Statistik - SoSe 2025

Inhaltsverzeichnis

I Wahrscheinlichkeitsrechnung **2**

1 Grundbegriffe **2**
1.1 Ergebnismenge und Ereignisse . . . 2
1.1.1 Rechenregeln für Mengen . . . 4
1.2 Wahrscheinlichkeitsmaße . . . 5
1.2.1 Rechenregeln für Wahrscheinlichkeitsmaße . . . 7

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit **9**
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

3 Zufallsvariablen **14**
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 . . . 19
3.4.2 Binomialverteilung . . . 20
3.4.3 Geometrische Verteilung . . . 21
3.4.4 Poisson-Verteilung . . . 21
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

4 Erwartungswert, Varianz und Kovarianz **25**
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

5 Grenzwertsätze **33**
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

II Schließende Statistik **36**

6 Parameterschätzung **36**
6.1 Erwartungstreue und Konsistenz . . . 36
6.1.1 Konsistenz . . . 38
6.2 Die Maximum-Likelihood-Methode . . . 39

7 Konfidenzintervalle **42**

Teil I
Wahrscheinlichkeitsrechnung

1 Grundbegriffe

1.1 Ergebnismenge und Ereignisse

Alle möglichen Ergebnisse eines Zufallsexperiments fasst man zur Ergebismenge Ω (Groß Omega) zusammen.

Bsp. 1.1: Zweimaliges Werfen eines Würfels

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

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

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

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 Ergebis 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 Ω 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* Ω 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.
$A \backslash B = A \cap B^c$	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
$B \backslash A = B \cap A^c = \{4,5,6\} \cap \{1,3,5\} = \{5\}$	Augenzahl ist mind. 4, aber nicht gerade.
$A \backslash B = A \cap B^c = \{1,3,5\} \cap \{4,5,6\} = \{1,3\}$	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. Ω 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 reelwertige Funktion P heißt **Wahrscheinlichkeitsmaß** auf Ω falls sie die folgenden Axiome der Wahrscheinlichkeitsrechnung erfüllt:

(i) $P(A) \geq 0$ für alle Ereignisse $A < \Omega$

(ii) $P(\Omega) = 1$

(iii) σ-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). (Ω, 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 reelwertig

(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 Ω, so ist

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i_1}^{n} |A_n|,$$

also

$$P(\bigcup_{i=1}^{n} A_i) = \frac{|\bigcup_{i=1}^{n} A_i|}{|\Omega|} = \frac{\sum_{i_1}^{n} |A_n|}{|\Omega|} = \sum_{i=1}^{n} \frac{|A_i|}{|\Omega|} = \sum_{i=1}^{n} P(A_i).$$

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

$$\Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcup_{i=1}^{N} A_i)$$

$$\Rightarrow \sum_{i=1}^{N} P(A_i) = \sum_{i=1}^{\infty} P(A_i)$$

$$P(A_i) = P(\emptyset) = \frac{|\emptyset|}{|\Omega|} = 0 \text{ für alle } i > N$$

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 $B = \{(1,1),(2,2),\ldots,(6,6)\}$ 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 $\Omega = \{0,1,\ldots,36\}$ ist die W'keit, dass eine ungerade Zahl fällt,

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

....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: $n_1, n_2, \ldots, n_k$ sind unabhängig.) Die Anzahl der Möglichkeiten, alle k Aufgaben zu erledigen, ist $n_1 \cdot n_2 \cdot \cdots \cdot n_k$

Bsp. 1.7: Eine faire Münze wird n-mal geworfen. Ergebnisse sind Folgen von Kopf(K) und Zahl(Z) der Länge n: $\Omega = \{(x_1,...,x_n) : x_i \in \{K,Z\}, i = 1,...,n\}$. 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“?

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

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

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 $\Omega = \{(x_1,...,x_k) : x_i \in \{1,...,365\}, i = 1,...,k\}$, also $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{365}$ für alle $A \subset \Omega$.

Sei $B = \{(x_1, ..., x_k) \in \Omega : \text{mind. 2 der } x_i \text{ stimmen überein}\}$

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

1.2.1 Rechenregeln für Wahrscheinlichkeitsmaße

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

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

1.3 **Falls $\mathbf{A \subset B}$, dann $\mathbf{P(B \backslash A) = P(B) - P(A)}$**, denn:
Falls $A \subset B$, dann

$$\begin{aligned}
B &= B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c) \\
&\overset{A \subseteq B}{=} A \cup (B \cap A^c) \\
\Rightarrow P(B) &\overset{(iii)}{=} P(A) + P(B \cap A^c) = P(A) + P(B \backslash A) \Rightarrow (1.3)
\end{aligned}$$

