# Wahrscheinlichkeitsrechnung

## 1 Grundbegriffe

### 1.1 Ergebnismenge und Ereignisse

Alle möglichen Ergebnisse eines Zufallsexperiments fasst man zur Ergebnismenge  $\Omega$  (Groß Omega) zusammen.

**Bsp. 1.1:** Zweimaliges Werfen eines Würfels

$$\begin{aligned}\Omega &= \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\} \\ &= \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\} \\ &= \{1, \dots, 6\} \times \{1, \dots, 6\} = \{1, \dots, 6\}^2\end{aligned}$$

**Bsp. 1.2:** Ziehung der Lottozahlen (6 aus 49)

$$\Omega = \{(\omega_1, \dots, \omega_6) : \omega_i \in \{1, 2, \dots, 49\}, 1 \leq i \leq 6, \omega_1 < \omega_2 < \dots < \omega_6\}$$

**Bsp. 1.3:** Verkaufszahlen. Ein Laden erhält morgens 3 Tageszeitungen  $Z_1, Z_2, Z_3$ , und zwar 100 bzw. 200 bzw. 250 Stück. Die verkauften Anzahlen sind dann als Ergebnis eines Zufallsexperiments zu interpretieren.

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \omega_3) : \omega_1 \in \{0, 1, \dots, 100\}, \omega_2 \in \{0, 1, \dots, 200\}, \omega_3 \in \{0, 1, \dots, 250\}\} \\ &= \{0, 1, \dots, 100\} \times \{0, 1, \dots, 200\} \times \{0, 1, \dots, 250\}\end{aligned}$$

Ein Ereignis ist eine Teilmenge der Ergebnismenge. Ereignisse werden zunächst verbal beschrieben und lassen sich dann als Teilmengen von  $\Omega$  auffassen.

*In Bsp. 1.1:*

- Ereignis E: Augensumme ist 10; zugeordnete Teilmenge von  $\Omega : E = \{(4, 6), (5, 5), (6, 4)\}$
- E: Pasch;  $E = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
- E: mindestens eine 6;  $E = \{(6, 1), (6, 2), \dots, (6, 6), (5, 6), \dots, (1, 6)\}$

*In Bsp. 1.3:*

Ereignis A: Von jeder der 3 Zeitungen werden mindestens 50 Stück verkauft

Teilmenge:  $A = \{(\omega_1, \omega_2, \omega_3) \in \Omega : \omega_1 \leq 50, \omega_2 \leq 50, \omega_3 \leq 50\} = \{50, 51, \dots, 100\} \times \{50, 51, \dots, 200\} \times \{50, 51, \dots, 250\}$

Ist  $\omega \in \Omega$  das betrachtete Ergebnis eines Zufallsexperiments und ist  $E \subset \Omega$  ein Ereignis, so sagt man *E ist eingetreten*, falls  $\omega \in E$  und *E ist nicht eingetreten*, falls  $\omega \notin E$ .

Übertrage Begriffe der Mengenlehre auf das zufällige Eintreten von Ereignissen. (Dies wird durch die Zuordnung *Ereignis*  $\Leftrightarrow$  *Teilmenge von*  $\Omega$  möglich.)

Mengenschreibweise	Interpretation für Ereignisse
$A = \Omega$	A ist ein sicheres Ereignis (A tritt sicher ein)
$A = \emptyset$	A ist ein unmögliches Ereignis (A tritt nicht ein)
$A^c = \{\omega \in \Omega, \omega \notin A\}$ $\bar{A} = A^c$ (andere Schreibweise)	Komplementärereignis zu A (A tritt nicht ein)
$A \cap B$	A geschnitten B. (A und B treten ein.)
$\bigcap_{i=1}^n A_i = A_1 \cap \dots \cap A_n$	Durchschnitt der Ereignisse $A_1, \dots, A_n$ (Jedes dieser Ereignisse $A_1, \dots, A_n$ tritt ein.)
$A \cap B = \emptyset$	A und B sind disjunkt. (A und B treten sicher nicht zusammen ein.)
$A \cup B$	Vereinigung von A und B. (A oder B tritt ein. „Oder“ ist nicht ausschließend.)
$\bigcup_{i=1}^n A_i = A_1 \cup \dots \cup A_n$	Mindestens eines der Ereignisse tritt ein.
$A \subset B$	Implikation: Aus dem Eintreten von A folgt das Eintreten von B.
$A \setminus B = A \cap B^c$	A tritt ein, aber B tritt nicht ein.

Einelementige Teilmengen  $\{\omega\} \subset \Omega$  (die nur einen Punkt enthalten) nennt man Elementarereignisse.

**Bsp. 1.4:** Einmaliges Werfen eines Würfels;  $\Omega = \{1, \dots, 6\}$

Ereignis A: „Augenzahl ist gerade“  $\Rightarrow A = \{2, 4, 6\}$ , Ereignis B: „Augenzahl ist mindestens 4“  $\Rightarrow B = \{4, 5, 6\}$

Schreibweise	Interpretation
$A^c = \{1, 3, 5\}$	Augenzahl ist ungerade
$B^c = \{1, 2, 3\}$	Augenzahl ist kleiner als 4
$A \cap B = \{4, 6\}$	Augenzahl ist gerade und mind. 4
$A \cup B = \{2, 4, 5, 6\}$	Augenzahl ist gerade oder mind. 4
$B \setminus A = B \cap A^c = \{4, 5, 6\} \cap \{1, 3, 5\} = \{5\}$	Augenzahl ist mind. 4, aber nicht gerade.
$A \setminus B = A \cap B^c = \{1, 3, 5\} \cap \{4, 5, 6\} = \{1, 3\}$	Augenzahl ist nicht gerade, aber mind. 4

### 1.1.1 Rechenregeln für Mengen

Für beliebige Mengen A,B,C gilt:

- Kommutativgesetze:  $A \cup B = B \cup A$ ,  $A \cap B = B \cap A$
- Assoziativgesetze:  $A \cup (B \cup C) = (A \cup B) \cup C$ ;  $A \cap (B \cap C) = (A \cap B) \cap C$
- Distributivgesetze:  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ;  $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- Komplement:  $(A^c)^c = A$
- De Morgansche Regeln:  $(A \cup B)^c = A^c \cap B^c$ ,  $(A \cap B)^c = A^c \cup B^c$

Hinweis: Um das Komplement verwenden zu können benötigt man eine bestimmte Obermenge.

Allgemeiner: Für beliebige  $A_1, \dots, A_n$  und B gilt:

- Kommutativ- und Assoziativgesetze gelten, da die Reihenfolge keine Rolle spielt.
- $(\bigcup_{i=1}^n A_i) \cap B = \bigcup_{i=1}^n (A_i \cap B)$

- $(\bigcap_{i=1}^n A_i) \cup B = \bigcap_{i=1}^n (A_i \cup B)$
- $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c, (\bigcap_{i=1}^n A_i)^c = \bigcup_{i=1}^n A_i^c$
- Für beliebige Mengen A,B gilt  $A \subset B \Leftrightarrow A^c \supset B^c$

Beweis für  $(\bigcup_{i=1}^n A_i)^c = \bigcap_{i=1}^n A_i^c$ , wobei die Komplemente bzgl.  $\Omega$  gebildet werden.  
Für jedes  $\omega \in \Omega$  gilt:

$$\begin{aligned} \omega \in (\bigcup_{i=1}^n A_i)^c &\Leftrightarrow \omega \notin \bigcup_{i=1}^n A_i \\ &\Leftrightarrow \omega \text{ ist in keiner der Mengen } A_1, \dots, A_n \text{ enthalten.} \\ &\Leftrightarrow \omega \in A_1^c \text{ und } \omega \in A_2^c \text{ und } \dots \text{ und } \omega \in A_n^c \\ &\Leftrightarrow \omega \in \bigcap_{i=1}^n A_i^c \end{aligned}$$

## 1.2 Wahrscheinlichkeitsmaße

**Def. 1.1:** Es sei  $\Omega \neq \emptyset$  einer Ergebnismenge. Eine für alle Ereignisse  $A \subset \Omega$  definierte reellwertige Funktion P heißt **Wahrscheinlichkeitsmaß** auf  $\Omega$  falls sie die folgenden Axiome der Wahrscheinlichkeitsrechnung erfüllt:

- (i)  $P(A) \geq 0$  für alle Ereignisse  $A \subset \Omega$
- (ii)  $P(\Omega) = 1$
- (iii)  $\sigma$ -Additivität:  $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$  gilt für jede endliche oder unendliche Folge von paarweise disjunkten Ereignissen  $A_1, A_2, \dots \subset \Omega$

$P(A)$  heißt *Wahrscheinlichkeit* von A (probability).  $(\Omega, P)$  heißt *Wahrscheinlichkeitsmodell* für das Zufallsexperiment. (i)-(iii) heißen *Axiome der Wahrscheinlichkeitsrechnung* oder *Kolmogorov-Axiome*.

**Bsp. 1.5:** Zweimaliges Werfen eines Würfels,  $\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$

Für  $A \subset \Omega$  sei

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{36},$$

wobei  $|A|$  = Anzahl der Elemente von A. Dann ist P ein W-Maß, denn P ist reellwertig

- (i)  $P(A) \geq 0$  für alle  $A \subset \Omega$
- (ii)  $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$
- (iii) Sind  $A_1, \dots, A_n$  paarweise disjunkte Teilmengen von  $\Omega$ , so ist

$$|\bigcup_{i=1}^n A_i| = \sum_{i=1}^n |A_i|,$$

also

$$P(\bigcup_{i=1}^n A_i) = \frac{|\bigcup_{i=1}^n A_i|}{|\Omega|} = \frac{\sum_{i=1}^n |A_i|}{|\Omega|} = \sum_{i=1}^n \frac{|A_i|}{|\Omega|} = \sum_{i=1}^n P(A_i).$$

Für jede unendliche Folge von paarweise disjunkten Teilmengen  $A_1, \dots, A_n$  von  $\Omega$  existiert ein  $N \in \mathbb{N}$ , so dass  $A_i = \emptyset$  für alle  $i > N$ , da  $\Omega$  endlich ist. Es gilt also:

$$\begin{aligned} \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^N A_i\right) \\ \Rightarrow \sum_{i=1}^N P(A_i) &= \sum_{i=1}^{\infty} P(A_i) \\ P(A_i) &= P(\emptyset) = \frac{|\emptyset|}{|\Omega|} = 0 \text{ für alle } i > N \end{aligned}$$

Für das Ergebnis A: „Augensumme ist 10“, also  $A = \{(4, 6), (5, 5), (6, 4)\}$  ergibt sich

$$P(A_i) = \frac{|A|}{|\Omega|} = \frac{3}{36} = \frac{1}{12}.$$

Für das Ergebnis B: „Pasch“, also  $B = \{(1, 1), (2, 2), \dots, (6, 6)\}$  ergibt sich

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

**Laplace-Wahrscheinlichkeit:** Ist allgemeiner  $\Omega \neq \emptyset$  eine beliebige endliche Menge und haben alle Ergebnisse (genauer: alle Elementarereignisse) dieselbe Wahrscheinlichkeit, dann folgt aus Axiomen (ii) und (iii)

$$P(\{w\}) = \frac{1}{|\Omega|}$$

und daher wegen Axiom (iii)

$$P(A) = \frac{|A|}{|\Omega|}$$

Also

$$P(A) = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl aller möglichen Ergebnisse}}$$

$P(A)$  heißt *Laplace-Wahrscheinlichkeit* von A.

**Bsp. 1.6** Roulette. Wenn  $\Omega = \{0, 1, \dots, 36\}$  ist die W'keit, dass eine ungerade Zahl fällt,

$$P(\{1, 3, 5, \dots, 35\}) = \frac{|\{1, 3, 5, \dots, 35\}|}{|\Omega|} = \frac{18}{37}.$$

....Grundlegendes Prinzip des Zählens ... was heißt das?

**Multiplikationsregel:** Betrachte k Aufgaben. Nimm an, die 1. Aufgabe kann auf  $n_1$  Arten erledigt werden; danach kann die 2. Aufgabe auf  $n_2$  Arten erledigt werden; ...; danach kann die k-te Aufgabe erledigt werden. (Annahme:  $n_1, n_2, \dots, n_k$  sind unabhängig.) Die Anzahl der Möglichkeiten, alle k Aufgaben zu erledigen, ist  $n_1 \cdot n_2 \cdot \dots \cdot n_k$

**Bsp. 1.7:** Eine faire Münze wird n-mal geworfen. Ergebnisse sind Folgen von Kopf(K) und Zahl(Z) der Länge n:  $\Omega = \{(x_1, \dots, x_n) : x_i \in \{K, Z\}, i = 1, \dots, n\}$ . Verwende Laplace-Wahrscheinlichkeiten:  $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{2^n}$  für alle  $A \in \Omega$ . Wie groß ist die Wahrscheinlichkeit von  $A_j$ : „K fällt im j-ten Wurf zum 1. Mal“?

$$A_j = \{(x_1, \dots, x_n) : x_i = Z \text{ für } 1 \leq i \leq j-1; x_j = K; x_i \in \{K, Z\} \text{ für } j+1 \leq i \leq n\}$$

$$|A_j| = 1^{j-1} \cdot 1 \cdot 2^{n-j} = 2^{n-j} \rightarrow P(A_j) = \frac{|A_j|}{|\Omega|} = \frac{2^{n-j}}{2^n}$$

**Bsp. 1.8:** Geburtstagsproblem. W'keit, dass von k Personen 2 oder mehr am selben Tag im Jahr Geburtstag haben?  $2 \leq k \leq 365$  ohne (29.2) und verwende Laplace-W'keiten mit  $\Omega = \{(x_1, \dots, x_k) : x_i \in \{1, \dots, 365\}, i = 1, \dots, k\}$ , also  $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{365^k}$  für alle  $A \subset \Omega$ . Sei  $B = \{(x_1, \dots, x_k) \in \Omega : \text{mind. 2 der } x_i \text{ stimmen überein}\}$

$$\begin{aligned} |B^c| &= \{(x_1, \dots, x_k) \in \Omega : x_i \neq x_j \text{ für } i \neq j\} \\ &= 365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - k + 1) \\ &= \prod_{i=0}^{k-1} (365 - i) \\ \rightarrow P_k &= P(B) = \frac{|\Omega| - |B^c|}{|\Omega|} = 1 - \frac{1}{365^k} \cdot \prod_{i=0}^{k-1} (365 - i) \\ &= 1 - \prod_{i=0}^{k-1} \left(1 - \frac{i}{365}\right) \text{ z.B. } P_{30} = 0,705 \end{aligned}$$

### 1.2.1 Rechenregeln für Wahrscheinlichkeitsmaße

$$1.1 \quad \mathbf{P(A^c)} = \mathbf{1 - P(A)}, \text{ denn } 1 = P(\Omega) = P(A \cup A^c) \stackrel{(iii)}{=} P(A) + P(A^c) \Rightarrow (1.1)$$

$$1.2 \quad \mathbf{P(\emptyset)} = \mathbf{0}, \text{ denn } P(\emptyset) = P(\Omega^c) \stackrel{(1.1)}{=} 1 - P(\Omega) = 1 - 1 = 0$$

$$1.3 \quad \text{Falls } \mathbf{A \subset B}, \text{ dann } \mathbf{P(B \setminus A) = P(B) - P(A)}, \text{ denn:}$$

Falls  $A \subset B$ , dann

$$\begin{aligned} B &= B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c) \\ &\stackrel{A \subset B}{=} A \cup (B \cap A^c) \\ \Rightarrow P(B) &\stackrel{(iii)}{=} P(A) + P(B \cap A^c) = P(A) + P(B \setminus A) \Rightarrow (1.3) \end{aligned}$$

**Beachte!:**  $P(B \setminus A) = P(B) - P(A)$  gilt im Allgemeinen nicht ohne die Voraussetzung  $A \subset B$ . Bsp.:  $A = \Omega, B = \emptyset$

$$\begin{aligned} P(B \setminus A) &= P(\emptyset \cap \Omega^c) = P(\emptyset) \stackrel{(1.2)}{=} 0 \\ P(B) - P(A) &= P(\emptyset) - P(\Omega) = 0 - 1 = -1 \neq 0 \end{aligned}$$

$$1.4 \quad \text{Falls } \mathbf{A \subset B}, \text{ dann } \mathbf{P(A) \leq P(B)} \text{ (*Monotonie*)}, \text{ denn:}$$

$$\text{Falls } A \subset B, \text{ dann } P(B) - P(A) \stackrel{(1.3)}{=} P(B \setminus A) \stackrel{(i)}{\geq} 0$$

$$1.5 \quad \text{Verallgemeinerung von (1.3): } \mathbf{P(B \setminus A) = P(B) - P(A \cap B)}, \text{ denn:}$$

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap (A \cup A^c)) \\ &= P((B \cap A) \cup (B \cap A^c)) \\ &= P(A \cap B) + P(B \cap A^c) \\ &= P(A \cap B) + P(B \setminus A) \\ &\Rightarrow (1.5) \end{aligned}$$

1.6  $\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$ , denn:

$$\begin{aligned}(A \cup B) \setminus B &= (A \cup B) \cap B^c \\ &= (A \cup B) \cap B^c \\ &= (A \cap B^c) \cup (B \cap B^c) \\ &= A \setminus B\end{aligned}$$

und

$$\begin{aligned}A \setminus (A \cap B) &= A \cap (A \cap B)^c \\ &= A \cap (A^c \cup B^c) \\ &= (A \cap A^c) \cup (A \cap B^c) \\ &= A \setminus B.\end{aligned}$$

Daher ist

$$\begin{aligned}(A \cup B) \setminus B &= A \setminus (A \cap B) \\ \Rightarrow P((A \cup B) \setminus B) &= P(A \setminus (A \cap B)) \\ \stackrel{(1.3)}{\Rightarrow} P(A \cup B) - P(B) &= P(A) - P(A \cap B) \Rightarrow (1.6)\end{aligned}$$

1.7 Erweiterung von 1.6:

$$\begin{aligned}\mathbf{P(A \cup B \cup C)} &= \mathbf{P(A) + P(B) + P(C)} \\ &\quad - \mathbf{P(A \cap B) - P(A \cap C) - P(B \cap C)} \\ &\quad + \mathbf{P(A \cap B \cap C)}\end{aligned}$$

1.8 Für alle Ereignisse  $A_1, A_2, \dots$  gilt

$$\begin{aligned}\mathbf{P\left(\bigcup_{i=1}^n A_i\right)} &\leq \sum_{i=1}^n \mathbf{P(A_i)} \\ \mathbf{P\left(\bigcup_{i=1}^{\infty} A_i\right)} &\leq \sum_{i=1}^{\infty} \mathbf{P(A_i)}\end{aligned}$$

Dies ist als **Boolesche Ungleichung** bekannt. Bei Disjunktheit gilt Gleichheit. Beweis in der Übung.