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

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

1.4 **Falls $\mathbf{A \subset B}$, dann $\mathbf{P(A) \leq P(B)}$ *(Monotonie)***, denn:
Falls $A \subset B$, dann $P(B) - P(A) \overset{(1.3)}{=} P(B \backslash A) \overset{(i)}{\geq} 0$

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

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

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

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

und

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

Daher ist

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

1.7 Erweiterung von 1.6:

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

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

$$\mathbf{P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)}$$
$$\mathbf{P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)}$$

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 Ω nicht diskret, so ist es aus mathematischen Gründen in manchen Andwendungen nicht möglich, das W.-Maß P für alle Teilmengen von Ω zu definieren. Jedoch lässt sich der Definitionsbereich von P stets so groß wählen, dass alle praktisch interessanten Teilmengen von Ω erfasst sind. Ist Ω diskret, so lässt sich das W.-Maß stets für alle Teilmengen von Ω 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 Ω eine Ergebnismenge und P ein W'Maß auf Ω. $B \subset \Omega$ 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 Ω (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 Ω 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:

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

2.1.3 Satz von Bayes

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

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

Denn:

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

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 gil $P(A_1) = 0,6$, $P(A_2) = 0,3$, $P(A_3) = 0,1$.
Für das Ereignis B: „E-mail enthält das Wort ‚gratis'“gelte $?(B|A_1) = 0,8$, $P(B|A_2) = 0,05$, $P(B|A_3) = 0,05$. (W'keiten wurden durch Beobachtung des Postfachs errechnet.)
⇒ 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 ($A_1, \ldots, A_n$ bilden eine Zerlegung von Ω, $B \subset \Omega$, $P(B) > 0$) bezeichnet man $P(A_1), P(A_2), \ldots, P(A_n)$ auch als *a-priori-Wahrscheinlichkeiten* und $P(A_1|B), P(A_2|B), \ldots, P(A_n|B)$ 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

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1) \cdot P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{0,99 \cdot 0,03}{0,99 \cdot 0,003 + 0,02 \cdot 0,997} \\ &= 0,13 \qquad \text{(a posteriori)} \end{aligned}$$

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):

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

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 Resulatate zu erhalten.

Bsp.2.6: Zweimaliger Würfelwurf.
A: „erste Augenzahl gerade“
B: „zweite Augenzahl ungerade“
C: „beide Augenzahlen gerade oder beide ungerade“
$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$:

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

$\Rightarrow A, B, C$ 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“$\dot{\text{D}}$ie 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 ω abhängt. Genauer:

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

Bsp. 3.1: Zweimaliger Münzwurf

Ω	$X(w)$ =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 $\omega \in \Omega = \{\omega_1, \ldots, \omega_4\}$, dann lassen sich die W'keiten

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

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 $X_1, \ldots, X_n$ heißen unabhängig, falls

$$(*) \qquad P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdot \cdots \cdot P(X_n \in B_n)$$

für alle $B_1, \ldots, B_n \in \mathbb{R}$.
(b) Ist $X_1, X_2, \ldots$ eine unendliche Folge von ZV, dann heißen $X_1, X_2, \ldots$ unabhängig, falls $(*)$ für alle $n \geq 2$ gilt.

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

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

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

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

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

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

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

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

$|\Omega| = 6^n, P(A) = \frac{|A|}{|\Omega|}$ für alle $A \in \Omega$
X_i = Ergebnis des i-ten Wurfs
$X_i((\omega_1, \ldots, \omega_n)) = \omega_i$
Behauptung: $X_1, \ldots, X_n$ sind unabhängig. Nach Satz 3.1 reicht es zu zeigen:
$P(X_1 = b_1, \ldots, X_n = b_n) = P(X_1 = b_1) \cdot \cdots \cdot P(X_n = b_n)\ \forall\, b_1, b_2, \ldots, b_n \in \mathbb{R}$.

Fall 1: Ein b_i ist keine mögliche Augenzahl $\Rightarrow$ beide Seiten $= 0$
Fall 2: $b_1, \ldots, b_n \in \{1, \ldots, 6\}$:

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

und

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

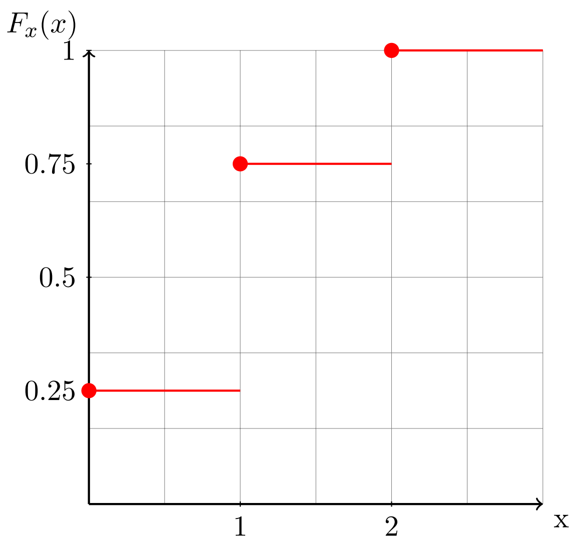
$\Rightarrow P(X_1 = b_1) \cdot \cdots \cdot P(X_n = b_n) = \frac{1}{6^n} = P(X_1 = b_1, \ldots, X_n = b_n)$.
Es folgt, dass $X_1, \ldots, X_n$ unabhängig sind.

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

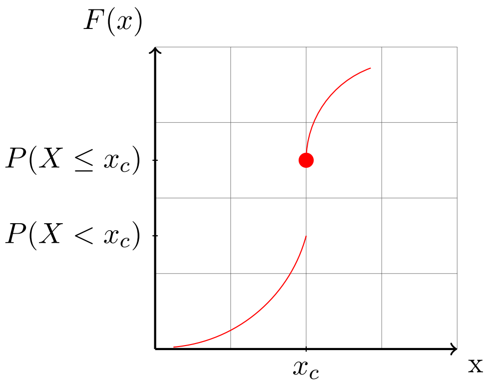
(c) F ist rechtsseitig stetig.

(d) $P(X > x) = 1 - F(x), x \in \mathbb{R}$

(e) $P((a < X \leq b)) =$

(f) $P(X < x) = F(x-)$, wobei $F(x-) = lim_{y \to x} F(y)$ 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.



$P(X = x) = P(X \leq x_c) - P(X < x_c)$ (s. Sprung in der Grafik)
$P(X = x) = 0 \Leftrightarrow$ F ist stetig an der Stelle X
$P(X = x) = 0$ für alle $x \in \mathbb{R} \Leftrightarrow$ F ist stetig auf $\mathbb{R}$

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

Beweis:

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

(b),(c) Übungsaufgabe 21

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

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

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

Bem. 3.2: Man kann zeigen:

(a) Jede Funktion $F : \mathbb{R} \to \mathbb{R}$, 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 $B \in \mathbb{R}$.
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

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

definierte Funktion $f : \mathbb{R} \to \mathbb{R}$ 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.,

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

Dann gilt für jedes $B \in \mathbb{R}$,

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

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

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

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

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) σ-Additivität
Seien $B_1, B_2, \dots \subset \mathbb{R}$ paarweise disjunkt
$\Rightarrow P_x(B_1 \cup B_2 \cup \dots) = P(X \in B_1 \cup B_2 \cup \dots) = P(\{X \in B_1\} \cup \{X \in B_2\} \cup \dots)$
$\overset{(\sigma\text{-Add.})}{=} \sum_i P(X \in B_i) = \sum_i P_x(B_i)$.

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 $f : \mathbb{R} \to \mathbb{R}$ gibt, so dass

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

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

Sei X eine stetige ZV mit Dichte f und VF F
$\to$ F ist stetig $\overset{(Satz 3.2(g))}{\Rightarrow} P(X = x) = 0$ für alle $x \in \mathbb{R}$.
$\Rightarrow$ Für $a \leq b$ gilt

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

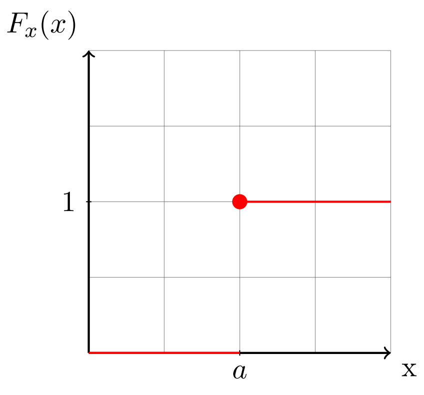
Ist die Dichte f an der Stelle x_c stetig, so gilt

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

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

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

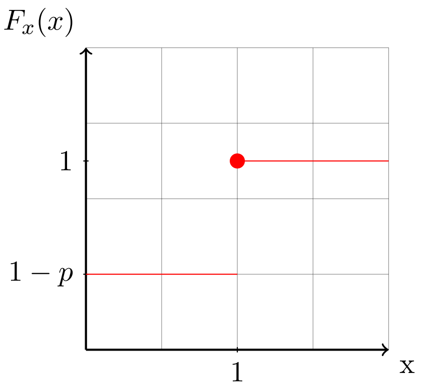
heißt *Bernoulli-verteilt* mit Parameter p.
X wird auch als Bernoulli-Variable bezeichnet. ($0 \le p \le 1$)
Bez. X∼Ber(p).