**Bsp. 1.9:** Ein Geschäft verkauft zwei Produkte. Für die zwei Ereignisse  $A_i$ : „Produkt ist (am Abend) ausverkauft“,  $i = 1, 2$  ist bekannt:  $P(A_1) = \frac{1}{8}$ ,  $P(A_2) = \frac{1}{10}$ ,  $P(A_1 \cap A_2) = \frac{1}{20}$ . Wie groß sind die W'keiten von:

$B_1$  : „Produkt 1 ausverkauft, aber Produkt 2 nicht“? (Lös:  $\frac{3}{40}$ )



$B_2$  : „genau ein Produkt ausverkauft“? (Lös:  $\frac{5}{40}$ )

$B_3$  : „mindestens ein Produkt ausverkauft“? (Lös:  $\frac{7}{40}$ )

$B_4$  : „höchstens ein Produkt ausverkauft“? (Lös:  $\frac{19}{20}$ )

**Bem. 1.1:** Ist  $\Omega$  nicht diskret, so ist es aus mathematischen Gründen in manchen Anwendungen nicht möglich, das W.-Maß  $P$  für alle Teilmengen von  $\Omega$  zu definieren. Jedoch lässt sich der Definitionsbereich von  $P$  stets so groß wählen, dass alle praktisch interessanten Teilmengen von  $\Omega$  erfasst sind. Ist  $\Omega$  diskret, so lässt sich das W.-Maß stets für alle Teilmengen von  $\Omega$  definieren (vgl. Satz 1.1).

## 2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

### 2.1 Bedingte Wahrscheinlichkeiten

Oft von Interesse: W'keit des Ereignisses unter der Annahme, dass ein anderes Ereignis eintritt bzw. eingetreten ist. Dafür definieren und berechnen wir sog. *bedingte Wahrscheinlichkeiten*.

**Bsp. 2.1:** 100 Personen

	mit Bachelorabschluss	ohne Bachelorabschluss	gesamt:
hohes Einkommen	21	19	40
niedriges Einkommen	15	45	60
	36	64	100

Wird von den 100 Personen eine zufällig ausgewählt, ergeben sich für die Ereignisse  $H$ : „hohes Einkommen“ und  $B$ : „mit Bachelorabschluss“ die Laplace-W'keiten:

$$P(H) = \frac{40}{100}, P(B) = \frac{36}{100}, P(H \cap B) = \frac{21}{100}$$

Intuitiv: W'keit für  $H$ , wenn bekannt ist, dass die Person einen Bachelorabschluss hat.

Quotient: Anzahl der Personen mit hohem Einkommen von denen mit Bachelorabschluss / Anzahl der Personen mit Bachelorabschluss =  $\frac{21}{36}$ . Bedingte W'keit ist also hier:

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

Entsprechend definiert man allgemeiner

**Def. 2.1:** Sei  $\Omega$  eine Ergebnismenge und  $P$  ein W'Maß auf  $\Omega$ .  $B \subset \Omega$  sei ein Ereignis mit  $P(B) > 0$ . Dann heißt

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

die *bedingte Wahrscheinlichkeit* von  $A$  unter der Bedingung  $B$ .

In Bsp. 2.1:

$$P(H|B) = \frac{P(H \cap B)}{P(B)} = \frac{21}{36}$$
$$P(B|H) = \frac{P(B \cap H)}{P(H)} = \frac{21}{40}$$

**Bem. 2.1:** Manchmal ist es einfach, direkt  $P(A|B)$  und  $P(B)$  zu bestimmen, um damit  $P(A \cap B)$  zu bestimmen:  $(A \cap B) = P(B) \cdot P(A|B)$

**Bsp. 2.2:** Eine Urne enthält  $s$  schwarze und  $r$  rote Kugeln. Zwei Kugeln werden nacheinander ohne Zurücklegen entnommen. Es seien

$A_1$ : „erste entnommene Kugel ist schwarz“

$A_2$ : „zweite entnommene Kugel ist rot“

Berechne  $P(A_1 \cap A_2)$ :

$$\begin{aligned} P(A_1) &= \frac{s}{r+s} \\ P(A_2|A_1) &= \frac{r}{r+s-1} \\ P(A_2 \cap A_1) &= P(A_1) \cdot P(A_2|A_1) = \frac{s \cdot r}{(r+s)(r+s-1)} \end{aligned}$$

Verallgemeinerung von 2.1:

### 2.1.1 Multiplikationssatz für bedingte Wahrscheinlichkeiten

Für  $n$  Ereignisse  $A_1, \dots, A_n \subset \Omega$  gilt

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

sofern  $P(A_1 \cap \dots \cap A_{n-1}) > 0$ .

Denn: Die Bedingung  $P(A_1 \cap \dots \cap A_{n-1}) > 0$  stellt sicher, dass die bed. W'keiten definiert sind, und es gilt

$$\begin{aligned} &P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \dots \cdot \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap \dots \cap A_n) \end{aligned}$$

**Bem. 2.2:** Betrachtet man bei festem  $B \subset \Omega$  mit  $P(B) > 0$   $P(A|B)$  als Funktion von  $A$ , dann ist  $P(A|B)$  ein Wahrscheinlichkeitsmaß auf  $\Omega$  (s.Tut., A14).

Daher lassen sich auch für  $P(A|B)$  die Eigenschaften (1.1)-(1.8) benutzen. Beispielsweise gilt

$$\begin{aligned} P(A^c|B) &= 1 - P(A|B) \\ P(A_1|B) &\leq P(A_2|B), \text{ falls } A_1 \subset A_2 \\ P(\emptyset|B) &= 0 \\ P(A_1 \cup A_2|B) &= P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B) \end{aligned}$$

### 2.1.2 Satz von der totalen Wahrscheinlichkeit

Seien  $A_1, \dots, A_n$  Ereignisse, die eine Zerlegung von  $\Omega$  bilden [d.h.  $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$  und  $A_i \cap A_j = \emptyset$  für  $i \neq j$ ] und es sei  $P(A_i) > 0$  für  $i = 1, \dots, n$ . Dann gilt

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i) \text{ für alle Ereignisse } B \subset \Omega$$

Denn:

$$\begin{aligned}
 \sum_{i=1}^n P(B|A_i) \cdot P(A_i) &= \sum_{i=1}^n P(B \cap A_i) \\
 &= P\left(\bigcup_{i=1}^n (B \cap A_i)\right) \\
 &= P\left(B \cap \left(\bigcup_{i=1}^n A_i\right)\right) \\
 &= P(B)
 \end{aligned}$$

### 2.1.3 Satz von Bayes

Seien  $A_1, \dots, A_n$  Ereignisse, die eine Zerlegung von  $\Omega$  bilden und es sei  $P(A_i) > 0$  für  $i = 1, \dots, n$ . Dann gilt für jedes  $B \subset \Omega$  mit  $P(B) > 0$

$$P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)} \quad \text{für } k = 1, \dots, n$$

Denn:

$$\begin{aligned}
 \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)} &\stackrel{(2.1.2)}{=} \frac{P(B|A_k) \cdot P(A_k)}{P(B)} \\
 &\stackrel{(B.2.1)}{=} \frac{P(B \cap A_k)}{P(B)} \\
 &\stackrel{(D.2.1)}{=} P(A_k|B)
 \end{aligned}$$

**Bsp. 2.3:** Betrachte für eingehende Emails die folgenden 3 Ereignisse:

$A_1$ : „Spam“

$A_2$ : „niedrige Priorität“

$A_3$ : „hohe Priorität“

$A_1, A_2, A_3$  bildet eine Zerlegung. Es gilt  $P(A_1) = 0,6$ ,  $P(A_2) = 0,3$ ,  $P(A_3) = 0,1$ .

Für das Ereignis  $B$ : „E-mail enthält das Wort ‚gratis‘“ gelte  $P(B|A_1) = 0,8$ ,  $P(B|A_2) = 0,05$ ,  $P(B|A_3) = 0,05$ . (W'keiten wurden durch Beobachtung des Postfachs errechnet.)

$\Rightarrow$  Was ist  $P(B)$ ?

Von dem Satz der totalen W'keit gilt

$$\begin{aligned}
 P(B) &= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) \\
 &= 0,8 \cdot 0,6 + 0,05 \cdot 0,3 + 0,05 \cdot 0,1 \\
 &= 0,5
 \end{aligned}$$

Angenommen, eine E-mail ist eingegangen, die das Wort „gratis“ enthält. Wie groß ist die W'keit, dass es sich um Spam handelt?

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{\sum_{i=1}^3 P(B|A_i)} = \frac{0,8 \cdot 0,6}{0,5} = 0,96$$

Analog,

$$P(A_2|B) = \frac{P(B|A_2) \cdot P(A_2)}{\sum_{i=1}^3 P(B|A_i)} = \frac{0,05 \cdot 0,3}{0,5} = 0,03$$
$$P(A_3|B) = \frac{P(B|A_3) \cdot P(A_3)}{\sum_{i=1}^3 P(B|A_i)} = \frac{0,05 \cdot 0,1}{0,5} = 0,01$$

**Bem. 2.2:** In den Situationen des Satzes von Bayes ( $A_1, \dots, A_n$  bilden eine Zerlegung von  $\Omega$ ,  $B \subset \Omega$ ,  $P(B) > 0$ ) bezeichnet man  $P(A_1), P(A_2), \dots, P(A_n)$  auch als *a-priori-Wahrscheinlichkeiten* und  $P(A_1|B), P(A_2|B), \dots, P(A_n|B)$  als *a-posteriori-Wahrscheinlichkeiten*.

**Bsp. 2.4:** Ein medizinischer Test für eine Krankheit liefert bei erkrankten Personen mit hoher W'keit 0,99 das richtige Resultat „positiv“. Bei Gesunden liefert der Test mit geringer W'keit 0,02 das falsche Resultat „positiv“. 0,3% der Personen aus der Bevölkerung sind krank (a priori). Wie groß ist die bedingte W'keit, dass jemand krank ist (a posteriori), wenn das Testergebnis positiv ist?

Seien

$A_1$ : „krank“

$A_2$ : „gesund Priorität“

$B$ : „positiv“

Geg. also  $P(A_1) = 0,003$  (a priori),  $P(A_2) = 1 - 0,003 = 0,997$ ,  $P(B|A_1) = 0,99$ ,  $P(B|A_2) = 0,02$ .

Formel von Bayes liefert

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$
$$= \frac{0,99 \cdot 0,003}{0,99 \cdot 0,003 + 0,02 \cdot 0,997}$$
$$= 0,13 \quad (\text{a posteriori})$$

## 2.2 Unabhängigkeit

Intuitiv: Ereignisse A und B sind unabhängig, falls die Kenntnis über das Eintreten des einen keine Information über die W'keit des Eintretens des anderen liefert:

$P(A|B) = P(A)$  und  $P(B|A) = P(B)$ ,  $P(A), P(B) > 0$ . Beide Gleichungen sind Äquivalent zu  $P(A \cap B) = P(A) \cdot P(B)$ . Nehme diese Gleichung für die Definition:

**Def. 2.2:** Zwei Ereignisse  $A, B \subset \Omega$  heißen *unabhängig*, wenn

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

**Bsp. 2.5:** Zweimaliger Würfelwurf. Der erste Wurf ist unabhängig vom Zweiten.

Also sind z.B.

$A_1$ : „im ersten Wurf 6“

$A_2$ : „im zweiten Wurf mind. 3“

unabhängige Ereignisse und  $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{6} \cdot \frac{2}{3} = \frac{1}{9}$  anwendbar.

**Bem. 2.4:** Sind A und B unabhängig, dann gilt auch

- (i)  $A$  und  $B^c$  sind unabhängig.

(ii)  $A^c$  und  $B$  sind unabhängig.

(iii)  $A^c$  und  $B^c$  sind unabhängig.

Nachweis von (i):

$$\begin{aligned}P(A \cap B^c) &\stackrel{(1.5)}{=} P(A) - P(A \cap B) \\&= P(A) - P(A) \cdot P(B) \\&= P(A) \cdot (1 - P(B)) \\&= P(A) \cdot P(B^c)\end{aligned}$$

**Def. 2.3:** Sei  $A_1, A_2, \dots$  eine endliche oder unendliche Folge von Ereignissen.  $A_1, A_2, \dots$  heißen *paarweise unabhängig*, falls

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$$

für jedes Paar von Indizes  $i \neq j$ .

$A_1, A_2, \dots$  heißen *unabhängig*, falls für jede endliche Auswahl von verschiedenen Indices  $i_1, \dots, i_k$  gilt

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

**Bem.2.5:** Unabhängigkeit  $\Rightarrow$  paarweise Unabhängigkeit, aber die Implikation gilt NICHT in die andere Richtung. Paarweise Unabhängigkeit ist außerdem oft zu schwach, um interessante Resultate zu erhalten.

**Bsp.2.6:** Zweimaliger Würfelwurf.

A: „erste Augenzahl gerade“

B: „zweite Augenzahl ungerade“

C: „beide Augenzahlen gerade oder beide ungerade“

$P(A) = \frac{1}{2}, P(B) = \frac{1}{2}, P(C) = \frac{1}{2}$ :

$$\begin{aligned}P(A \cap B) &= \frac{3 \cdot 3}{36} = \frac{1}{4} = P(A) \cdot P(B) \\P(A \cap C) &= \frac{3 \cdot 3}{36} = \frac{1}{4} = P(A) \cdot P(C) \\P(B \cap C) &= \frac{3 \cdot 3}{36} = \frac{1}{4} = P(B) \cdot P(C)\end{aligned}$$

$\Rightarrow A, B, C$  sind paarweise unabhängig. Aber  $A \cap B \cap C = \emptyset \Rightarrow P(A \cap B \cap C) = 0 < P(A) \cdot P(B) \cdot P(C) \Rightarrow A, B, C$  sind nicht unabhängig.

**Bem.2.6:** Sind  $A_1, \dots, A_n$  unabhängige Ereignisse, und ist für jedes  $i \in 1, \dots, n$   $B_i = A_i$  oder  $B_i = A_i^c$ , dann sind  $B_1, \dots, B_n$  unabhängig.

**Bsp.2.7:** Drei Personen werden gefragt, ob sie einem bestimmten Vorschlag zustimmen. Jede Person antwortet mit W'keit 0,8 „nein“ und mit W'keit 0,2 „ja“. Die Antworten sind unabhängig. Wie groß ist dann die W'keit, dass alle drei Personen dieselbe Antwort geben?

Sei  $A_i$  das Ereignis. Person  $i$  antwortet „ja“  $i = 1, 2, 3$ . Dann gilt  $P(A_1) = P(A_2) = P(A_3) = 0,2$  und  $A_1, A_2, A_3$  sind unabhängig.

Gesuchte W'keit ist

$$\begin{aligned} P([A_1 \cap A_2 \cap A_3] \cup [A_1^c \cap A_2^c \cap A_3^c]) &= P(A_1 \cap A_2 \cap A_3) + P(A_1^c \cap A_2^c \cap A_3^c) \\ &= P(A_1)P(A_2)P(A_3) + P(A_1^c)P(A_2^c)P(A_3^c) \\ &= 0,2^3 + 0,8^3 = 0,52 \end{aligned}$$

### 3 Zufallsvariablen

#### 3.1 Zufallsvariablen

Eine *Zufallsvariable* beschreibt eine reellwertige Größe, die vom Zufall, d.h. von  $\omega$  abhängt. Genauer:

**Def. 3.1:** Sei  $\Omega$  eine Ergebnismenge. Eine Zufallsvariable (ZV)  $X$  ist eine Abbildung von  $\Omega$  nach  $\mathbb{R}$ ,  $X : \Omega \rightarrow \mathbb{R}$ .

**Bsp. 3.1:** Zweimaliger Münzwurf

$\Omega$	$X(\omega)$ = Anzahl von K
$\omega_1 = (K, K)$	$X(\omega_1) = 2$
$\omega_2 = (K, K)$	$X(\omega_2) = 1$
$\omega_3 = (K, K)$	$X(\omega_3) = 1$
$\omega_4 = (K, K)$	$X(\omega_4) = 0$

Ist  $P(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{4}$  für alle  $\omega \in \Omega = \{\omega_1, \dots, \omega_4\}$ , dann lassen sich die W'keiten

$$\begin{aligned} P(X = n) &= \text{W'keit, dass die ZV } X \text{ den Wert } n \text{ annimmt} \\ &= \text{W'keit, dass } n\text{-mal Kopf fällt} \\ &= P(\{\omega \in \Omega : X(\omega) = n\}) \quad n = 0, 1, 2, \dots \end{aligned}$$

berechnen.

$$\begin{aligned} P(X = 0) &= P(\{\omega_4\}) = \frac{1}{4} \\ P(X = 1) &= P(\{\omega_2, \omega_3\}) = \frac{1}{2} \\ P(X = 2) &= P(\{\omega_1\}) = \frac{1}{4} \\ P(X = 3) &= P(\emptyset) = 0 \end{aligned}$$

Notation: Für ZV  $X, Y$

$$\begin{aligned} \{X = a\} &= \{\omega \in \Omega : X(\omega) = a\}, \quad a \in \mathbb{R} \\ \{X < a\} &= \{\omega \in \Omega : X(\omega) < a\} \\ \{X \in A\} &= \{\omega \in \Omega : X(\omega) \in A\}, \quad A \in \mathbb{R} \\ \{X \in A, Y \in B\} &= \{\omega \in \Omega : X(\omega) \in A \text{ und } Y(\omega) \in B\}, \quad A, B \in \mathbb{R} \\ P(X \geq a) &= P(\{\omega \in \Omega : X(\omega) \geq a\}) \\ P(X \leq a, Y \leq b) &= P(\{\omega \in \Omega : X(\omega) \leq a \text{ und } Y(\omega) \leq b\}), \quad a, b \in \mathbb{R} \\ P(X > Y) &= P(\{\omega \in \Omega : X(\omega) > Y(\omega)\}) \end{aligned}$$

usw.

## 3.2 Unabhängigkeit

Bisher: Unabhängigkeit von Ereignissen.

Jetzt: Unabhängigkeit von Zufallsvariablen

**Def. 3.2:**

(a) ZV  $X_1, \dots, X_n$  heißen unabhängig, falls

$$(*) \quad P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \dots \cdot P(X_n \in B_n)$$

für alle  $B_1, \dots, B_n \in \mathbb{R}$ .

(b) Ist  $X_1, X_2, \dots$  eine unendliche Folge von ZV, dann heißen  $X_1, X_2, \dots$  unabhängig, falls

(\*) für alle  $n \geq 2$  gilt.

**Bem. 3.1:** Sind  $X_1, \dots, X_n$  unabhängige ZV und sind  $A_1, \dots, A_n$  Ereignisse, so dass Eintreten von  $A_i$  nur von  $X_i$  abhängt,  $i = 1, \dots, n$ , dann sind  $A_1, \dots, A_n$  unabhängig.

**Bsp. 3.2:** Seien  $X_1, X_2, X_3$  unabhängige ZV. Dann sind die Ereignisse  $A_1 = \{X_1 < 3\}$ ,  $A_2 = \{X_2^2 > 7\}$ ,  $A_3 = \{4 < |X_3| < 7\}$  unabhängig.

Wählt man in (\*) z.B.  $B_1 = \{x \in \mathbb{R} : x < 3\}$ ,  $B_2 = \{x \in \mathbb{R} : x^2 < 7\}$ ,  $B_3 = \mathbb{R}$  ergibt sich:

$$\begin{aligned} P(A_1 \cap A_2) &= P(X_1 \in B_1, X_2 \in B_2, X_3 \in B_3) \\ &= P(X_1 \in B_1) \cdot P(X_2 \in B_2) \cdot P(X_3 \in B_3) \\ &= P(A_1) \cdot P(A_2) \cdot 1 \end{aligned}$$

### 3.2.1 Satz 3.1:

$X_1, \dots, X_n$  seien ZV, die jeweils höchstens abzählbar viele Werte annehmen. Dann gilt:  $X_1, \dots, X_n$  sind unabhängig, gdw

$$P(X_1 = b_1, \dots, X_n = b_n) = P(X_1 = b_1) \dots P(X_n = b_n)$$

für alle  $b_1, \dots, b_n \in \mathbb{R}$ .

**Bsp. 3.3:**  $X_1$  sei ZV mit  $P(X_1 = -1) = P(X_1 = 0) = P(X_1 = 1) = \frac{1}{3}$  und ZV  $X_2$  sei definiert durch  $X_2 = X_1^2 + 2$ .

Für  $b_1 = 0$  und  $b_2 = 2$  gilt

$$\begin{aligned} P(X_1 = b_1) &= P(X_1 = 0) = \frac{1}{3} \\ P(X_2 = b_2) &= P(X_1^2 + 2 = 2) = P(X_1 = 0) = \frac{1}{3} \end{aligned}$$

$$\begin{aligned} \Rightarrow P(X_1 = b_1, X_2 = b_2) &= P(X_1 = 0, X_1^2 + 2 = 2) \\ &= P(X_1 = 0) = \frac{1}{3} \end{aligned}$$

und daher  $P(X_1 = b_1, X_2 = b_2) \neq P(X_1 = b_1) \cdot P(X_2 = b_2)$ .

$\Rightarrow X_1$  und  $X_2$  sind nicht unabhängig.

**Bsp. 3.4:** n-maliger Würfelwurf (mit Laplace-Würfel)

$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, \dots, 6\}; 1 \leq i \leq n\}$

$|\Omega| = 6^n, P(A) = \frac{|A|}{|\Omega|}$  für alle  $A \in \Omega$

$X_i$  = Ergebnis des i-ten Wurfs

$X_i((\omega_1, \dots, \omega_n)) = \omega_i$

Behauptung:  $X_1, \dots, X_n$  sind unabhängig. Nach Satz 3.1 reicht es zu zeigen:

$P(X_1 = b_1, \dots, X_n = b_n) = P(X_1 = b_1) \cdot \dots \cdot P(X_n = b_n) \forall b_1, b_2, \dots, b_n \in \mathbb{R}.$

**Fall 1:** Ein  $b_i$  ist keine mögliche Augenzahl  $\Rightarrow$  beide Seiten = 0

**Fall 2:**  $b_1, \dots, b_n \in \{1, \dots, 6\}$ :

$$\begin{aligned} P(X_1 = b_1, \dots, X_n = b_n) &= P(\{\omega \in \Omega : X_1(\omega) = b_1, \dots, X_n(\omega) = b_n\}) \\ &= P(\{\omega \in \Omega : \omega_1 = b_1, \dots, \omega_n = b_n\}) \\ &= P(\{b_1, \dots, b_n\}) \\ &= \frac{1}{|\Omega|} = \frac{1}{6^n} \end{aligned}$$

und

$$\begin{aligned} P(X_i = b_i) &= P(\{\omega \in \Omega : X_i(\omega) = b_i\}) \\ &= P(\{\omega \in \Omega : \omega_i = b_i\}) \\ &= \frac{6^{n-1}}{6^n} = \frac{1}{6}, i = 1, \dots, 6 \end{aligned}$$

$\Rightarrow P(X_1 = b_1) \cdot \dots \cdot P(X_n = b_n) = \frac{1}{6^n} = P(X_1 = b_1, \dots, X_n = b_n).$

Es folgt, dass  $X_1, \dots, X_n$  unabhängig sind.

### 3.3 Verteilungsfunktionen, W'keitsfunktionen und Dichten

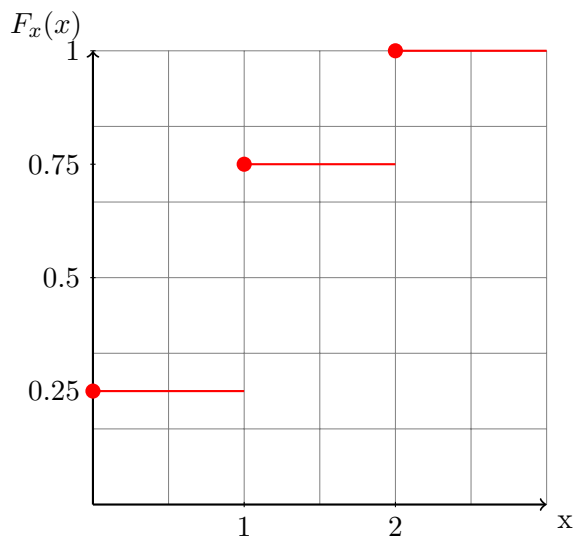
**Def. 3.3:** Sei  $X$  eine ZV. Die Verteilungsfunktion (VF)  $F = F_x$  von  $x$  ist definiert durch

$$F(x) = F_x(x) = P(X \leq x) \forall x \in \mathbb{R}.$$

**Bsp. 3.5:** Für  $X$  wie in Bsp. 3.1:  $P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{1}{2}, P(X = 2) = \frac{1}{4}.$

$$\Rightarrow F_x(x) = P(X \leq x) = \begin{cases} P(\emptyset) = 0 & x < 0 \\ P(X = 0) = \frac{1}{4} & 0 \leq x < 1 \\ P(X = 0) + P(X = 1) = \frac{3}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

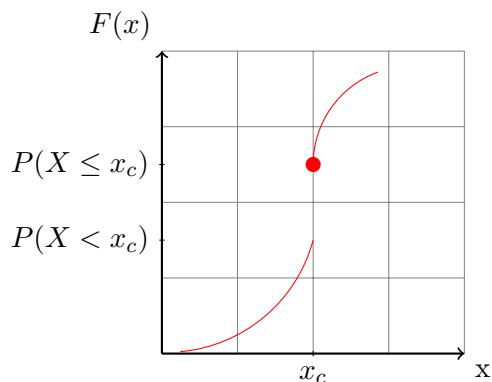




### 3.3.1 Satz 3.2: Eigenschaften von VF

Sei  $X$  eine ZV und  $F$  ihre VF. Dann gilt

- (a)  $F$  ist monoton wachsend.  
Für alle  $x, y \in \mathbb{R}$  mit  $x < y$  gilt  $F(x) \leq F(y)$ .
- (b)  $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$
- (c)  $F$  ist rechtsseitig stetig.
- (d)  $P(X > x) = 1 - F(x), x \in \mathbb{R}$
- (e)  $P((a < X \leq b)) =$
- (f)  $P(X < x) = F(x-)$ , wobei  $F(x-) = \lim_{y \rightarrow x} F(y)$  der linksseitige Grenzwert von  $F$  an der Stelle  $x$  ist, d.h., der Grenzwert der Werte  $F(y)$  wenn  $y$  sich  $x$  von links nähert.
- (g)  $P(X = x) = F(x) - F(x-) =$  Sprunghöhe von  $F$  an der Stelle  $x$   
Insb.



$$P(X = x) = P(X \leq x_c) - P(X < x_c) \text{ (s. Sprung in der Grafik)}$$

$$P(X = x) = 0 \Leftrightarrow F \text{ ist stetig an der Stelle } x$$

$$P(X = x) = 0 \text{ für alle } x \in \mathbb{R} \Leftrightarrow F \text{ ist stetig auf } \mathbb{R}$$

- (h)  $P(X \geq x) = 1 - F(x-)$   
 $P(a \leq X \leq b) = F(b) - F(a-), \quad a \leq b$

**Beweis:**

$$(a),(e) \text{ Für } a < b \text{ gilt } 0 \leq P(a < X \leq b) = P(\{X \leq b\} \setminus \{X \leq a\}) = \\ = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

(b),(c) Übungsaufgabe 21

$$(d) P(X > x) = P(\{X \leq x\}^c) = 1 - P(X \leq x) = 1 - F(x)$$

$$(f) P(X < x) = P(\bigcup_{n=1}^{\infty} \{X \leq x - \frac{1}{n}\}) \\ \text{Übungsaufgabe 20} = \lim_{n \rightarrow \infty} P(X \leq x - \frac{1}{n}) = \lim_{n \rightarrow \infty} F(x - \frac{1}{n}) = F(x-) \\ \{X \leq x - \frac{1}{n}\} \subset \{X \leq x - \frac{1}{n+1}\}$$

$$(h) P(X \leq x) = P(\{X < x\}^c) = 1 - P(X < x) \stackrel{(f)}{=} 1 - F(x-) \\ P(a \leq X \leq b) = P(\{x \leq b\} \setminus \{X < a\}) = P(X \leq b) - P(X < a) = F(b) - F(a-). \\ \text{Für } a = b = x \text{ ergibt sich (g).}$$

**Bem. 3.2:** Man kann zeigen:

(a) Jede Funktion  $F : \mathbb{R} \rightarrow \mathbb{R}$ , die (a) – (c) aus Satz 3.2 erfüllt, ist die VF einer geeigneten ZV.

(b) Seien  $X$  und  $Y$  ZF mit VF  $F_x$  und  $F_y$ . Dann gilt  
 $F_x(x) = F_y(x)$  für alle  $x \in \mathbb{R} \Leftrightarrow P(X \in B) = P(Y \in B)$  für alle  $B \in \mathcal{B}$ .  
 Falls  $F_x(x) = F_y(x)$ , dann sagt man „ $X$  und  $Y$  sind identisch verteilt“.

**3.3.2 Diskrete ZV**

**Def. 3.4:** Eine ZV  $X$  heißt *diskret*, falls  $X$  höchstens abzählbar viele Werte  $x_1, x_2, \dots$  annimmt.

Ist  $X$  eine diskrete ZV, dann heißt die durch

$$f(x) = P(X = x) \quad x \in \mathbb{R},$$

definierte Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  die *Wahrscheinlichkeitsfunktion* von  $X$ .

Sei  $X$  eine diskrete ZV mit W'funktion  $f$  und VF  $F$ . Bezeichne die möglichen Werte von  $X$  mit  $x_1, x_2, \dots$ , d.h.,

$$\{x_1, x_2, \dots\} = \{X(\omega) : \omega \in \Omega\}$$

Dann gilt für jedes  $B \in \mathcal{B}$ ,

$$P(X \in B) = P(\bigcup_{x \in B} \{X = x\}) = \sum_{x_i \in B} P(X = x_i) = \sum_{x_i \in B} f(x_i)$$

$$\sum_{x_i \in \mathbb{R}} f(x_i) = P(X \in \mathbb{R}) = 1$$

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i), \quad x \in \mathbb{R}$$

$$f(x) = P(X = x) \stackrel{(Satz 3.2(g))}{=} F(x) - F(x-) \quad x \in \mathbb{R}$$

Setze

$$P_x(B) = P(X \in B) \text{ für jedes } B \in \mathbb{R}.$$

Dann ist  $P_x$  ein W-Maß auf  $\mathbb{R}$ , denn:

(i)  $P_x(B) \geq 0$  für alle  $B \subset \mathbb{R}$

(ii)  $P_x(\mathbb{R}) = P(X \in \mathbb{R}) = 1$

(iii)  $\sigma$ -Additivität

Seien  $B_1, B_2, \dots \subset \mathbb{R}$  paarweise disjunkt

$$\begin{aligned} \Rightarrow P_x(B_1 \cup B_2 \cup \dots) &= P(X \in B_1 \cup B_2 \cup \dots) = P(\{X \in B_1\} \cup \{X \in B_2\} \cup \dots) \\ &\stackrel{(\sigma\text{-Add.})}{=} \sum_i P(X \in B_i) = \sum_i P_x(B_i). \end{aligned}$$

Das W-Maß  $P_x$  heißt die Verteilung von  $X$ .

### 3.3.3 Stetige ZV

**Def. 3.5:** Eine ZV  $X$  heißt *stetig*, falls es eine Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  gibt, so dass

$$\begin{aligned} f(x) &\geq 0 \quad \forall x \in \mathbb{R}, \\ \int_{-\infty}^{\infty} f(x) dx &= 1 \\ \text{und} \quad P(X \leq x) &= \int_{-\infty}^x f(u) du \quad \forall x \in \mathbb{R}. \end{aligned}$$

$f$  heißt *Dichte* oder *Wahrscheinlichkeitsdichte* von  $X$ .

Sei  $X$  eine stetige ZV mit Dichte  $f$  und VF  $F$

$\rightarrow F$  ist stetig  $\stackrel{(\text{Satz 3.2(g)})}{\Rightarrow} P(X = x) = 0$  für alle  $x \in \mathbb{R}$ .

$\Rightarrow$  Für  $a \leq b$  gilt

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X < b) + P(X = b) = P(a < X < b) \\ &= P(a \leq X < b) = P(a < X \leq b) = F(b) - F(a) \\ &= \int_{-\infty}^a f(x) dx - \int_{-\infty}^b f(x) dx = \int_a^b f(x) dx \end{aligned}$$

Ist die Dichte  $f$  an der Stelle  $x_c$  stetig, so gilt

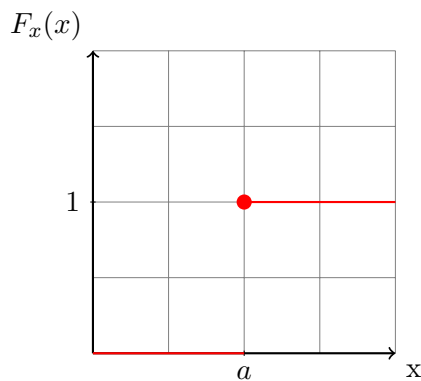
$$F'(x_c) = f(x_c)$$

$f$  lässt sich aus  $F$  gewinnen.

### 3.4 Beispiele diskreter Verteilungen

*Diskrete Verteilungen*, d.h. Verteilungen von diskreten ZV

Einfachster Fall:  $\exists a \in \mathbb{R}$  mit  $P(X = a) = 1$ .



### 3.4.1 Bernoulli-Verteilung

Eine ZV  $X$ , die nur die Werte 0 und 1 annimmt mit

$$P(X = 1) = p, P(X = 0) = 1 - p$$

heißt *Bernoulli-verteilt* mit Parameter  $p$ .

$X$  wird auch als Bernoulli-Variable bezeichnet. ( $0 \leq p \leq 1$ )

Bez.  $X \sim \text{Ber}(p)$ .