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 tritt ein
$X = I_A$ ist eine Bernoulli-Variable mit Parameter p=P(A).
Man sagt oft,

$$\text{Erfolg tritt ein, falls } X(\omega) = 1$$
$$\text{Misserfolg tritt ein, falls } X(\omega) = 0$$



3.4.2 Binomialverteilung

Seien $X_1, \dots, X_n$ unabhängig und jeweils $Ber(p)$ (Erfolgsw'keit jeweils p).
Anzahl der Erfolge

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

Bsp.: n Münzwürfe

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

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,

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

$\rightarrow$ 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. $P(S_n = 0) = \binom{n}{0} \cdot p^0 \cdot q^n = q^n, P(S_n = 1) = \binom{n}{n} \cdot p^n \cdot q^0 = p^n$.
S_n heißt *binomialverteilt* mit Parametern n(# Versuchen) und p(Erfolgsw'keit).
Bez. S_n ~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 ~Bin(5, $\frac{1}{3}$)

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

3.4.3 Geometrische Verteilung

Seien $X_1, X_2, \dots$ unabh., ~Ber(p), $0 < p \leq 1$
$X_i = 1 \leftrightarrow$ Erfolg im i-ten Versuch.
T = Anuahl der Versuche bis zum ersten Erfolg (einschließlich)

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

Für $k \geq 2$ ist

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

Mit $q = 1 - p$ also

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

Bez. T~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~Poi(λ).
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^λ.
Anwendung der Posison-Verteilung zu Approximation einer Binomialverteilung.
Sei X∼Bin(n,p). Ist n „groß“ und p „klein“ etwa $n \geq 50, p \leq \frac{1}{10}$ und $np \leq 10$, dann

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

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
∼Bin(n,p) mit $n = 3600$ und $p = 0,001$
⇒ Poisson-Approximation $[n \geq 50, p \leq \frac{1}{10}, np = 2,6 \leq 10]$
Approximiere Bin(n,p) durch Poi(λ) mit $\lambda = np = 3,6$.

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

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

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

Daher wird die Konvention

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

benutzt.
Für $X \in \mathbb{N}_0$ gilt

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

Eine ZV mit W-funktion f heißt *hypergeometrisch verteilt* mit Parametern N,M, und n.
Bez. X∼Hyp(N,M,n).

Bsp. 3.8: Lotto 6 aus 49
Jemand hat die Zahlen 1,2,...,6 getippt. W'keit für 3 Richtige?

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

$n = 6$ Kugeln werden ohne Zurücklegen gezogen.

$X =$ Anzahl richig getippter Zahlen.
W'keit für 3 Richtige:

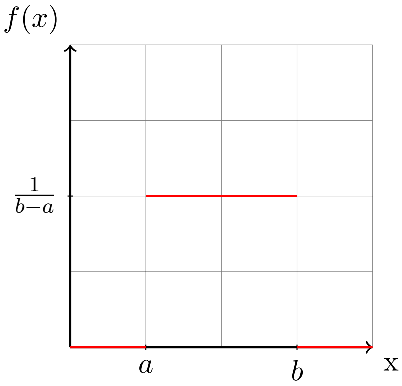
$$P(X = 3) = \frac{\binom{6}{3}\binom{49-6}{6-3}}{\binom{49}{6}} \approx 0,01765$$

3.5 Beispiele stetiger Verteilungen

3.5.1 Gleichverteilung

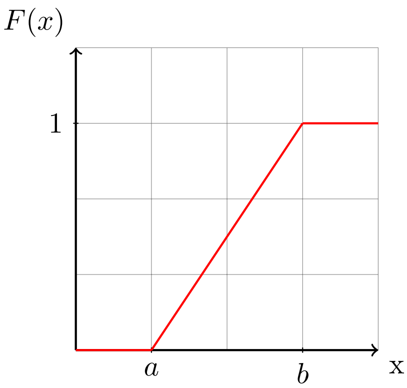
Gleichverteilung auf $[a, b]$, $\quad a < b$.
Dichte

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



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∼Uni(0,1), dann gilt für $0 \leq c \leq d \leq 1$

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

und für $c < 0 \leq d \leq 1$ gilt

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

3.5.2 Exponentialverteilung

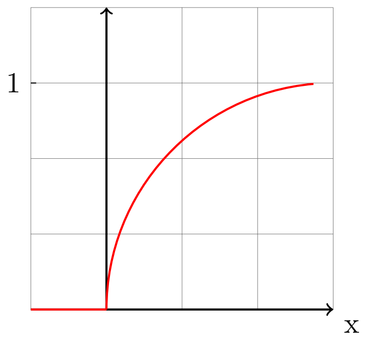
mit Parameter $\lambda > 0$
Bez. X$\sim$Exp(λ)
Diese Verteilung wird oft beutzt zur Modellierung von Wartezeiten oder Lebensdauern

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)

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

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

μ: Mittelwert/Erwartungswert einer Zufallsvariable mit obiger Dichte
σ^2: Varianz einer Zufallsvariable mit obiger Dichten
$N(0,1)$ heißt *Standardnormalverteilung*

Dichte

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

VF

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

$\rightarrow$ 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)
$P(Y \geq y)$ = VF der transformierten Zufallszahl

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

$\Rightarrow$ 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 $P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$, $x \in \mathbb{R}$
Dabei ist Φ 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 $x_1, x_2, \ldots$ die verschiedenen möglichen Werte von f, sei f die W'funktion von X.
Der *Erwartungswert* (EW) von X wird definiert durch

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

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

Bsp. 4.1:

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

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

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

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

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

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

(d) Für $X \sim Poi(\lambda)$ gilt:

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

4.1.1 Satz 4.1:

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

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

Bew.:

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

Bsp. 4.2: Ist $X \sim$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

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

Hier wird angenommenn, dass $\int_{-\infty}^{\infty} xf(X)dx < \infty$ ist.
Für nichtnegative stetige ZV X ist $E(X)$ immer wohldefiniert, aber evtl. ∞.

Bsp. 4.3:

(a) Sei $X \sim$Uni(a, b)

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

(b) Sei $X \sim$Exp(λ)

$$E(X) = \int_0^{\infty} x\lambda e^{-\lambda x}dx \quad (1)$$
$$= [-xe^{-\lambda x}]_0^{\infty} - \int_0^{\infty} e^{-\lambda x} \quad (2)$$
$$= 0 + [-\frac{1}{\lambda}e^{-\lambda x}]_0^{\infty} = \frac{1}{\lambda} \quad (3)$$
$$E(X) = \frac{1}{\lambda} \quad (4)$$
$$(5)$$

4.1.2 Satz 4.2:

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

(a)

$$E(aX + bY) = aE(X) + bE(Y)\text{für alle } a, b \in \mathbb{R} \text{ (Linearität)}$$
$$E(a + bX) = a + bE(X)\text{für alle } a, b \in \mathbb{R}$$
$$E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \cdot E(X_i)$$

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

$$E(X_1, \ldots, X_n) = E(X_1) \cdot \cdots \cdot E(X_n)$$

(c) Falls X eine nichtnegative ZV ist, dann gil

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

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

$$\Rightarrow E(X) \overset{Satz\ 4.2\ (a)}{=} E(X_1) + \cdots + E(X_n)$$

(b) $X \sim \text{N}(\rho, \sigma^2)$
X hat dieselbe Verteilung wie $\mu + \sigma Y$, wobei $Y \sim \text{N}(0,1)$

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

4.1.3 Satz 4.3:

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

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

Bsp. 4.5: X∼N(0,1)

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

partielle Integration

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

4.2 Varianz einer ZV

Die *Varianz* einer ZV X ist det. durch

$$Var(x) := E[(X - E(X))^2]$$

und die *Standardabweichung* von X ist

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

$Var(X)$ 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

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

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∼Ber(p)
Da $X = 0$ oder $X = 1$, folgt $X^2 = X$.
$\Rightarrow Var(X) \overset{Satz\ 4.4\ (a)}{=} E(X^2) - [E(X)]^2 = E(X) - [E(X)]^2 \overset{Bsp.\ 4.1\ (c)}{=} p - p^2 = p(1-p)$ (b)
X∼Bin(n,p)
$\Rightarrow X = X_1 + \dots + X_n$, X_i unabhängig, X_i ∼Ber(p)
$\Rightarrow Var(X) = Var(X_1 + \dots + X_n) \overset{Satz\ 4.4\ (e)}{=} Var(X_1) + \dots + Var(X_n) = np(1-p)$ (b)
X∼ $N(\mu, \sigma^2)$, Var(X)=?
Sei $Y \sim N(0, 1)$
$\Rightarrow$ X hat dieselbe Verteilung wie $\mu + \sigma Y$
$\Rightarrow Var(X) = Var(\mu + \sigma Y) \overset{Satz\ 4.4\ (c)}{=} \sigma^2 Var(Y) = \sigma^2(E(Y^2) - (E(Y))^2) = \sigma^2$

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

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

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

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

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

$\Rightarrow$ 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
$Cov(X,Y) = E(XY) - E(X)E(Y) = 0$
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 $X, Y, X_1, \ldots, X_n, Y_1, \ldots, Y_n$ ZV