Sei  $A$  ein Ereignis. Die Indikatorfunktion  $I_A$  von  $A$  (oder Indikatorvariable) ist definiert durch

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in A^c \end{cases}$$

Also

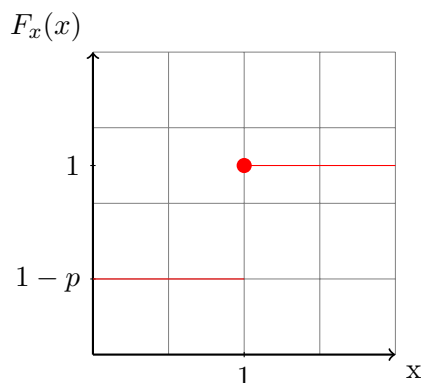
$$I_A(\omega) = 1 \Leftrightarrow A \text{ tritt ein}$$

$X = I_A$  ist eine Bernoulli-Variable mit Parameter  $p = P(A)$ .

Man sagt oft,

Erfolg tritt ein, falls  $X(\omega) = 1$

Misserfolg tritt ein, falls  $X(\omega) = 0$



### 3.4.2 Binomialverteilung

Seien  $X_1, \dots, X_n$  unabhängig und jeweils  $\text{Ber}(p)$  (Erfolgsw'keit jeweils  $p$ ).

Anzahl der Erfolge

$$S_n := X_1 + \dots + X_n$$

Bsp.: n Münzwürfe

$$X_i = \begin{cases} 1, & \text{im } i\text{-ten Wurf K} \\ 0, & \text{im } i\text{-ten Wurf Z} \end{cases}$$

$S_n$  = Anzahl der K-Würfe  $\Rightarrow$  Mögliche Werte von  $S_n : 0, 1, \dots, n$ .

$P(S_n = k) = ?$

Da  $X_1, \dots, X_n$  unabhängig sind,

$$\begin{aligned} & P(X_1 = 1, \dots, X_k = 1, X_{k+1} = 0, \dots, X_n = 0) \\ &= P(X_1 = 1) \cdot \dots \cdot P(X_k = 1) \cdot P(X_{k+1} = 0) \cdot \dots \cdot P(X_n = 0) \\ &= p^k \cdot q^{n-k} \quad (q := 1 - p) \end{aligned}$$

$\rightarrow$  hängt nur ab von der Anzahl der vorgegebenen Einsen, nicht von deren Position. Also

$$P(S_n = k) = p^k \cdot q^{n-k}$$

„Anzahl der Möglichkeiten, aus n Versuche k anzugeben, bei denen Erfolg eintreten soll“

$$\Rightarrow P(S_n = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}, k = 0, 1, \dots, n$$

wobei  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  (Binomialkoeffizient).

Insb.  $P(S_n = 0) = \binom{n}{0} \cdot p^0 \cdot q^n = q^n, P(S_n = 1) = \binom{n}{1} \cdot p^1 \cdot q^{n-1} = p \cdot q^{n-1}$ .

$S_n$  heißt *binomialverteilt* mit Parametern n (# Versuchen) und p (Erfolgsw'keit).

Bez.  $S_n \sim \text{Bin}(n, p)$

**Bsp. 3.6:** Wie groß ist die W'keit, dass von 5 E-Mails höchstens eine Spam ist, wenn bei jeder E-Mail die W'keit für Spam  $\frac{1}{3}$  ist?

$X$  = Anzahl der Spam E-Mails  $\sim \text{Bin}(5, \frac{1}{3})$

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{5}{0} \cdot \left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^5 + \binom{5}{1} \cdot \left(\frac{1}{3}\right)^1 \cdot \left(\frac{2}{3}\right)^4 \\ &= \frac{2^5}{3^5} + \frac{5!}{1!4!} \cdot \frac{1}{3} \cdot \left(\frac{2}{3}\right)^4 \end{aligned}$$

### 3.4.3 Geometrische Verteilung

Seien  $X_1, X_2, \dots$  unabh.,  $\sim \text{Ber}(p), 0 < p \leq 1$

$X_i = 1 \leftrightarrow$  Erfolg im i-ten Versuch.

$T$  = Anzahl der Versuche bis zum ersten Erfolg (einschließlich)

$$P(T = 1) = P(X_1 = 1) = p$$

Für  $k \geq 2$  ist

$$\begin{aligned} P(T = k) &= P(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= P(X_1 = 0) \cdot \dots \cdot P(X_{k-1} = 0) \cdot P(X_k = 1) = (1 - p)^{k-1} \cdot p \end{aligned}$$

Mit  $q = 1 - p$  also

$$P(T = k) = q^{k-1} \cdot p, k = 1, 2, \dots$$

Bez.  $T \sim \text{Geo}(p)$  Geometrische Verteilung mit Parameter  $0 < p \leq 1$ .

Beachte

$$\sum_{k=1}^{\infty} q^{k-1} \cdot p = p(1 + q + q^2 + \dots) = p \cdot \frac{1}{1 - q} = \frac{p}{p} = 1$$

### 3.4.4 Poisson-Verteilung

Eine ZV  $X$  heißt *Poisson-verteilt* mit Parameter  $\lambda > 0$ , falls

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Bez.  $X \sim \text{Poi}(\lambda)$ .

Beachte

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

über die Taylor-Reihe für  $e^{\lambda}$ .

Anwendung der Poisson-Verteilung zu Approximation einer Binomialverteilung.

Sei  $X \sim \text{Bin}(n, p)$ . Ist  $n$  „groß“ und  $p$  „klein“ etwa  $n \geq 50, p \leq \frac{1}{10}$  und  $np \leq 10$ , dann

$$P(X = k) \approx P(Y = k), \text{ wobei } Y \sim \text{Poi}(np)$$

**Bsp. 3.7:** Ein Callcenter erhält in jeder Sekunde mit W'keit 0,001 einen Anruf. Anrufe erfolgen unabhängig.

$X$  = Anzahl der Anrufe in einer Stunde

$\sim \text{Bin}(n, p)$  mit  $n = 3600$  und  $p = 0,001$

$\Rightarrow$  Poisson-Approximation [ $n \geq 50, p \leq \frac{1}{10}, np = 3,6 \leq 10$ ]

Approximiere  $\text{Bin}(n, p)$  durch  $\text{Poi}(\lambda)$  mit  $\lambda = np = 3,6$ .

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = P(Y = k) \quad \text{wobei } Y \sim \text{Poi}(\lambda)$$

k	0	1	2
$P(X = k)$	.02727	.09829	.1770
$P(Y = k)$	.02732	.09837	.1771

### 3.4.5 Hypergeometrische Verteilung

Warenlieferung enthalten  $N$  Stücken, davon  $M$  defekt ( $N-M$  intakt)

Es werden  $n$  Stücke ohne Zurücklegen zufällig gezogen, wobei  $n \leq N$ . Sei

$X$  = Anzahl der defekten Stücke in Stichprobe vom Umfang  $n$

Die W-funktion von  $X$  ist gegeben durch

$$f(x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, & x = 0, 1, \dots, n \\ 0, & \text{sonst} \end{cases}$$

Daher wird die Konvention

$$\binom{m}{k} = 0 \quad \text{falls } k < 0 \text{ oder } k > m$$

benutzt.

Für  $X \in \mathbb{N}_0$  gilt

$$\begin{aligned} f(X) > 0 &\Leftrightarrow 0 \leq x \leq M \text{ und } 0 \leq n - x \leq N - M \\ &\Leftrightarrow x \in \{\max(0, n - N + M), \dots, \min(n, M)\} \end{aligned}$$

Eine ZV mit W-funktion  $f$  heißt *hypergeometrisch verteilt* mit Parametern  $N, M$ , und  $n$ .  
 Bez.  $X \sim \text{Hyp}(N, M, n)$ .

**Bsp. 3.8:** Lotto 6 aus 49

Jemand hat die Zahlen  $1, 2, \dots, 6$  getippt. W'keit für 3 Richtige?

Warenlieferung	49 Kugeln	$N = 49$
defekt	richtige Zahl $(1, \dots, 6)$	$M = 6$
intakt	falsche Zahl $(7, \dots, 49)$	$N - M = 43$

$n = 6$  Kugeln werden ohne Zurücklegen gezogen.

$X$  = Anzahl richtig getippter Zahlen.

W'keit für 3 Richtige:

$$P(X = 3) = \frac{\binom{6}{3} \binom{49-6}{6-3}}{\binom{49}{6}} \approx 0,01765$$

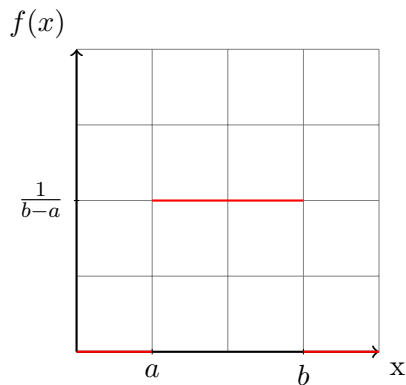
### 3.5 Beispiele stetiger Verteilungen

#### 3.5.1 Gleichverteilung

**Gleichverteilung** auf  $[a, b]$ ,  $a < b$ .

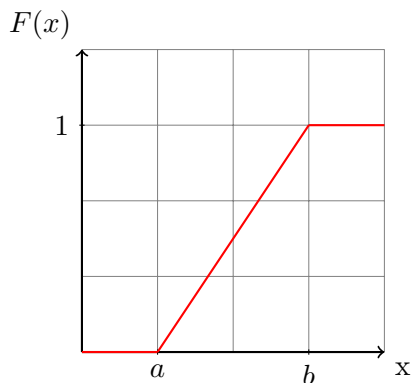
Dichte

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \in \mathbb{R} \setminus [a, b] \end{cases}$$



VF

$$F(x) = \int_{-\infty}^x f(u) du = \begin{cases} 0, & x < a \\ \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$



**Bsp. 3.9:** Ist  $X \sim \text{Uni}(0,1)$ , dann gilt für  $0 \leq c \leq d \leq 1$

$$P(c \leq X \leq d) = \int_c^d 1 dx = d - c$$

und für  $c < 0 \leq d \leq 1$  gilt

$$P(c \leq X \leq d) = \int_c^0 0 dx + \int_0^d 1 dx = d$$

### 3.5.2 Exponentialverteilung

mit Parameter  $\lambda > 0$

Bez.  $X \sim \text{Exp}(\lambda)$

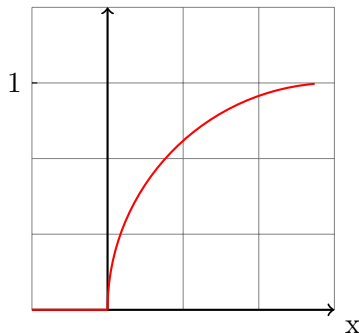
Diese Verteilung wird oft benutzt zur Modellierung von Wartezeiten oder Lebensdauern

Dichte

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

VF

$$F(x) = \int_{-\infty}^x f(u) du = \begin{cases} \int_0^x \lambda e^{-\lambda u} du = -e^{-\lambda u} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$



$$\int_{-\infty}^{\infty} f(u) du = \int_0^{\infty} \lambda e^{-\lambda x} = [e^{-\lambda x}]_0^{\infty} = 1$$

### 3.5.3 Normalverteilung

(oder Gaußverteilung)

$$N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$$

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

$\mu$ : Mittelwert/Erwartungswert einer Zufallsvariable mit obiger Dichte

$\sigma^2$ : Varianz einer Zufallsvariable mit obiger Dichten

$N(0,1)$  heißt *Standardnormalverteilung*

Dichte

$$\phi(X) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{X^2}{2}\right)$$



VF

$$\Phi(X) = \int_{-\infty}^x \phi(u) du$$

→ Abgeschlossen bezüglich affiner Transformation.

Beweis:

Sei  $X \sim N(0, 1)$ , seien  $a, b \in \mathbb{R}, b \neq 0$

$\Rightarrow Y := a + bX \sim N(a + b\mu, b^2\sigma^2)$

Nachweis für den Fall  $b > 0$  (Fall  $b < 0$  analog)

$P(Y \geq y) = \text{VF der transformierten Zufallszahl}$

$$\begin{aligned} P(Y \geq y) &= P(a + bX \leq y) = P(X \leq \frac{y-a}{b}) \\ &= F_x\left(\frac{y-a}{b}\right), y \in \mathbb{R} \\ \Rightarrow \frac{d}{dy} P(Y \geq y) &= F'_x\left(\frac{y-a}{b}\right) \frac{1}{b} = f\left(\frac{y-a}{b}\right) \frac{1}{b} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{\left(\frac{y-a}{b} - \mu\right)^2}{2\sigma^2}\right) \frac{1}{b} \\ &= \frac{1}{\sqrt{2\pi(b\sigma)^2}} \cdot \exp\left(-\frac{[y - (a + b \cdot \mu)]^2}{2(b\sigma)^2}\right) \end{aligned}$$

$\Rightarrow$  Dichte von Y ist die Dichte der  $N(a + b\mu, b^2\sigma^2)$ -Verteilung

Insb. gilt:

Falls  $X \sim N(0, 1)$ , dann  $a + bX \sim N(a, b^2)$

Falls  $X \sim N(\mu, \sigma^2)$ , dann  $\frac{X-\mu}{\sigma} \sim N(0, 1)$

und  $P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right), x \in \mathbb{R}$

Dabei ist  $\Phi$  die VF einer  $N(0, 1)$ -verteilten ZV.  $\frac{X-\mu}{\sigma}$  ist die zu X gehörende standardisierte ZV. Durch Standardisierung kann eine beliebige Normverteilung als Standard-Normalverteilung dargestellt werden.

## 4 Erwartungswert, Varianz und Kovarianz

### 4.1 Erwartungswert einer ZV

Sei X eine diskrete ZV, und seien  $x_1, x_2, \dots$  die verschiedenen möglichen Werte von f, sei f die W'funktion von X.

Der *Erwartungswert* (EW) von X wird definiert durch

$$E(X) = \sum_i x_i f(x_i) = \sum_i x_i P(X = x_i) \quad (\text{gewichtetes Mittel der } x_i)$$

Hier wird angenommen, dass  $E(X) = \sum_i x_i f(x_i) < \infty$  ist. Dies stellt sicher, dass der Erwartungswert wohldefiniert ist und nicht von der Summationsreihenfolge abhängt. Für nichtnegative diskrete ZV X ist  $E(X)$  immer nichtdefiniert, aber eventuell  $= \infty$ .

**Bsp. 4.1:**

(a) Sei  $\Omega = \{1, \dots, n\}$ ,  $P(A) = \frac{|A|}{|\Omega|}$  für  $A \subset \Omega$  und  $a_\omega = X(\omega)$ ,  $\omega \in \Omega = \mathbb{R} = 4$ .

$$\begin{aligned} \Rightarrow E(X) &= \sum_{x \in \{a_1, \dots, a_n\}} x P(X = x) \\ &= \sum_{x \in \{a_1, \dots, a_n\}} x \frac{|\{\omega \in \Omega : X(\omega) = x\}|}{n} \\ &= \frac{1}{n} \cdot \sum_{i=1}^n a_i \end{aligned}$$

(b) Ist  $P(X = a) = 1$ ,  $a \in \mathbb{R}$ , dann ist

$$E(X) = a \cdot P(X = a) = a$$

(c) Sei  $X \sim \text{Ber}(p)$ ,  $X = I_A$ ,  $P(A) = p$

$$\begin{aligned} \Rightarrow E(X) &= E(I_A) = 0 \cdot P(I_A = 0) + 1 \cdot P(I_A = 1) \\ &= 1 \cdot P(A) = p \end{aligned}$$

(d) Für  $X \sim \text{Poi}(\lambda)$  gilt:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \stackrel{j=k-1}{=} e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} \\ &= e^{-\lambda} \lambda \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{=e^\lambda} = \lambda \end{aligned}$$

#### 4.1.1 Satz 4.1:

Ist  $X$  eine ZV, die nur Werte in  $\{0, 1, 2, \dots\}$  annimmt, dann gilt

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n)$$

**Bew.:**

$$\begin{aligned} E(X) &= \sum_{n=1}^{\infty} n P(X = n) = P(X = 1) \\ &\quad + P(X = 2) + P(X = 2) \\ &\quad + P(X = 3) + P(X = 3) + P(X = 3) \\ &\quad + \dots \\ &= P(X > 0) + P(X > 1) + P(X > 2) + \dots \\ &= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots \end{aligned}$$

**Bsp. 4.2:** Ist  $X \sim \text{Geo}(p)$ , dann

$$P(X > n) = (1 - p)^n, n = 0, 1, \dots$$

und

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} (1 - p)^n = \frac{1}{p}$$

Sei  $X$  eine stetige Zufallsvariable mit Dichte  $f$ . Dann ist der Erwartungswert von  $X$  definiert durch

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Hier wird angenommen, dass  $\int_{-\infty}^{\infty} x f(x) dx < \infty$  ist.

Für nichtnegative stetige ZV  $X$  ist  $E(X)$  immer wohldefiniert, aber evtl.  $\infty$ .

**Bsp. 4.3:**

(a) Sei  $X \sim \text{Uni}(a, b)$

$$\begin{aligned} \Rightarrow E(X) &= \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} \\ &= \frac{a+b}{2} \text{Mittelpunkt von } [a, b] \end{aligned}$$

(b) Sei  $X \sim \text{Exp}(\lambda)$

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx \tag{1}$$

$$= [-x e^{-\lambda x}]_0^{\infty} - \int_0^{\infty} e^{-\lambda x} dx \tag{2}$$

$$= 0 + \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda} \tag{3}$$

$$E(X) = \frac{1}{\lambda} \tag{4}$$

$$\tag{5}$$

#### 4.1.2 Satz 4.2:

Seien  $X, Y, X_1, \dots, X_n$  beliebige ZV (diskret oder stetig).