(a) $Cov(X,Y) = Cov(Y,X)$

(b) $Cov(X,X) = Var(X)$

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

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

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

Bew. (c):

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

Bew. (d):

$$
\begin{aligned}
&Cov(\sum_i a_i X_i, \sum_j b_j X_j) \\
=&E(\sum_i \sum_j a_i b_i X_i Y_i) - (\sum_i a_i E(X_i))(\sum_j b_j E(Y_j)) \\
=&\sum_i \sum_j a_i b_j E(X_i Y_j) - a_i b_j E(X_i) E(Y_j) \\
\underset{(a)}{=}&\sum_i \sum_j a_i b_j Cov(X_i, Y_j)
\end{aligned}
$$

Bew. (e):

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

4.4 Ungleichungen

4.4.1 Satz 4.6: Markov-Ungleichung

Sei X eine nicht-negative ZV. Dann gilt

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

Beweis: Definiere eine Bernoulli-ZV durch

$$Y(\omega) := \begin{cases} 1 \text{ falls } X(\omega) \geq t \\ 0 \text{ falls } X(\omega) < t \end{cases}$$

$$\Rightarrow Y(\omega) \leq \frac{X(\omega)}{t} \quad \forall \omega$$

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

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

Bsp. 4.8: Sei X ~Poi(1) und $t > 1$.

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

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

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

und

$$E(e^{uX}) \underset{Satz4.3}{=} e^{-1} \sum_{k=0}^{\infty} \frac{(e^u)^k}{k!} = e^{-1}\exp(e^u - 1)$$

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

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

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

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 $\sigma^2 := Var(X)$. Dann gilt

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

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

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

Beweis:

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

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

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

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

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

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

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

$$\bar{X_n} := \frac{X_1 + \cdots + X_n}{n}$$

Für große N

$$E(\bar{X}_n) = \frac{1}{n}(E(X_1) + \cdots + E(X_n)) = \mu$$

$$Var(\bar{X}_n) = \frac{1}{n^2}(Var(X_1) + \cdots + Var(X_n)) = \frac{\sigma^2}{n} \to 0(\text{für } n \to \infty)$$

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

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

Zentraler Grenzwertsatz beschreibt die Verteilung der standardisierten ZV

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

für $n \to \infty$

5.1 Gesetz der großen Zahlen

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

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

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

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

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 $X_1, X_2, \ldots$ *konvergiert stochastisch* gegen eine ZV X, geschrieben $X_n \xrightarrow{P} X$, falls für jedes $\epsilon > 0$

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

gilt.

5.1.1 Satz 5.1: Schwaches Gesetz der großen Zahlen

Seien $X_1, X_2, \ldots$ 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.:

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

Für jedes $\varepsilon > 0$ gilt

$$\begin{aligned} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \varepsilon) \\ (\text{Tschebyscheff}) \leq \frac{Var(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\varepsilon^2} \underset{\text{Satz 4.7}}{==} \frac{1}{n^2\varepsilon^2} = \sum_{i=1}^{n} Var(X_i) \\ = \frac{Var(X_1)}{n\varepsilon^2} \to 0 (n \to \infty) \end{aligned}$$

Bsp. 5.2: Seien $X_1, X_2, \ldots$ unabhängig, identisch verteilte ZV, $X_i \sim$Ber(p)
$\to E(X_1) = p$ 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 $\to$ Erfolgsw'keit

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

$$Y_i(\omega) = \begin{cases} 1 & \text{falls } X_i(\omega) \in a \\ 0 & \text{falls } X_i(\omega) \notin a \end{cases}$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} E(Y_1) = P(X_1 \in A)$$

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

5.1.2 Satz 5.2: Starkes Gesetz der großen Zahlen

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

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

5.2 Zentraler Grenzwertsatz

5.2.1 Satz 5.2 Zentraler Grenzwertsatz ZGS

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

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

Dann gilt

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

wobei $Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{\frac{-x^2}{2}} dx =$ 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

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

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 $X_1, X_2, \ldots$ unabhängig $\sim$Ber(p), $0 < p < 1 \to S_n = X_1 + \cdots + X_n$ $\sim$Bin(n, p), $E(S_n) = np$, $Var(S_n) = np(1-p)$

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

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

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

(a) Exakte Lösung

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

(b) Normalverteilungsapproximation $X = S_n$ $\sim$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 $\sqrt{np(1-p)} = \sqrt{\frac{19}{2}} = 3,0822$

$$\begin{aligned}
P(5 \leq S_n \leq 15) &= P(\frac{5-np}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{15-np}{\sqrt{np(1-p)}}) \\
&= P(-1,6222 \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq 1,6222) \\
\text{(Über ZGS)} &\approx \Phi(1,6222) - \Phi(-1,6222) \\
&\quad \Phi(-x) = 1 - \Phi(x) \text{ für alle } x \in \mathbb{R} \text{ s. Tutorium A40} \\
&\Rightarrow P(5 \leq S_n \leq 15) \approx 2\Phi(1,6222) - 1 \approx 2 \cdot 0,9474 - 1 \approx 0,8948
\end{aligned}$$