(a)

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

$$E(a + bX) = a + bE(X) \text{ für alle } a, b \in \mathbb{R}$$

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \cdot E(X_i)$$

(b) Falls  $X_1, \dots, X_n$  unabhängig sind, dann gilt

$$E(X_1, \dots, X_n) = E(X_1) \cdot \dots \cdot E(X_n)$$

(c) Falls  $X$  eine nichtnegative ZV ist, dann gilt

$$E(X) \geq 0 \text{ mit Gleichheit genau dann, wenn } P(X = 0) = 1$$

(d) Falls  $P(X \leq Y) = 1$ , dann  $E(X) \leq E(Y)$  (Monotonie)

**Bsp. 4.4:**

(a)  $X \sim \text{Bin}(n, p)$

Benutze Darstellung  $X = X_1 + \dots + X_n, X_i \sim \text{Ber}(p)$

$$\Rightarrow E(X) \stackrel{\text{Satz 4.2 (a)}}{=} E(X_1) + \dots + E(X_n)$$

(b)  $X \sim N(\mu, \sigma^2)$

$X$  hat dieselbe Verteilung wie  $\mu + \sigma Y$ , wobei  $Y \sim N(0, 1)$

$$\Rightarrow E(X) = E(\mu + \sigma Y) = \mu + \sigma E(Y)$$

$$E(Y) = \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 0, \text{ da } ye^{-\frac{y^2}{2}} \text{ ungerade.}$$

**4.1.3 Satz 4.3:**

Seien  $X$  eine ZV und  $g: \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion. Dann gilt

$$E[g(X)] = \begin{cases} \sum_i g(x_i) P(X = x_i) & \text{falls } X \text{ diskret ist mit möglichen Werten } x_1, x_2, \dots \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{falls } X \text{ stetig ist und Dichte } f \text{ hat} \end{cases}$$

**Bsp. 4.5:**  $X \sim N(0, 1)$

$$\begin{aligned} E(X^2) &\stackrel{\text{Satz 4.3 mit } g(x)=x^2}{=} \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x)(-xe^{-\frac{x^2}{2}}) dx \\ &= \frac{1}{\sqrt{2\pi}} ([-xe^{-\frac{x^2}{2}}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx) \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

partielle Integration

$$\begin{aligned} u(x) &= -xv(x) = e^{-\frac{x^2}{2}} \\ u'(x) &= -1v'(x) = -xe^{-\frac{x^2}{2}} \end{aligned}$$

**4.2 Varianz einer ZV**

Die *Varianz* einer ZV  $X$  ist def. durch

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

und die *Standardabweichung* von  $X$  ist

$$\sigma(X) := \sqrt{\text{Var}(X)}$$

$\text{Var}(X)$  und  $\sigma(X)$  sind Maße für die Streuung der Verteilung von  $X$  um  $E(X)$

#### 4.2.1 Satz 4.4:

Sei  $X$  eine ZV. Dann gilt

- (a)  $Var(X) = E(X^2) - [E(X)]^2$
- (b) Ist  $P(X=a)=1$  für eine Konstante  $A \in \mathbb{R}$ , dann  $Var(X) = c$ .  
Ist  $Var(X) = c$ , dann folgt  $P(X = E[X]) = 1$ , d.h. mit W'keit 1 ist  $X$  gleich der Konstanten  $E(X)$ .
- (c)  $Var(a + bX) = Var(bX) = b^2 Var(X)$  für alle  $a, b \in \mathbb{R}$
- (d) Sei  $Var(X) \in (0, \infty)$ . Für die standardisierte ZV

$$X* = \frac{X - E(X)}{\sqrt{Var(X)}}$$

gilt  $E(X*) = 0, Var(X*) = 1$ .

- (e) Sind  $X_1, \dots, X_n$  unabhängige ZV, dann gilt

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Beweis nur für (a):

$$\begin{aligned} Var(X) &= E[(X - E(X))^2] \\ &= E[X^2 - 2XE(X) + [E(X)]^2] \\ &= E(X^2) - 2E[XE(X)] + E[[E(X)]^2] \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

**Bsp. 4.6:** (a)  $X \sim \text{Ber}(p)$

Da  $X = 0$  oder  $X = 1$ , folgt  $X^2 = X$ .

$$\Rightarrow Var(X) \stackrel{\text{Satz 4.4 (a)}}{=} E(X^2) - [E(X)]^2 = E(X) - [E(X)]^2 \stackrel{\text{Bsp. 4.1 (c)}}{=} p - p^2 = p(1 - p) \quad (\text{b})$$

$X \sim \text{Bin}(n, p)$

$\Rightarrow X = X_1 + \dots + X_n, X_i$  unabhängig,  $X_i \sim \text{Ber}(p)$

$$\Rightarrow Var(X) = Var(X_1 + \dots + X_n) \stackrel{\text{Satz 4.4 (e)}}{=} Var(X_1) + \dots + Var(X_n) = np(1 - p) \quad (\text{b})$$

$X \sim N(\mu, \sigma^2), Var(X) = ?$

Sei  $Y \sim N(0, 1)$

$\Rightarrow X$  hat dieselbe Verteilung wie  $\mu + \sigma Y$

$$\Rightarrow Var(X) = Var(\mu + \sigma Y) \stackrel{\text{Satz 4.4 (c)}}{=} \sigma^2 Var(Y) = \sigma^2 (E(Y^2) - (E(Y))^2) = \sigma^2$$

#### 4.3 Kovarianz und Korrelation

Sind  $X$  und  $Y$  ZV mit EW  $\mu_x = E(X), \mu_y = E(Y) \in \mathbb{R}$ .

Die Kovarianz von  $X$  und  $Y$  ist def. durch

$$Cov(X, Y) := E[(X - \mu_x)(Y - \mu_y)]$$

Der *Korrelationskoeffizient* von  $X$  und  $Y$  ist def. durch

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

falls  $\text{Var}(X), \text{Var}(Y) \in (0, \infty)$

$\rho(X, Y)$  ist ein Maß für den linearen Zusammenhang zwischen  $X$  und  $Y$ .

Es gilt

$$-1 \leq \rho(X, Y) \leq 1$$

$\rho(X, Y) = 1 \Leftrightarrow$  es ex.  $A \in \mathbb{R}, b > 0$  mit  $P(Y = a + bX) = 1$ .  $\rho(X, Y) = -1 \Leftrightarrow$  es ex.  $A \in \mathbb{R}, b < 0$   $P(Y = a + bX) = 1$ .

$X$  und  $Y$  heißen *unkorreliert*, falls  $\rho(X, Y) = 0$ .

Es gilt

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY - \mu_x Y - X\mu_y + \mu_x \mu_y) \\ &= \end{aligned}$$

$\Rightarrow$  Sind  $X$  und  $Y$  unabhängig, dann sind  $X$  und  $Y$  unkorreliert. Umkehrung gilt nicht.

**Bsp. 4.7:** Sei  $P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}, Y := X^2$

$\Rightarrow E(X) = 0, E(XY) = E(X^3) = 0$ , also

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

$X$  und  $Y$  sind unkorreliert, aber

$$P(X = 0|Y = 0) = 1 \neq P(X = 0)$$

$X$  und  $Y$  sind nicht unabhängig.

#### 4.3.1 Satz 4.5:

Seien  $X, Y, X_1, \dots, X_n, Y_1, \dots, Y_n$  ZV

$$(a) \text{Cov}(X, Y) = \text{Cov}(Y, X)$$

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

$$(c) \text{Cov}(a + bX, c + dY) = bd\text{Cov}(X, Y) \text{ für alle } a, b, c, d \in \mathbb{R}$$

$$(d) \text{Cov}(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j) \quad \forall a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{R}$$

$$(e) \text{Var}(\sum_{i=1}^m X_i) = \sum_{i=1}^m \text{Var}(X_i) + 2 \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

**Bew. (c):**

$$\begin{aligned} &\text{Cov}(a + bX, c + dY) \\ &= E[(a + bX - E(a + bX))(c + dY - E(c + dY))] \\ &= E[b(X - E(X))d(Y - E(Y))] \\ &= b \cdot d \cdot \text{Cov}(X, Y) \end{aligned}$$

**Bew. (d):**

$$\begin{aligned}
& Cov\left(\sum_i a_i X_i, \sum_j b_j X_j\right) \\
&= E\left(\sum_i \sum_j a_i b_j X_i Y_j\right) - \left(\sum_i a_i E(X_i)\right)\left(\sum_j b_j E(Y_j)\right) \\
&= \sum_i \sum_j a_i b_j E(X_i Y_j) - a_i b_j E(X_i) E(Y_j) \\
&\stackrel{(a)}{=} \sum_i \sum_j a_i b_j Cov(X_i, Y_j)
\end{aligned}$$

**Bew. (e):**

$$\begin{aligned}
Var\left(\sum_{i=1}^m X_i\right) &\stackrel{(b)}{=} Cov\left(\sum_{i=1}^m X_i, \sum_{j=1}^m X_j\right) \\
&\stackrel{(d)}{=} \sum_{j=1}^m Cov(X_i, X_j) \\
&= \sum_{i=1}^m Cov(X_i, X, i) + \sum_{i \neq j} Cov(X_i, X_j) \\
&= \sum_{i=1}^m Var(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)
\end{aligned}$$

## 4.4 Ungleichungen

### 4.4.1 Satz 4.6: Markov-Ungleichung

Sei  $X$  eine nicht-negative ZV. Dann gilt

$$P(X \geq t) \leq \frac{E(X)}{t} \quad \forall t > 0$$

**Beweis:** Definiere eine Bernoulli-ZV durch

$$\begin{aligned}
Y(\omega) &:= \begin{cases} 1 & \text{falls } X(\omega) \geq t \\ 0 & \text{falls } X(\omega) < t \end{cases} \\
\Rightarrow Y(\omega) &\leq \frac{X(\omega)}{t} \quad \forall \omega
\end{aligned}$$

- Für  $X(\omega) < t$  gilt  $Y(\omega) = 0 \leq \frac{X(\omega)}{t}$
- Für  $X(\omega) \geq t$  gilt  $Y(\omega) = 1 \leq \frac{X(\omega)}{t}$

$$\Rightarrow P(X \geq t) \stackrel{Bsp. 4.1}{=} E(Y) \stackrel{Satz 4.2(d)}{\leq} E\left(\frac{X}{t}\right) \stackrel{Satz 4.2(a)}{=} \frac{1}{t} E(X)$$

**Bsp. 4.8:** Sei  $X \sim \text{Poi}(1)$  und  $t > 1$ .

$$P(X \geq t) \underset{\text{Markov}}{\leq} \frac{E(X)}{t} \underset{\text{Bsp. 4.1(d)}}{=} \frac{1}{t}$$

Die obere Schranke lässt sich verbessern. Für jedes  $u > 0$  gilt

$$P(X \geq t) = P(e^{uX} \geq e^{ut}) \underset{\text{Markov}}{\leq} \frac{E(e^{uX})}{e^{ut}} \underset{(\text{monotone Transf.})}{}{}$$

und

$$E(e^{uX}) \underset{\text{Satz 4.3}}{=} e^{-1} \sum_{k=0}^{\infty} \frac{(e^u)^k}{k!} = e^{-1} \exp(e^u - 1)$$

$$\Rightarrow P(X \geq t) \leq \exp(e^u - 1 - ut) \quad \forall u > 0$$

$e^u - 1 - ut$  wird minimal für  $u = \log(t)$  und für dieses  $u$  ergibt sich

$$P(X \geq t) \leq \exp(t - 1 - t \cdot \log(t)) = \left(\frac{e}{t}\right)^t \cdot e^{-1}$$

z.B.  $P(X \geq 5) \leq 0,01747, P(X \geq 10) \leq 8,103 \cdot 10^{-7}$

#### 4.4.2 Satz 4.7: Tschebyscheff-Ungleichung

Sei  $X$  eine ZV,  $\mu := E(X) \in \mathbb{R}$  und  $\sigma^2 := \text{Var}(X)$ . Dann gilt

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \quad \forall t > 0$$

Ist  $\sigma = \sqrt{\text{Var}(X)} \in (0, \infty)$ , dann gilt

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \forall k > 0$$

**Beweis:**

$$P(|X - \mu| \geq t) = P((X - \mu)^2 \geq t^2) \underset{\text{Markov}}{\leq} \frac{E[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}$$

Mit  $t = k\sigma$  ergibt sich die andere Behauptung.

**Bsp. 4.9:**  $n = 10.000$  Münzwürfe  $X = \text{Anzahl der Würfe, in denen Zahl fällt} \sim \text{Bin}(n, \frac{1}{2})$

$$\begin{aligned} P(4.750 < X < 5.250) &= P(|X - 5000| < 250) \\ &= 1 - P(|X - E(X)| \geq 250) \\ &\underset{\text{Tschebyscheff}}{\geq} 1 - \frac{\text{Var}(X)}{250^2} = 1 - \frac{\frac{n}{4}}{250^2} \\ &= 0,96 \end{aligned}$$

**Bsp. 4.10:** Sei  $x$  eine ZV,  $\mu = E(X), \sigma = \sqrt{\text{Var}(X)} \in (0, \infty)$

Für den  $k\sigma$ -Bereich von  $X$

$$\mu - k\sigma, \mu + k\sigma$$



gilt

$$P(X \notin [\mu - k\sigma, \mu + k\sigma]) \leq \frac{1}{k^2}$$

und

$$P(X \in [\mu - k\sigma, \mu + k\sigma]) \geq 1 - \frac{1}{k^2}, k = 2, 3, \dots$$

W'keit, dass eine ZV ... das k-fache ihrer Standardabweichung von ihrem EW ... mind.  $1 - \frac{1}{k^2}$ .

Dies gilt für große Verteilungen .

Zum Vergleich: Ist  $X \sim N(\mu, \sigma^2)$ , dann

$$\begin{aligned} P(X \in [\mu - k\sigma, \mu + k\sigma]) &= P(-k \leq \frac{X - \mu}{\sigma} \leq k) = \Phi(k) - \Phi(-k) \\ (\text{Tutorium A40}) &= 2\Phi(k) - 1 \\ &\approx \begin{cases} 0,6827, k = 1 \\ 0,9545, k = 2 \\ 0,9973, k = 3 \end{cases} \end{aligned}$$

## 5 Grenzwertsätze

Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte ZV,  $\mu = E(X_1), \sigma = \sqrt{Var(X_1)}$   
Untersuche das Verhalten des Stichprobenmittel:

$$\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$$

Für große N

$$\begin{aligned} E(\bar{X}_n) &= \frac{1}{n}(E(X_1) + \dots + E(X_n)) = \mu \\ Var(\bar{X}_n) &= \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n)) = \frac{\sigma^2}{n} \rightarrow 0 (\text{für } n \rightarrow \infty) \end{aligned}$$

Intuitiv, Verteilung von  $\bar{X}_n$  konzentriert sich für  $n \rightarrow \infty$  in der Nähe von  $\mu$ ,

$\bar{X}_n$  konvergiert gegen  $\mu$  (Gesetz d. großen Zahlen)

Zentraler Grenzwertsatz beschreibt die Verteilung der standardisierten ZV

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

für  $n \rightarrow \infty$

### 5.1 Gesetz der großen Zahlen

Wichtiger Spezialfall:  $X_n \xrightarrow{P} a$  für eine Konstante  $a$ .  
d.h.

$$\lim_{n \rightarrow \infty} P(a - \epsilon < X_n < a + \epsilon) = 1 \forall \epsilon > 0$$

**Bsp. 5.1** Seien  $U_1, U_2, \dots$  unabhängig,  $U_i \sim U_{ni}(0, 1)$

$$X_n := \min\{U_1, \dots, U_n\}, n \in \mathbb{N}$$

Beh.  $X_n \xrightarrow{P} 0$   
 Für jedes  $\varepsilon > 0$  gilt

$$P(|X_n - 0| \geq \varepsilon) = P(X_n \geq \varepsilon) = \begin{cases} (1 - \varepsilon)^2, & \text{falls } 0 < \varepsilon < 1 \\ 0, & \text{falls } \varepsilon \geq 1 \end{cases}$$

**Def. 5.1:** Eine Folge von ZV  $X_1, X_2, \dots$  konvergiert stochastisch gegen eine ZV  $X$ , geschrieben  $X_n \xrightarrow{P} X$ , falls für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

gilt.

### 5.1.1 Satz 5.1: Schwaches Gesetz der großen Zahlen

Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte ZV mit  $\mu = E(X_1)$  und  $Var(X_1) < \infty$ . Dann folgt

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$$

**Bew.:**

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

Für jedes  $\varepsilon > 0$  gilt

$$\begin{aligned} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \\ (\text{Tschebyscheff}) &\leq \frac{Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}{\varepsilon^2} \stackrel{\text{Satz 4.4}}{=} \frac{1}{n^2 \varepsilon^2} = \sum_{i=1}^n \frac{Var(X_i)}{n^2 \varepsilon^2} \\ &= \frac{Var(X_1)}{n \varepsilon^2} \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

**Bsp. 5.2:** Seien  $X_1, X_2, \dots$  unabhängig, identisch verteilte ZV,  $X_i \sim \text{Ber}(p)$   
 $\rightarrow E(X_1) = p$  und nach Satz 5.1 gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} p$$

relative Häufigkeit der Erfolge in den ersten  $n$  Experimenten  $\rightarrow$  Erfolgsw'keit

**Bsp. 5.3:** Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte ZV, sei  $A \in \mathbb{R}$  Wende Satz 5.1 an auf Bernoulli-ZV

$$\begin{aligned} Y_i(\omega) &= \begin{cases} 1 & \text{falls } X_i(\omega) \in A \\ 0 & \text{falls } X_i(\omega) \notin A \end{cases} \\ \Rightarrow \frac{1}{n} \sum_{i=1}^n Y_i &\xrightarrow{P} E(Y_1) = P(X_1 \in A) \end{aligned}$$

relative Häufigkeit von „ $X_i \in A$ “ in den ersten  $n$  Experimenten  $\rightarrow$  W'keit von „ $X_i \in A$ “

### 5.1.2 Satz 5.2: Starkes Gesetz der großen Zahlen

Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte ZV mit  $\mu \in E(X_1) \in \mathbb{R}$ . Dann gilt

$$P(\{\omega \in \Omega : n \xrightarrow{\lim} \infty \frac{1}{n}(X_1(\omega) + \dots + X_n(\omega) = \mu)\}) = 1$$

## 5.2 Zentraler Grenzwertsatz

### 5.2.1 Satz 5.2 Zentraler Grenzwertsatz ZGS

Seien  $X_1, X_2, \dots$  unabhängige, identisch verteilte ZV. Sei

$$\begin{aligned}\sigma^2 &= \text{Var}(X_1) \in (0, \infty) \\ \mu &= E(X_1) \\ S_n &= X_1 + \dots + X_n \\ S_n^* &= \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}\end{aligned}$$

Dann gilt

$$n \xrightarrow{\lim} \infty P(a \leq S_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a), -\infty \leq a \leq b \leq \infty$$

wobei  $\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{x^2}{2}} dx = \text{VF}$  einer  $N(0,1)$ -verteilten ZV ( $\Phi(-\infty) = 0, \Phi(\infty) = 1$ )

**Bem. 5.1:** Unter den Voraussetzungen von Satz 5.3 gilt auch

$$n \xrightarrow{\lim} \infty P(a < S_n^* < b) = n \xrightarrow{\lim} \infty P(a < S_n^* \leq b) = n \xrightarrow{\lim} \infty P(a \leq S_n^* < b) = \Phi(b) - \Phi(a)$$

für  $-\infty \leq a < b \leq \infty$ .

ZGS kann oft benutzt werden, um W'keiten, die mit Summen von unabhängigen, identisch verteilten ZV gebildet werden, zu approximieren.

**Bsp. 5.4:** Normalverteilungsapproximation der Binomialverteilung Seien  $X_1, X_2, \dots$  unabhängig  $\sim \text{Ber}(p)$ ,  $0 < p < 1 \rightarrow S_n = X_1 + \dots + X_n \sim \text{Bin}(n, p)$ ,  $E(S_n) = np$ ,  $\text{Var}(S_n) = np(1-p)$

$$\rightarrow P(a \leq S_n^* \leq b) = P(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) \approx \Phi(b) - \Phi(a)$$

Faustregel: Approximation ist anwendbar falls  $np \geq 5$  und  $n(1-p) \geq 5$ .

**Bsp. 5.5:** Sei  $X \sim \text{Bin}(200, \frac{1}{20})$ ,  $P(5 \leq X \leq 15) = ?$

(a) Exakte Lösung

$$P(5 \leq X \leq 15) = \sum_{k=5}^{15} \binom{200}{k} \left(\frac{1}{20}\right)^k \left(\frac{19}{20}\right)^{200-k} = 0,9292$$

(b) Normalverteilungsapproximation  $X = S_n \sim \text{Bin}(n, p)$  mit  $n = 200$  und  $p = \frac{1}{20}$

Approximation  $P(5 \leq S_n \leq 15)$  wie in Bsp. 5.4

Bringe gesuchte W'keit auf die Form

$$P(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b)$$

wobei  $np = 10$  und  $\sqrt{np(1-p)} = \sqrt{\frac{19}{2}} = 3,0822$

$$\begin{aligned} P(5 \leq S_n \leq 15) &= P\left(\frac{5 - np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{15 - np}{\sqrt{np(1-p)}}\right) \\ &= P(-1,6222 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq 1,6222) \end{aligned}$$

(Über ZGS)  $\approx \Phi(1,6222) - \Phi(-1,6222)$

$\Phi(-x) = 1 - \Phi(x)$  für alle  $x \in \mathbb{R}$  s. Tutorium A40

$$\Rightarrow P(5 \leq S_n \leq 15) \approx 2\Phi(1,6222) - 1 \approx 2 \cdot 0,9474 - 1 \approx 0,8948$$

## Teil II

# Schließende Statistik

## 6 Parameterschätzung

Sei  $X$  eine ZV mit (ganz oder teilweise) unbekannter Verteilung (die Verteilung der Grundgesamtheit).  
Gesucht: Information über die Verteilung von  $X$ .

Betrachte dazu (*Zufalls-*)*Stichprobe* vom Umfang  $n$ :  $(X_1, \dots, X_n)$  D.h. Stichprobenvariablen  $(X_1, \dots, X_n)$  sind unabhängige ZV, die jeweils dieselbe Verteilung haben wie  $X$ .

$n$  unabhängige Wiederholungen von  $X$ .

Konstruiere mit Hilfe der  $X_1, \dots, X_n$  Schätzer für interessierenden Parameter der Verteilung, z.B.  $\vartheta = E(X)$ , oder  $\vartheta = Var(X)$ , oder, falls bekannt ist, dass  $X \sim \text{Exp}(\lambda)$ , aber  $\lambda$  unbekannt,  $\vartheta = \lambda$ .

### 6.1 Erwartungstreue und Konsistenz

Sei  $(X_1, \dots, X_n)$  eine Stichprobe. Sei  $T_{\mathbb{R}}^n \rightarrow \mathbb{R}^p$ . Dann heißt die Zufallsvariable

$$T(X_1, \dots, X_n)$$

eine *Stichprobenfunktion* oder *Statistik*. Wird  $T$  zum Schätzen eines Parameters  $\vartheta \in \mathbb{R}^p$  verwendet, so nennt man  $T(X_1, \dots, X_n)$  eine *Schätzfunktion* für (das Schätzen von)  $\vartheta$ .

In Anwendungen: Erhebe Stichprobe, d.h. Beobachtungswerte  $x_1, \dots, x_n \in \mathbb{R}$  = Realisationen von  $X_1, \dots, X_n$ .

$T(x_1, \dots, x_n)$  ist ein Schätzwert (keine ZV) für  $\vartheta$ .

Im Folgenden gelte für

$\Theta$  = Menge der möglichen Parameterwerte

$\Theta \in \mathbb{R}$  eindimensional

**Def. 6.1:** Eine Schätzfunktion  $\hat{\vartheta} = T(X_1, \dots, X_n)$  für  $\vartheta \in \Theta \in \mathbb{R}$  heißt *erwartungstreu* (e-treu) für  $\vartheta$ , falls

$$E(\hat{\vartheta}) = \vartheta \quad \forall \vartheta \in \Theta$$

Der *Bias* oder die *Verzerrung* von  $\hat{\vartheta}$  ist definiert durch

$$\text{Bias}(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta$$

**Bsp. 6.1:** Schätze  $\mu = E(X)$ .

Für das Stichprobenmittel

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

gilt

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

So ist  $\bar{X}_n$  e-treue Schätzfunktion für  $\mu = E(X)$ .

$\bar{X}_n$  hat unter allen e-treuen linearen Schätzfunktionen also Schätzfunktionen der Form

$$T(X_1, \dots, X_n) = \sum_{i=1}^n c_i X_i \quad \text{mit } c_1, \dots, c_n \in \mathbb{R}$$

die kleinste Varianz, denn

Sei  $T(X_1, \dots, X_n) = \sum_{i=1}^n c_i X_i$  e-treu für  $\mu$

$$\Rightarrow \mu = E[T(X_1, \dots, X_n)] = \sum_{i=1}^n c_i E(X_i) = \mu \sum_{i=1}^n c_i, \text{ also } \sum_{i=1}^n c_i = 1$$

und mit  $\sigma^2 = \text{Var}(X_i)$  gilt

$$\begin{aligned} \text{Var}(T(X_1, \dots, X_n)) &= \sum_{i=1}^n c_i^2 \text{Var}(X_i) \\ &= \sigma^2 \sum_{i=1}^n (c_i - \frac{1}{n} + \frac{1}{n})^2 \\ &= \sigma^2 \left[ \sum_{i=1}^n (c_i - \frac{1}{n})^2 + \frac{2}{n} \sum_{i=1}^n (c_i - \frac{1}{n}) + \frac{n}{n^2} \right] \\ &\geq \frac{\sigma^2}{n} = \text{Var}(\bar{X}_n) \end{aligned}$$

**Bsp. 6.2:** Schätze  $\sigma^2 = \text{Var}(X)$

$$\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Ist  $\tilde{S}^2$  e-treu für  $\sigma^2$ ?

$$\begin{aligned} E(\tilde{S}^2) &= \frac{1}{n} E\left[\left(\sum_{i=1}^n X_i^2\right) - 2\bar{X}_n\left(\sum_{i=1}^n X_i\right) + n\bar{X}_n^2\right] \\ &= \frac{1}{n} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}_n^2)\right) \end{aligned}$$

$$E(X_i^2) = \text{Var}(X_i) + [E(X_i)]^2 = \sigma^2 + \mu^2$$

$$E(\bar{X}_n^2) = \text{Var}(\bar{X}_n) + [E(\bar{X}_n)]^2 = \frac{\sigma^2}{n} + \mu^2$$

$$\Rightarrow E(\tilde{S}^2) = \frac{1}{n}(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)) = \frac{n-1}{n}\sigma^2$$

$\tilde{S}^2$  ist nicht e-treu für  $\sigma^2$ .

$$Bias(\tilde{S}^2) = E(\tilde{S}^2) - \sigma^2 = -\frac{1}{n}\sigma^2$$

Aber: Die *Stichprobenvarianz*

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, n \geq 2$$

ist e-treu für  $\sigma^2 = Var(X)$ , denn  $E(S_n^2) = E(\frac{n}{n-1}S^2) = \sigma^2$

**Def. 6.2:** Die *mittlere quadratische Abweichung* (oder der mittlere quadratische Fehler) einer Schätzfunktion  $\hat{\vartheta}$  für  $\vartheta$  ist definiert durch

$$MSE(\hat{\vartheta}) := E[(\hat{\vartheta} - \theta)^2]$$

**Bem. 6.1:**

(a) Ist  $\hat{\vartheta}$  e-treu für  $\theta$ , dann gilt

$$MSE(\hat{\vartheta}) := E[(\hat{\vartheta} - \theta)^2] = Var(\hat{\vartheta})$$

In Bsp. 6.1

$$MSE(\bar{X}_n) = Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

(b) Für bel. Schätzfunktionen  $\hat{\vartheta}$  gilt

$$MSE(\hat{\vartheta}) = [Bias(\hat{\vartheta})]^2 + Var(\hat{\vartheta})$$

denn mit  $t := E(\hat{\vartheta})$  gilt

$$\begin{aligned} MSE(\hat{\vartheta}) &= E[(\hat{\vartheta} - t + t - \theta)^2] \\ &= E[(\hat{\vartheta} - t)^2] + 2(t - \vartheta)E(\hat{\vartheta} - t) + (t - \vartheta)^2 \\ &= Var(\hat{\vartheta}) + [Bias(\hat{\vartheta})]^2 \end{aligned}$$

### 6.1.1 Konsistenz

Seien  $X_1, X_2, \dots$  unabh. identisch verteilte ZV.

Betrachte Folge von Schätzfunktionen  $f \rightarrow$  der Schätzer eines Parameters:

$$T_1(X_1), T_2(X_1, X_2), T_3(X_1, X_2, X_3)$$

Für jedes  $n$  sei  $T_n(X_1, \dots, X_n)$  Schätzfunktion basierend auf Stichprobe vom Umfang  $n$ .

Wird für  $n \rightarrow \infty$  Schätzgenauigkeit bel. gut?

**Def. 6.3:** Eine Folge von Schätzfunktionen für  $\vartheta$ :

$$T_1(X_1), T_2(X_1, X_2),$$

heißt *konsistent* für den Parameter  $\vartheta$ , falls

$$T_n(X_1, \dots, X_n) \xrightarrow{P} \vartheta \quad (n \rightarrow \infty)$$

**Bsp. 6.3:** Seien  $X_1, X_2, \dots$  unabh., identisch verteilte ZV mit unbekanntem EW  $\mu = E(X_i)$ . Nach dem schwachen Gesetz der großen Zahlen gilt

$$\bar{X}_n \xrightarrow{P} \mu$$

Also ist die Folge  $(\bar{X}_n)_{n=1}^\infty$  konsistent für  $\mu$ .

**Bem. 6.2:** Seien  $X_1, X_2, \dots$  unabh., identisch verteilte ZV,

$$\hat{\vartheta}_n = T_n(X_1, \dots, X_n)$$

und  $MSE(\hat{\vartheta}_n) \rightarrow 0 \forall n \rightarrow \infty$ .

$\Rightarrow$  Für jedes  $\varepsilon > 0$  gilt

$$P(|\hat{\vartheta}_n - \vartheta| \geq \varepsilon) = P(|\hat{\vartheta}_n - \vartheta|^2 \geq \varepsilon^2) \underset{\text{Markov}}{\leq} \frac{E(|\hat{\vartheta}_n - \vartheta|^2)}{\varepsilon^2} = \frac{MSE(\hat{\vartheta}_n)}{\varepsilon^2} \rightarrow 0$$

$$\Rightarrow \hat{\vartheta}_n \xrightarrow{P} \vartheta \text{ und } (\hat{\vartheta}_n)_{n=1}^\infty \text{ ist konsistent.}$$

Mit Bem. 6.1(b) folgt:

Falls  $Bias(\hat{\vartheta}_n) \rightarrow 0$  und  $Var(\hat{\vartheta}_n) \rightarrow 0$ , dann ist  $(\hat{\vartheta}_n)_{n=1}^\infty$  konsistent.

## 6.2 Die Maximum-Likelihood-Methode

Methode zur Konstruktion von Schätzfunktionen für unbekannten Parameter einer Verteilung.  
Sei

$$\Theta = \text{Menge der möglichen Parameterwerte.}$$

Sei  $X$  eine ZV, deren Verteilung nur bis auf einen Parameter  $\vartheta$  bekannt ist. D.h. die Verteilung von  $X$  wird beschrieben durch W-Maß  $P_\vartheta$ , das von  $\vartheta$  abhängt.

$X_1, \dots, X_n$  seien unabh. ZV, die jeweils dieselbe Verteilung wie  $X$  haben. Schätze  $\vartheta$  mit Hilfe von  $(X_1, \dots, X_n)$

**Bsp. 6.4:** Es sei bekannt, dass  $X$  eine Poisson-Verteilung hat,  $X \sim \text{Poi}(\lambda)$ , aber  $\lambda$  unbekannt ist.

Schätze  $\vartheta = \lambda$

Diskreter Fall:

Idee: Für Beobachtungswerte  $x_1, \dots, x_n$  der Zufallsstichprobe, wähle als Schätzwert  $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n)$  dasjenige  $\vartheta \in \Theta$ , für das die W'keit,  $x_1, \dots, x_n$  zu beobachten, maximal wird. (ML-Methode)

Sei  $X$  eine diskrete ZV mit W'keitsfunktion  $f(x|\vartheta)$ , d.h.  $P_\vartheta(X = x) = f(x|\vartheta)$ .

Dann gilt

$$\begin{aligned} P_\vartheta(X_1 = x_1, \dots, X_n = x_n) &= P_\vartheta(X_1 = x_1) \cdot \dots \cdot P_\vartheta(X_n = x_n) \\ &= f(x_1|\vartheta) \cdot \dots \cdot f(x_n|\vartheta) \text{ für alle } x_1, \dots, x_n \end{aligned}$$

Die Funktion

$$L(\vartheta) = L(\vartheta|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\vartheta)$$

heißt *Likelihood-Funktion*. Sie wird als Funktion von  $\vartheta$  aufgefasst bei geg. Werten  $x_1, \dots, x_n$ .

**Def. 6.4:** Für Stichprobenrealisation  $(x_1, \dots, x_n)$  sei  $\hat{\vartheta} = \hat{\vartheta}(x_1, \dots, x_n) \in \Theta$  so, dass

$$(*) \quad L(\hat{\vartheta}|x_1, \dots, x_n) = \max_{\vartheta \in \Theta} L(\vartheta|x_1, \dots, x_n)$$

Dann ist

$\hat{\vartheta}(x_1, \dots, x_n)$  ein *ML-Schätzwert* für  $\vartheta$ .

$\hat{\vartheta}(X_1, \dots, X_n)$  eine *ML-Schätzfunktion* für  $\vartheta$ .