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: $(X_1, \ldots, X_n)$ D.h. Stichprobenvariablen $(X_1, \ldots, X_n)$ sind unabhängige ZV, die jeweils dieselbe Verteilung haben wie X.
n unabhängige Wiederholungen von X.
Konstruiere mit Hilfe der $X_1, \ldots, X_n$ Schätzer für interesierenden Parameter der Verteilung, z.B. $\vartheta = E(X)$, oder $\vartheta = Var(X)$, oder, falls bekannt ist, dass $X \sim$Exp(λ), aber λ unbekannt, $\vartheta = \lambda$.

6.1 Erwartungstreue und Konsistenz

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

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

eine *Stichprobenfunktion* oder *Statistik*. Wird T zum Schätzen eines Parameters $\vartheta \in \mathbb{R}^p$ verwendet, so nennt man $T(X_1, \ldots, X_n)$ eine *Schätzfunktion* für (das Schätzen von) ϑ.
In Anwendungen: Erhebe Stichprobe, d.h. Beobachtungswerte $x_1, \ldots, X_n \in \mathbb{R} =$ Realisationen von $X_1, \ldots, X_n$.
$T(X_1, \ldots, X_n)$ ist ein Schätz<u>wert</u> (keine ZV) für ϑ.
Im Folgenden gelte für

$$\begin{aligned}
&\Theta = \text{Menge der möglichen Parameterwerte} \\
&\Theta \in \mathbb{R}\text{eindimensional}
\end{aligned}$$

Def. 6.1: Eine Schätzfunktion $\hat{\vartheta} = T(X_1, \ldots, X_n)$ für $\vartheta \in \Theta \in \mathbb{R}$ heißt *erwartungstreu* (e-treu) für ϑ, 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 X_n e-treue Schätzfunktion für $\mu = E(X)$.

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

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

die kleinste Varianz, denn
Sei $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} c_i X_i$ e-treu für μ

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

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

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

Bsp. 6.2: Schätze $\sigma^2 = Var(X)$

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

Ist $\tilde{S^2}$ e-treu für σ^2?

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

$$\begin{aligned} E(X_i^2) &= Var(X_i) + [E(X_i)]^2 = \sigma^2 + \mu^2 \\ E(\bar{X}_n^2) &= Var(\bar{X}_n) + [E(\bar{X}_n)]^2 = \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

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

$\tilde{S^2}$ ist nicht e-treu für σ^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 ϑ ist definiert durch

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

Bem. 6.1:

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

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

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

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

6.1.1 Konsistenz

Seien $X_1, X_2, \ldots$ unabh. identisch verteilte ZV.
Betrachte Folge von Schätzfunktionen $f \to$ der Schätzer eines Parameters:

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

Für jedes n sei $T_n(X_1, \ldots, X_n)$ Schätzfunktion basierend auf Stichprobe vom Umfang n.
Wird für $n \to \infty$ Schätzgenauigkeit bel. gut?

Def. 6.3: Eine Folge von Schätzfunktionen für ϑ:

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

heißt *konsistent* für den Parameter ϑ, falls

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

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 μ.

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

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

und $MSE(\hat{\vartheta}_n) \to 0 \forall n \to \infty$.
$\Rightarrow$ Für jedes $\varepsilon > 0$ gilt

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

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

Mit Bem. 6.1(b) folgt:
Falls $Bias(\hat{\vartheta}_n) \to 0$ und $Var(\hat{\vartheta}_n) \to 0$, dann ist $(\hat{\vartheta}_n)_{n=1}^{\infty}$ konsistent.

6.2 Die Maximum-Likelihood-Methode

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

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

Sei X eine ZV, deren Verteilung nur bis auf einen Parameter ϑ bekannt ist. D.h. die Verteilung von X wird beschrieben durch W-Maß P_ϑ, das von ϑ abhängt.
$X_1, \dots, X_n$ seien unabh. ZV, die jeweils dieselbe Verteilung wie X haben. Schätze ϑ mit Hilfe von $(X_1, \dots, X_n)$

Bsp. 6.4: Es sei bekannt, dass X eine Poisson-Verteilung hat, $X \sim$Poi(λ), aber λ unbekannt ist.
Schätze $\vartheta = \lambda$

<u>Diskreter Fall:</u>
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

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

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 ϑ aufgefasst bei geg. Werten $x_1, \ldots, x_n$.