**Bem. 6.3**

- (a) Im Maximierungsproblem  $(*)$  sind  $x_1, \dots, x_n$  fest und es wird bzgl.  $\vartheta$  maximiert.
- (b) Evtl. existiert kein  $\hat{\vartheta} \in \Theta$  oder es existieren mehrere  $\hat{\vartheta} \in \Theta$ , für die  $(*)$  gilt.
- (c) Ist  $L(\vartheta|x_1, \dots, x_n)$  differenzierbar nach  $\vartheta$ , so findet man  $\hat{\vartheta}$  oft durch Differenzieren.  
Ist  $\vartheta \in \Theta \subset \mathbb{R}$ , betrachte

$$\frac{\partial}{\partial \vartheta} L(\vartheta|x_1, \dots, x_n) = 0$$

Beachte: Maximum kann am Rand liegen  $\Theta = [a, b]$

Ist  $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta \subset \mathbb{R}^2$ , betrachte die Gleichungen

$$\frac{\partial}{\partial \vartheta_1} L(\vartheta_1, \vartheta_2|x_1, \dots, x_n) = 0, \quad \frac{\partial}{\partial \vartheta_2} L(\vartheta_1, \vartheta_2|x_1, \dots, x_n) = 0$$

- (d) Oft einfacher:  
Maximierung der Log-Likelihood-Funktion

$$l(\vartheta) = l(\vartheta|x_1, \dots, x_n) = \ln L(\vartheta|x_1, \dots, x_n) \quad (\text{wobei gilt } \frac{d}{dx} \ln x = \frac{1}{x})$$

$l(\vartheta)$  hat an derselben Stelle  $\hat{\vartheta}$  ein Maximum wie  $L(\vartheta)$ , da log streng monoton wachsend ist gilt

$$L(\hat{\vartheta}) \geq L(\vartheta) \Leftrightarrow l(\hat{\vartheta}) \geq l(\vartheta) \forall \vartheta$$

$$\frac{\partial}{\partial \vartheta} \log L(\vartheta|x_1, \dots, x_n) = \frac{\partial}{\partial \vartheta} \log \prod_{i=1}^n f(x_i|\vartheta) = \frac{\partial}{\partial \vartheta} \sum_{i=1}^n \log f(x_i|\vartheta)$$

- (e) Ist  $\hat{\vartheta}$  eine ML-Schätzfunktion für  $\vartheta$  und ist  $g(\vartheta)$  eine Funktion von  $\vartheta$ , dann ist  $g(\hat{\vartheta})$  eine ML-Schätzfunktion für  $g(\vartheta)$ .

**Bsp. 6.5** Sei  $X \sim \text{Poi}(\lambda)$ ,  $\lambda$  unbekannt,  $\lambda \geq 0$ . Bestimt ML-Schätzfunktion für  $\lambda$ . (für Stichprobe vom Umfang  $n$ )

$$f(x|\lambda) = P_\lambda(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \dots$$

$$X \sim \text{Poi}(0) \text{ bedeutet } P_0(X = x) = \begin{cases} 1, & x = 0 \\ 0, & x > 0 \end{cases}$$



1. Bestimme Likelihood-Funktion

$$\begin{aligned} L(\lambda|x_1, \dots, x_n) &= f(x_1|\lambda) \cdots f(x_n|\lambda) \\ &= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} \\ &= \frac{1}{x_1! \cdots x_n!} \lambda^{x_1 + \cdots + x_n} e^{-n\lambda}, \quad \text{wobei } x_1, \dots, x_n \in \mathbb{N}_0 \end{aligned}$$

2. Maximiere Likelihood-Funktion bzgl.  $\lambda$

Betrachte Log-Likelihood-Funktion

$$\begin{aligned} \log L(\lambda|x_1, \dots, x_n) &= \log \frac{1}{x_1! \cdots x_n!} + (x_1 + \cdots + x_n) \log \lambda - n\lambda \\ \frac{\partial}{\partial \lambda} \log L(\lambda|x_1, \dots, x_n) &= (x_1 + \cdots + x_n) \frac{1}{\lambda} - n \stackrel{!}{=} 0 \\ \Leftrightarrow \lambda &= \frac{x_1 + \cdots + x_n}{n} \\ \frac{\partial^2}{\partial \lambda^2} \log L(\lambda|x_1, \dots, x_n) &= -(x_1 + \cdots + x_n) \frac{1}{\lambda^2} < 0 \text{ sofern } x_1 + \cdots + x_n > 0 \end{aligned}$$

$\Rightarrow$  An der Stelle  $\lambda = \frac{x_1 + \cdots + x_n}{n}$  wird  $L(\lambda|x_1, \dots, x_n)$  maximal, d.h.

Falls  $x_1 + \cdots + x_n = 0$ , dann  $x_1 = \cdots = x_n = 0$

$$\Rightarrow L(\lambda|x_1, \dots, x_n) = 1 \cdot \lambda^0 \cdot e^{-n\lambda} = e^{-n\lambda}, \text{ wird maximal f\"ur } \lambda = 0 = \frac{x_1 + \cdots + x_n}{n}$$

$\Rightarrow$  in jedem Fall ist ML-Schätzwert geg. durch

$$\hat{\lambda}(x_1, \dots, x_n) = \frac{x_1 + \cdots + x_n}{n} \quad \text{ML-Schätzfunktion ist } \frac{x_1 + \cdots + x_n}{n}$$

Stetiger Fall

ZV  $X$  habe Dichte  $f(x|\vartheta)$  mit unbekanntem Parameter,  $\vartheta \in \Theta$ .

Likelihood-Funktion ist dann definiert durch

$$L(\vartheta) = L(\vartheta|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\vartheta)$$

und die Definition von ML-Schätzwert und ML-Schätzfunktion ist analog zu Def. 6.4. Auch Bem. 6.3 gilt analog.

**Bsp. 6.6:**  $X \sim N(\mu, \sigma^2)$ . Zunächst sei  $\sigma^2 > 0$  bekannt.

Bestimme ML-Schätzfunktion für  $\mu$ .

Dichte

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Likelihood-Funktion

$$\begin{aligned} L(\mu|x_1, \dots, x_n) &= f(x_1|\mu) \cdots f(x_n|\mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right) \\ \Rightarrow \log L(\mu|x_1, \dots, x_n) &= n \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

Maximiere bzgl.  $\mu$

$$\begin{aligned}\frac{\partial}{\partial \mu} \log L(\mu|x_1, \dots, x_n) &= \frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu) = \frac{1}{\sigma^2} \left[ \left( \sum_{i=1}^n x_i \right) - n\mu \right] \stackrel{!}{=} 0 \\ \Leftrightarrow \mu &= \frac{1}{n} \sum_{i=1}^n x_i\end{aligned}$$

und

$$\begin{aligned}\frac{\partial^2}{\partial \mu^2} \log L(\mu|x_1, \dots, x_n) &= -n < 0 \\ \Rightarrow L(\mu|x_1, \dots, x_n) &\text{ wird maximal f\"ur } \mu = \frac{1}{n} \sum_{i=1}^n x_i\end{aligned}$$

$\Rightarrow$  ML-Sch\"atzfunktion f\"ur  $\mu$  ist  $\frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$  sowohl wenn  $\sigma^2$  bekannt ist als auch wenn  $\sigma^2$  unbekannt ist.

ML-sch\"atzfunktion f\"ur  $\sigma^2$  ist

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 &\text{ falls } \mu \text{ bekannt} \\ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &\text{ falls unbekannt}\end{aligned}$$

ML-Sch\"atzfunktion f\"ur  $\sigma$

$$\begin{aligned}\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2} &\text{ falls } \mu \text{ bekannt} \\ \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} &\text{ falls } \mu \text{ unbekannt}\end{aligned}$$

**Bsp. 6.7:**  $X \sim \text{Uni}(0, \vartheta)$ ,  $\vartheta > 0$  unbekannt

Dichte

$$f(x|\vartheta) = \begin{cases} \frac{1}{\vartheta}, & x \in [0, \vartheta] \\ 0, & \text{sonst} \end{cases}$$

Likelihood-funktion f\"ur  $x_1, \dots, x_n \geq 0$

$$\begin{aligned}L(\vartheta|x_1, \dots, x_n) &= f(x_1|\vartheta) \cdot \dots \cdot f(x_n|\vartheta) = \begin{cases} \left(\frac{1}{\vartheta}\right)^n, & x \in [0, \vartheta] \forall i = 1, \dots, n \\ 0, & \text{sonst} \end{cases} \\ &= \begin{cases} \left(\frac{1}{\vartheta}\right)^n, & \text{falls } \vartheta \geq \max x_1, \dots, x_n \\ 0, & \text{sonst} \end{cases}\end{aligned}$$

## 7 Konfidenzintervalle

Ziel: Gebe (m\"oglichst kurzes) Intervall an, das interessierenden unbekannten Parameter mit hoher W'keit enth\"alt.

Sei  $X$  ZV,  $(X_1, \dots, X_n)$  Stichprobe,  $\vartheta$  unbekannter Parameter der Verteilung von  $X$ ,  $\vartheta \in \Theta \in \mathbb{R}$

Sei  $\alpha \in (0, 1)$  eine vorgegebene Irrtumsw'keit.  
Bestimme Stichprobenfunktionen

$$A = \alpha(X_1, \dots, X_n) \text{ und } B = \beta(X_1, \dots, X_n)$$

so dass

$$\alpha(x_1, \dots, x_n) \leq \beta(x_1, \dots, x_n) \text{ f\"ur alle } x_1, \dots, x_n$$

und so dass das Zufallsintervall  $[A, B]$  den Parameter  $\vartheta$  mit W'keit  $1 - \alpha$  enth\"alt, d.h.

$$P(A \leq \vartheta \leq B) = 1 - \alpha, \quad \vartheta \in \Theta$$

Das Zufallsintervall  $[A, B]$  hei\u00dft dann ein  $(1 - \alpha)$ -Konfidenzintervall f\"ur  $\vartheta$  oder ein Konfidenzintervall f\"ur  $\vartheta$  zum Konfidenzniveau  $1 - \alpha$ .

$1 - \alpha$  ist die *\"Uberdeckungs w'keit* von  $[A, B]$ .

Ein Konfidenzintervall  $[A, B]$  hei\u00dft *symmetrisch*, falls

$$P(A > \vartheta) = P(B < \vartheta)$$

Allgemeiner Ansatz:  $\alpha \in (0, 1)$

$$A = a(X_1, \dots, X_n), B = b(X_1, \dots, X_n)$$

$[A, B]$  ist ein  $(1 - \alpha)$ -Konfidenzintervall f\"ur  $\vartheta$ , falls

$$P(A \leq \vartheta \leq B) = 1 - \alpha$$

symmetrisch, falls

$$P(A > \vartheta) = P(B < \vartheta)$$

Die Realisierung des Konfidenzintervalls f\"ur Beobachtungswerte  $x_1, \dots, x_n$  ist das Sch\"atzintervall

$$a(x_1, \dots, x_n), b(x_1, \dots, x_n)$$

Konstruiere Konfidenzintervall f\"ur den unbekannten Parameter  $\mu$  einer  $N(\mu, \sigma^2)$ -Verteilung

Fall A:  $\sigma^2$  bekannt

Benutze folgende Eigenschaften:

- (i) Sind  $X_1, \dots, X_n$  unabh\"angig und jeweils  $\sim N(\mu, \sigma^2)$ , dann

$$\bar{X}_n = \frac{x_1 + \dots + x_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ vgl. Tut. 9, A49 f\"ur } n = 2$$

und

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1)$$

- (ii) Sei  $\Phi$  die VF eines  $N(0, 1)$ -verteilten ZV

Dann gibt es f\"ur jedes  $0 < \alpha < 1$  eine eindeutige Zahl  $x_\alpha \in \mathbb{R}$ , so dass  $\Phi(x_\alpha) = \alpha$ .  $x_\alpha$  hei\u00dft das  $\alpha$ -Quantil oder  $\alpha$ -Fraktil der  $N(0, 1)$ -Verteilung.

Sei nun  $X \sim N(\mu, \sigma_0^2)$  ( $\mu$  unbekannt,  $\sigma_0^2 > 0$  bekannt). Sei  $(X_1, \dots, X_n)$  Stichprobe  
Konstruiere Konfidenzintervall f\"ur  $\mu$  symmetrisch zur ML-Sch\"atzfunktion  $\bar{X}_n$

Betrachte das Intervall der Form  $[\bar{X}_n - \Delta, \bar{X}_n + \Delta]$  und bestimme  $\Delta$   
 Sei  $\alpha \in (0, 1)$  vorgegebene

$$\begin{aligned}
 1 - \alpha &= P(\bar{X}_n - \Delta \leq \mu \leq \bar{X}_n + \Delta) \\
 &= P(-\Delta \leq \bar{X}_n - \mu \leq \Delta) \\
 &= P\left(\frac{-\sqrt{n}}{\sigma_0} \Delta \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0} \leq \frac{\sqrt{n}}{\sigma_0} \Delta\right) \\
 &= \Phi\left(\frac{\sqrt{n}}{\sigma_0} \Delta\right) - \Phi\left(\frac{-\sqrt{n}}{\sigma_0} \Delta\right) \\
 &= 2\Phi\left(\frac{\sqrt{n}}{\sigma_0} \Delta\right) - 1 \\
 \Leftrightarrow 1 - \frac{\alpha}{2} &= \Phi\left(\frac{\sqrt{n}}{\sigma_0} \Delta\right)
 \end{aligned}$$

Sei  $c$  das  $(1 - \frac{\alpha}{2})$ -Quantil der  $N(0,1)$ -Verteilung. Dann muss

$$\begin{aligned}
 c = \frac{\sqrt{n}}{\sigma_0} \Delta \text{ gelten} &\Rightarrow \Delta = \frac{c\sigma_0}{\sqrt{n}} \\
 &\Rightarrow [\bar{X}_n - \frac{c\sigma_0}{\sqrt{n}}, \bar{X}_n + \frac{c\sigma_0}{\sqrt{n}}]
 \end{aligned}$$

ist ein Konfidenzintervall für  $\mu$  zum Konfidenzniveau  $1 - \alpha$ .  
 Es ist symmetrisch, denn

$$P(\bar{X}_n - \frac{c\sigma_0}{\sqrt{n}} > \mu) = P(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0} > c) = 1 - \Phi(c) = \frac{\alpha}{2}$$

und

$$P(\bar{X}_n + \frac{c\sigma_0}{\sqrt{n}} < \mu) = P(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0} < -c) = \Phi(-c) = 1 - \Phi(c) = \frac{\alpha}{2}$$

**Bsp. 7.1** Sei  $X \sim N(\mu, \sigma^2)$  mit  $\sigma_0 = 0,1$ . Stichpunkte vom Umfang  $n = 4$  Ein Schätzintervall für  $\mu$  wird wie folgt berechnet:

$$x_1 = 5.09, x_2 = 5.22, x_3 = 5.02, x_4 = 5.15$$

1. Lege Konfidenzniveau  $1 - \alpha$  fest.

$$1 - \alpha = 0,95$$

2. Bestimme  $(1 - \frac{\alpha}{2})$ -Quantil  $c$  für  $\Phi(x)$

$$\begin{aligned}
 \alpha &= 0,05, 1 - \frac{\alpha}{2} = 0,975 \\
 \Rightarrow c &= \Phi^{-1}(0,975) = 1,96
 \end{aligned}$$

3. Berechne  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n x_i$

$$\bar{X}_n = 5,12$$

4. Berechne  $\frac{c-\sigma_0}{\sqrt{n}}$ .

$$\frac{1,96 - 0,1}{\sqrt{4}} = 0,98$$

5. Gebe Schätzintervall

$$[\bar{X}_n - \frac{c - \sigma_0}{\sqrt{n}}, \bar{X}_n + \frac{c - \sigma_0}{\sqrt{n}}]$$

an.

$$[5,022, 5,218]$$

Fall B:  $\mu$  und  $\sigma^2$  unbekannt

Sei  $X \sim N(\mu, \sigma^2)$  ( $X_1, \dots, X_n$ ) Stichprobe

Konstruiere Konfidenzintervall für  $\mu$  bei unbekanntem  $\sigma$ : Gehe ähnlich vor wie in Fall A, aber ersetze unbekanntes  $\sigma$  durch Stichproben-Standardabweichung

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

Benutze, dass die Verteilung der ZV

$$Y = \sqrt{n} = \frac{\bar{X}_n - \mu}{S_n}$$

eine sogenannte (Student'sche) *t-Verteilung mit  $n-1$  Freiheitsgraden* ist,  $Y \sim t(n-1)$ .

Symm.  $(1-\alpha)$ -Konfidenzintervall für  $\mu$  ist dann geg. durch

$$\bar{X}_n - \frac{aS_n}{\sqrt{n}}, \bar{X}_n + \frac{aS_n}{\sqrt{n}}$$

**Bsp. 7.2:** Verwende Stichprobe aus Bsp. 7.1 und bestimme Schätzintervall für unbekannten Parameter  $\mu$  einer  $N(\mu, \sigma^2)$ -Verteilung, wenn auch  $\sigma$  unbekannt ist.

1. Lege Konfidenzniveau  $1-\alpha$  fest

$$\text{z.B. } 1-\alpha = 0,95$$

2. Bestimme  $(1-\frac{\alpha}{2})$ -Quantil  $a$  der  $t(n-1)$ -Verteilung

$$1 - \frac{\alpha}{2} = 0.975, n-1 = 3 : \text{ Tabelle liefert } a = 3.1824$$

3. Berechne  $\bar{x}_n$  und  $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$

$$\bar{X}_4 = 5.12$$

$$S_4 = \sqrt{\frac{1}{3} \sum_{i=1}^4 (X_i - 5.12)^2}$$

4. Berechne  $\frac{a \cdot s_n}{\sqrt{n}}$

$$\frac{3.1824 \cdot 0.08524}{\sqrt{4}} = 0.1356$$

5. Gebe Schätzintervall

$$[\bar{X}_n - \frac{a \cdot s_n}{\sqrt{n}}, \bar{X}_n + \frac{a \cdot s_n}{\sqrt{n}}]$$

an.

$$[4.9844, 5.2556]$$

## 8 Testverfahren

### 8.1 Einleitung, Definitionen

Formuliere Hypothese  $H_0$  über unbekannte Verteilung. Verwende Ergebnis einer Stichprobe, um zu entscheiden, ob  $H_0$  für richtig gehalten wird bzw. abgelehnt wird. **Bsp. 8.1**

(a) Gegeben sei eine Münze,  $p = P(\text{„Zahl“})$

Hypothese  $H_0 : p = 0,5$ , Münze ist fair

Alternativ  $H_1 : p \neq 0,5$ , Münze ist nicht fair

(b) Sei  $X \sim (N(\mu, \sigma^2))$  mit  $\mu$  unbekannt

Hypothese  $H_0 : \mu \leq 5$

Alternative  $H_1 : \mu > 5$

Mögliche Entscheidungen

1.  $H_0$  wird abgelehnt (und  $H_1$  wird akzeptiert)
2.  $H_0$  wird nicht abgelehnt

Ein *Test* ist ein Verfahren, das für jedes mögliche Ergebnis der Stichprobe entscheidet, ob  $H_0$  abgelehnt wird oder nicht.

Mögliche Fehler:

Fehler 1. Art:  $H_0$  wird abgelehnt, obwohl  $H_0$  richtig ist.

Fehler 2. Art:  $H_1$  wird nicht abgelehnt, obwohl  $H_0$  falsch ist.

Die W'keiten für beide Fehlerarten lassen sich im Allgemeinen nicht gleichzeitig kontrollieren, wenn der Stichprobenumfang gegeben ist.  $\Rightarrow$  Man gibt also eine obere Schranke  $\alpha$  ( $0 < \alpha < 1$ ) für die W'keit eines Fehlers 1. Art vor und minimiert die andere W'keit unter dieser Bedingung.

$\alpha$  heißt (*Signifikanz-*)*Niveau*

Typischerweise  $\alpha = 0,05$  oder  $\alpha = 0,01$ .

#### Allgemeine Testsituation

$\vartheta \in \Theta$  Parameter der Verteilung einer ZV  $X$

Nullhypothese  $H_0$ :  $\vartheta \in \Theta_0$ , wobei  $\Theta_0 \subset \Theta$ ,  $\emptyset \neq \Theta_0 \neq \Theta$ .

Alternative  $H_1$ :  $\vartheta \in \Theta_1 := \Theta \setminus \Theta_0$

In Bsp. 8.1(a):  $\Theta = [0, 1]$

$$\begin{aligned} H_0 : p &= \frac{1}{2} & \Theta_0 &= \left\{ \frac{1}{2} \right\} \\ H_1 : p &\neq \frac{1}{2} & \Theta_1 &= \Theta \setminus \Theta_0 = \left[ 0, \frac{1}{2} \right) \cup \left( \frac{1}{2}, 1 \right] \end{aligned}$$

$(X_1, \dots, X_n)$  sei Zufallsstichprobe.  $X_1, \dots, X_n$  unabhängige ZV, die jeweils dieselbe Verteilung haben wie  $X$ .

Sei  $B$  = Menge aller möglichen Realisationen von  $(X_1, \dots, X_n)$ , also  $B \subset \mathbb{R}^n$

**Test**  $\tau$  wird eindeutig beschrieben durch die Menge  $K = K_\tau \subset B$  Lehne  $H_0$  genau dann ab, wenn  $(X_1(\omega), \dots, X_n(\omega)) \in K$

$K$  heißt *kritischer Bereich* oder *Ablehnungsbereich*

$t$  heißt **Test zum Niveau**  $\alpha$  ( $0 < \alpha < 1$ ), falls

$$P_\vartheta(\tau \text{ lehnt } H_0 \text{ ab}) = P_\vartheta((X_1, \dots, X_n) \in K) \leq \alpha \text{ für alle } \vartheta \in \Theta_0$$

Beachte:  $\vartheta \in \Theta_0 \Rightarrow H_0$  trifft zu

$\Rightarrow$  W'keit eines Fehlers 1. Art

$$= P_\tau(\vartheta) = P_\vartheta(\tau \text{ lehnt } H_0 \text{ ab}) \leq \alpha$$

Die Funktion

$$\beta_\tau(\vartheta) := P_\vartheta(\tau \text{ lehnt } H_0 \text{ ab}), \vartheta \in \Theta$$

heißt **Gütefunktion** von  $\tau$ .

$\tau$  ist Test zum Niveau  $\alpha \Leftrightarrow \beta_\tau(\vartheta) \leq \alpha$  für alle  $\vartheta \in \Theta_0$ .

Erwünscht: Unter dieser Bedingung sollte  $\beta_2(\vartheta)$  für  $\vartheta \in \Theta_1$  möglichst groß sein, da dann W'keit für Fehler 2. Art möglichst klein wird. Denn:  $\vartheta \in \Theta_1 \Rightarrow H_0$  falsch

$\Rightarrow$  W'keit für Fehler 2. Art

$$P_\vartheta(\tau \text{ lehnt } H_0 \text{ nicht ab}) = 1 - \beta_\tau(\vartheta)$$

### 8.1.1 Satz 8.1

Seien  $\tau_1, \tau_2$  Tests zum Niveau  $\alpha$  für  $H_0 : \vartheta \in \Theta_0$  gegen  $H_1 : \vartheta \in \Theta_1$  mit kritischen Bereichen  $K_1, K_2$ .

Gilt  $K_1 \subset K_2$ , dann folgt  $\beta_{\tau_1}(\vartheta) \leq \beta_{\tau_2}(\vartheta)$  für alle  $\vartheta \in \Theta$

**Beweis:**

$$\beta_{\tau_1}(\vartheta) = P_\vartheta(\tau_1 \text{ lehnt } H_0 \text{ ab}) = P_\vartheta((X_1, \dots, X_n) \in K_1)$$

$$K_1 \subset K_2 \leq P_\vartheta((X_1, \dots, X_n) \in K_2) = P_\vartheta(\tau_2 \text{ lehnt } H_0 \text{ ab}) = \beta_{\tau_2}(\vartheta)$$

**Fazit:** Wähle bei geg. Niveau den kritischen Bereich möglichst groß.

## 8.2 Testen von Hypothesen über den Mittelwert einer $N(\mu, \sigma^2)$ -Verteilung

**Fall A:**  $\mu$  unbekannt,  $\sigma^2$  bekannt

$X \sim N(\mu, \sigma^2)$ ,  $X_1, \dots, X_n$  unabhängig, jeweils  $\sim N(\mu, \sigma^2)$

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

wobei  $\mu_0 \in \mathbb{R}$  gegeben ist. Sei  $\alpha \in (0, 1)$  vorgegebenes Niveau

Idee: Lehne  $H_0$  ab, falls  $\bar{X}_n$  „zu groß“.

Betrachte kritischen Bereich der Form

$$K = \{(x_1, \dots, x_n) \in \mathbb{R}^n : \frac{1}{n}(x_1 + \dots + x_n) > \gamma\}$$

Bestimme  $\gamma$ :

Niveaubedingung: (die *W'keit*, dass fälschlicherweise abgelehnt wird)

$$(*) P_{\mu}(\bar{X}_n > \gamma) \leq \alpha \text{ für alle } \mu \leq \mu_0$$

Wähle  $K$  so groß wie möglich, also  $\gamma$  so klein wie möglich unter der Bedingung  $(*)$  (Satz 8.1)

Vereinfache die Bedingung

Für jedes  $\gamma \in \mathbb{R}$  ist

$$P_{\mu}(\bar{X}_n > \gamma) = P_{\mu}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > \frac{\gamma - \mu}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{\gamma - \mu}{\sigma/\sqrt{n}}\right)$$

streng monoton wachsend in  $\mu$ .

Daher gilt

$$(*) \Leftrightarrow (**) P_{\mu_0}(\bar{X}_n > \gamma) \leq \alpha$$

Das kleinste  $\gamma \in \mathbb{R}$ , das  $(**)$  erfüllt, ist gegeben durch

$$P_{\mu_0}(\bar{X}_n > \gamma) = \alpha \quad (\text{da } P_{\mu_0}(\bar{X}_n > \gamma) \text{ monoton fallend in } \gamma)$$

Auflösen nach  $\gamma$ :

$$\begin{aligned} \alpha &= P_{\mu_0}(\bar{X}_n > \gamma) = 1 - \Phi\left(\frac{\gamma - \mu_0}{\sigma/\sqrt{n}}\right) \\ &\Leftrightarrow \Phi\left(\sqrt{n}\frac{\gamma - \mu_0}{\sigma}\right) = 1 - \alpha \\ &\Leftrightarrow \sqrt{n}\frac{\gamma - \mu_0}{\sigma} = \Phi^{-1}(1 - \alpha) =: u_{1-\alpha} \\ &\Leftrightarrow \gamma = \mu_0 + \frac{\sigma}{\sqrt{n}}u_{1-\alpha} \end{aligned}$$

Test zum Niveau  $\alpha$  für das Testen von

$$H_0 : \mu \leq \mu_0 \text{ gegen } H_1 : \mu > \mu_0$$

**lehnt  $H_0$  genau dann ab, wenn**

$$\sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} > u_{1-\alpha},$$

wobei  $u_{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der  $N(0,1)$ -Verteilung ist.

Ähnlich zeigt man für vorgegebenes  $\mu_0 \in \mathbb{R}, 0 < \alpha < 1$  und bekanntes  $\sigma > 0$

Test zum Niveau  $\alpha$  für das Testen von

$$H_0 : \mu \geq \mu_0 \text{ gegen } H_1 : \mu < \mu_0$$

**lehnt  $H_0$  genau dann ab, wenn**

$$\sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma} < u_{1-\alpha},$$

wobei  $\Phi(u_{1-\alpha}) = 1 - \alpha$ .



Test zum Niveau  $\alpha$  für das Testen von

$$H_0 : \mu = \mu_0 \text{ gegen } H_1 : \mu \neq \mu_0$$

lehnt  $H_0$  genau dann ab, wenn

$$\sqrt{n} \frac{|\bar{X}_n - \mu_0|}{\sigma} > u_{1-\frac{\alpha}{2}},$$

wobei  $\Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ .

Die drei Tests heißen *Einstichproben-Gauß-Tests*.

**Bsp. 8.2:** Sei  $X \sim N(\mu, 0.16)$

Eine Stichprobe vom Umfang  $n=10$  liefert für das Stichprobenmittel den Wert

$$x_{10} = \bar{50},294$$

Teste

$$H_0 : \mu = 50 \text{ gegen } H_1 : \mu \neq 50$$

zum Niveau  $\alpha = 0.05$ .

Hier ist  $\mu_0 = 50, \sigma = \sqrt{0,16} = 0,4, u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$ .

$$\begin{aligned} \Rightarrow \sqrt{n} \frac{|\bar{X}_{10} - \mu_0|}{\sigma} &= \frac{0.294}{0.4} \sqrt{10} \\ &= 2.324 > 1.96 = u_{1-\frac{\alpha}{2}} \\ &\Rightarrow \text{lehne } H_0 \text{ ab.} \end{aligned}$$

**Fall B:**  $\mu$  und  $\sigma^2$  unbekannt Ersetze im Gauß-Test

$$\sigma \text{ durch } S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$$

und Quantil der  $N(0,1)$ -Verteilung durch Quantil der  $t(n-1)$ -Verteilung.  
Dies liefert im Fall  $\sigma$  unbekannt.

Test zum Niveau  $\alpha$  für das Testen von

- (a)  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$
- (b)  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$
- (c)  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$

lehnt  $H_0$  genau dann ab, wenn

- (a)  $\sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} > v_{1-\alpha}$
- (b)  $\sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n} < -v_{1-\alpha}$
- (c)  $\sqrt{n} \left| \frac{\bar{X}_n - \mu_0}{S_n} \right| > v_{1-\frac{\alpha}{2}}$

wobei  $v_{1-\alpha}$  (bzw.  $v_{1-\frac{\alpha}{2}}$ ) das  $1 - \alpha$  (bzw.  $1 - \frac{\alpha}{2}$ )-Quantil der  $t(n-1)$  Verteilung ist. Diese Tests heißen (Einstichproben-) *t-Tests*.

**Bem. 8.1:** Signifikanzniveau  $\alpha$  kontrolliert (ist eine obere Schranke für) nur die

W'keit  $H_0$  abzulehnen, wenn  $H_0$  richtig ist (Fehler 1. Art).

Die W'keit

$H_0$  nicht abzulehnen, wenn  $H_0$  falsch ist (Fehler 2. Art)

ist nicht kontrolliert.

Wähle daher  $H_0$  so, dass Fehler 1. Art der Gravierendere ist.

Möchte man eine Aussage „**statistisch widerlegen**“, wählt man die Aussage als  $H_0$ .

Möchte man eine Aussage „**statistisch nachweisen/untermauern/bestätigen**“, wählt man die Aussage als  $H_1$ .

Wird  $H_0$  abgelehnt, ist  $H_1$  statistisch nachgewiesen zum Niveau  $\alpha$ . Man sagt das Testergebnis ist statistisch signifikant zum Niveau  $\alpha$ .

Wird  $H_0$  nicht abgelehnt, heißt das Testergebnis statistisch nicht signifikant zum Niveau  $\alpha$ .

Es gilt nicht „ $H_0$  ist verifiziert“.

**Bem. 8.2:** Falls für einen gegebenen Datensatz  $\alpha_0 \in (0, 1)$  so ist, dass

$H_0$  zum Niveau  $\alpha$  abgelehnt wird  $\forall \alpha > \alpha_0$

und

$H_0$  zum Niveau  $\alpha$  nicht abgelehnt wird  $\forall \alpha < \alpha_0$

dann heißt  $\alpha_0$  *p-Wert* oder empirisches Signifikanzniveau. *p-Wert* kann als Maß für die Glaubwürdigkeit von  $H_0$  interpretiert werden. Je kleiner der *p-Wert*, desto weniger glaubwürdig ist  $H_0$ .

In Bsp. 8.3,  $H_0 : \mu = 50$  wird zum Niveau  $\alpha$  genau dann abgelehnt, wenn

$$\begin{aligned} 2.324 &> u_{1-\frac{\alpha}{2}} \\ \Leftrightarrow \Phi(2.324) &> \Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \\ \Leftrightarrow \alpha &> 2(1 - \Phi(2.324)) = 0.02 \end{aligned}$$

### 8.3 Der $\chi^2$ -Anpassungstest

Sind  $Z_1, \dots, Z_n$  unabhängig  $N(0, 1)$ -verteilte ZV, dann heißt die Verteilung von

$$V := Z_1^2 + \dots + Z_n^2$$

*Chi-Quadrat-Verteilung mit  $n$  Freiheitsgraden.* ( $V \sim \chi^2(n)$ )

#### Testproblem

Sei  $X$  eine diskrete ZV mit bekannten möglichen Werten

$$a_1, \dots, a_k$$

und unbekannten W'keiten

$$p_i = P(X = a_i), \quad 1 \leq i \leq k$$

Teste

$$H_0 : p_i = \bar{p}_i \text{ für } i = 1, \dots, k$$

wobei  $\bar{p}_1, \dots, \bar{p}_k > 0$  gegebene Zahlen sind mit  $\bar{p}_1 + \dots + \bar{p}_k = 1$ .

Sei dazu  $(X_1, \dots, X_n)$  eine Zufallsstichprobe und

$$\begin{aligned} N_i &:= |\{j \in \{1, \dots, n\} : X_j = a_i\}| \\ &= \text{absolute Häufigkeit des Werts } a_i \text{ in der Stichprobe} \end{aligned}$$

für  $i = 1, \dots, k$ .

Ist  $H_0$  richtig, dann ist  $E(N_i) = n\bar{p}_i$  und  $(N_i - n\bar{p}_i)^2$  sollte nicht zu groß sein.

Theorie liefert folgende Aussage für die Statistik

$$Q := \sum_{i=1}^k \frac{(N_i - n\bar{p}_i)^2}{n\bar{p}_i}$$

Ist

$$n\bar{p}_i \geq 5, \quad 1 \leq i \leq k$$

dann hat  $Q$  annähernd eine  $\chi^2(k-1)$ -Verteilung, falls  $H_0$  richtig ist.

Ist also  $x_{1-\alpha}$  das  $(1-\alpha)$ -Quantil der  $\chi^2(k-1)$ -Verteilung, so ist

$$P(Q \leq x_{1-\alpha}) \approx 1 - \alpha$$

und

$$P(Q > x_{1-\alpha}) \approx \alpha, \quad \text{falls } H_0 \text{ richtig ist.}$$

Das Verfahren

Lehne  $H_0$  genau dann ab, wenn  $Q > x_{1-\alpha}$

ist also ein (approximativer) Test zum Testen von  $H_0$  zum Niveau  $\alpha$ .

Der Test ist ein  $\chi^2$ -Anpassungstest.

### Bsp. 8.3:

Teste Echtheit eines Würfels zum Niveau  $\alpha=0.01$ . Hier ist

$$a_1 = 1, \dots, a_6 = 6, \quad k = 6$$

$$\bar{p}_i = \frac{1}{6}, \quad 1 \leq i \leq 6$$

$$H_0 : p_i = \frac{1}{6}, \quad i = 1, \dots, 6$$

In  $n = 120$  Würfeln waren die Häufigkeiten der 6 Seiten

$$n_1 = 29, n_2 = 11, n_3 = 26, n_4 = 12, n_5 = 27, n_6 = 15$$

$n\bar{p}_i = 20 \geq 5$  für alle  $i \Rightarrow \chi^2$ -Test anwenden.

$$\begin{aligned} q &= \sum_{i=1}^6 \frac{(N_i - n\bar{p}_i)^2}{n\bar{p}_i} = \frac{1}{20} \sum_{i=1}^6 (n_i - 20)^2 = \frac{1}{20} (9^2 + 9^2 + 6^2 + 8^2 + 7^2 + 5^2) \\ &= 16,8 \end{aligned}$$

$(1-\alpha)$ -Quantil der  $\chi^2(5)$ -Verteilung:  $x_{0,99} = 15,09$

$$q = 16,8 > x_{0,99}$$

Lehne Nullhypothese „Würfel ist echt“ ab.