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

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

Dann ist

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

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

Bem. 6.3

(a) Im Maximierungsproblem (*) sind $x_1, \ldots, x_n$ fest und es wird bzgl. ϑ maximiert.

(b) Evtl. existiert kein $\hat{\vartheta} \in \Theta$ oder es existieren mehrere $\hat{\vartheta} \in \Theta$, für die (*) gilt.

(c) Ist $L(\hat{\vartheta}|x_1, \ldots, x_n)$ differenzierbar nach ϑ, so findet man $\hat{\vartheta}$ oft durch Differenzieren. Ist $\vartheta \in \Theta \in \mathbb{R}$, betrachte

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

Beachte: Maximum kann am Rand liegen $\Theta = [a, b]$
Ist $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta \in \mathbb{R}^2$, betrachte die Gleichungen

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

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

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

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

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

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

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

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

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

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

1. Bestimme Likelihood-Funktion

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

2. Maximiere Likelihood-Funktion bzgl. λ
Betrachte Log-Likelihood-Funktion

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

$\Rightarrow$ An der Stelle $\lambda = \frac{x_1+\dots+x_n}{n}$ wird $L(\lambda|x_1,\dots,x_n)$ maximal, d.h.
Falls $x_1+\dots+x_n=0$, dann $X_1=\dots=x_n=0$

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

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

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

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(μ,σ^2). Zunächst sei $\sigma^2>0$ bekannt.
Bestimme ML-Schätfunktion für μ.
Dichte

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

Likelihood-Funktion

$$L(\mu|x_1,\dots,x_n) = f(x_1|\mu)\dots f(x_n|\mu) = (\frac{1}{\sqrt{2\pi\sigma^2}})^2\exp(-\sum_{i=1}^n -\frac{(x-\mu)^2}{2\sigma^2}) \Rightarrow \log L(\mu|x_1,\dots,x_n) = n\log\frac{1}{\sqrt{2\pi\sigma^2}} -$$

Maximiere bzgl. μ

$$\frac{\partial}{\partial\mu}\log L(\mu|x_1,\ldots,x_n) = \frac{1}{2\sigma^2}\sum_{i=1}^n 2(x-\mu) = \frac{1}{\sigma^2}[(\sum_{i=1}^n x_i) - n\mu] \stackrel{!}{=} 0 \Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^n x_i$$

und

$$\frac{\partial^2}{\partial\mu^2}\log L(\mu|x_1,\ldots,x_n) = -n < 0$$

$$\Rightarrow L(\mu|x_1,\ldots,x_n) \text{ wird maximal für } \mu = \frac{1}{n}\sum_{i=1}^n x_i$$

$\Rightarrow$ ML-Schätzfunktion für μ ist $\frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n$ sowohl wenn σ^2 bekannt ist als auch wenn σ^2 unbekannt ist.
ML-schätzfunktion für σ^2 ist

$\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ falls μ bekannt

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

ML-Schätzfunktion für σ

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

Bsp. 6.7: X$\sim$Uni(0,ϑ), $\vartheta > 0$ unbekannt
Dichte

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

Likelihood-funktion für $x_1,\ldots,x_n \geq 0$

$$L(\vartheta|x_1,\ldots,x_n) = f(x_1|\vartheta)\cdot\cdots\cdot f(x_n|\vartheta) = \begin{cases} (\frac{1}{\vartheta})^n, & x \in [0,\vartheta] \forall i = 1,\ldots,n \\ 0, & \text{sonst} \end{cases}$$

$$= \begin{cases} (\frac{1}{\vartheta})^n, & \text{falls } \vartheta \geq \max x_1,\ldots,x_n \\ 0, & \text{sonst} \end{cases}$$

7 Konfidenzintervalle

Ziel: Gebe (möglichst kurzes) Intervall an, das interessierenden unbekannten Parameter mit hoher W'keit enthält.
Sei X ZV, $(X_1,\ldots,X_n)$ Stichprobe, ϑ unbekannter Parameter der Verteilung von X, $\vartheta \in \Theta \in \mathbb{R}$

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

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

so dass

$$\alpha(x_1,\ldots,x_n) \leq \beta(x_1,\ldots,x_n) \text{ für alle } x_1,\ldots,x_n$$

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

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

Das Zufallsintervall [A,B] heißt dann ein $(1-\alpha)$-*Konfidenzintervall* für ϑ oder ein Konfidenzintervall für ϑ zum *Konfidenzniveau* $1-\alpha$.
$1-\alpha$ ist die *Überdeckungsw'keit* von [A,B].
Ein Konfidenzintervall [A,B] heißt *symmetrisch,* falls

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