Stochastik 1 (PAM)

6c SJ 2018/2019

Version: 18. März 2019

1 Mengen

Begriffe

Eine *Menge* M ist eine Zusammenfassung von Objekten, die *Elemente* der Menge genannt werden.

Ist x ein Element von M, schreiben wir $x \in M$, andernfalls $x \notin M$.

Eine Menge kann auch keine Elemente enthalten. In diesem Fall wird sie *leere Menge* genannt und mit $\emptyset$ oder $\{\,\}$ bezeichnet.

Beschreibung von Mengen

Enthält die Menge M endlich viele Elemente $x_1, x_2, \ldots, x_n$, so können wir sie in *aufzählender Form* darstellen:

$$M = \{x_1, x_2, \ldots, x_n\}$$

Wenn M unendlich viele Elemente x_1, x_2, ... enthält, die sich durchnummerieren lassen, schreiben wir

$$M = \{x_1, x_2, x_3, \ldots\}$$

Eine solche Menge M wird *abzählbar unendlich* genannt.

Für grosse Mengen eignet sich die *beschreibende Form*:

$M = \{x \mid x \text{ hat Eigenschaft } E\}$

Beispiele:

- $M_1 = \{x \mid x \in \mathbb{N} \text{ und } x \text{ ist gerade}\} = \{2, 4, 6, \ldots\}$
- $M_2 = \{x \mid x \in \mathbb{R} \text{ und } 0 \leq x \leq 1\}$

M_2 hat so viele Elemente, dass diese sich nicht durchnummerieren lassen. Eine solche Menge wird *überabzählbar (unendlich)* genannt.

Teilmengen

Ist jedes Element einer Menge A auch Element einer Menge B, wird A *Teilmenge* von B genannt. Dafür schreibt man $A \subset B$ oder $B \supset A$.

Gilt $A \subset B$ und $B \subset A$, so sind beide Mengen *gleich* und wir schreiben $A = B$.

Es ist manchmal vorteilhaft eine *Grundmenge* Ω einzuführen, die alle Objekte enthält, die möglicherweise von Interesse sein können. In diesem Falle betrachtet man nur Mengen, die Teilmengen von Ω sind.

Mengenoperationen

Das *Komplement* einer Menge A bezüglich einer Grundmenge Ω ist die Menge aller Elemente in Ω die nicht in A liegen und wird mit $\overline{A}$ oder A^c bezeichnet.

Die *Vereinigungsmenge* zweier Mengen A und B ist die Menge aller Elemente, die in A oder in B liegen und wird mit $A \cup B$ bezichnet. („oder“ wird hier nicht im ausschliessenden Sinne gebraucht.)

Die *Schnittmenge* zweier Mengen A und B ist die Menge aller Elemente, die sowohl in A als auch in B liegen und wird mit $A \cap B$ bezeichnet.

Das *kartesische Produkt* zweier Mengen A und B ist die Menge aller Paare (x, y) mit $x \in A$ und $y \in B$ und wird mit $A \times B$ bezeichet.

Sind A und B Mengen, so wird die Menge aller Element aus A, die nicht in B liegen *Mengendifferenz* genannt und mit $A \setminus B$ [sprich: „A ohne B“] bezeichnet.

Die Mengendifferenz ist jedoch keine neue Operation, denn innerhalb einer Grundmenge Ω gilt: $A \setminus B = A \cap \overline{B}$.

Ist A eine Menge, so wird die Menge aller Teilmengen von A *Potenzmenge* genannt und mit $\mathcal{P}(A)$ oder 2^A bezeichnet. Damit sind die Elemente der Potenzmenge selbst Mengen. Beispiel:

$\mathcal{P}(\{1,2,3\}) = \big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\big\}$

Zusammenfassung der Mengenoperationen

$$
\begin{aligned}
\overline{A} &= \{x \mid x \in \Omega \text{ und } x \notin A\} \\
A \cup B &= \{x \mid x \in A \text{ oder } x \in B\} \\
A \cap B &= \{x \mid x \in A \text{ und } x \in B\} \\
A \setminus B &= \{x \mid x \in A \text{ und } x \notin B\} = A \cap \overline{B} \\
A \times B &= \{(x, y) \mid x \in A \text{ und } y \in B\} \\
\mathcal{P}(A) &= \{M \mid M \subset A\}
\end{aligned}
$$

Mengenalgebra

Die Mengenoperationen haben verschiedene Eigenschaften:

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: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Gesetze von De Morgan: $\overline{A \cup B} = \overline{A} \cap \overline{B}$
$\overline{A \cap B} = \overline{A} \cup \overline{B}$

Spezialfälle: $A \cup \emptyset = A \qquad A \cup A = A \qquad A \cup \Omega = \Omega$
$A \cap \emptyset = \emptyset \qquad A \cap A = A \qquad A \cap \Omega = A$

Mächtigkeit von Mengen

Ist A eine endliche Menge, so ist die *Mächtigkeit* von A die Anzahl der Elemente von A und wird mit $|A|$ abgekürzt.

Zwei Mengen A und B sind *gleichmächtig*, wenn es eine bijektive Abbildung $f\colon A \to B$ gibt.

Eine Menge, die gleichmächtig zu einer Teilmenge von $\mathbb{N}$ ist, heisst *abzählbar*. Eine Menge die nicht abzählbar ist, heisst *überabzählbar*.

2 Wahrscheinlichkeitsmodelle

Zufallsexperimente

Ein *Zufallsexperiment* ist ein Experiment, dessen Ausgang nicht vorhersehbar ist.

(a) Augenzahl beim Werfen eines Spielwürfels

(b) zweifacher Münzwurf

Stichprobenraum

Die Menge aller möglichen *Ergebnisse* (*outcomes*) wird *Stichprobenraum* (*sample space*) genannt und meist mit Ω bezeichnet.

(a) $\Omega = \{1, 2, 3, 4, 5, 6\}$

(b) $\Omega = \{ZZ, ZW, WZ, WW\}$ (Z: Zahl, W: Wappen)

Um es nochmals deutlich zu wiederholen: Die Elemente eines Stichprobenraums müssen jedes Ergebnis darstellen und sich gegenseitig ausschliessen (*collectively exhaustive*, *mutually exclusive*).

Wahl des Modell

Bei der Wahl eines Stichproberaums haben wir eine gewisse Freiheit, welche Faktoren in das Modell einbezogen werden.

Jemand könnte auf die Idee kommen, dass die Luftfeuchtigkeit einen Einfluss auf den Ausgang des doppelten Münzwurfs hat.

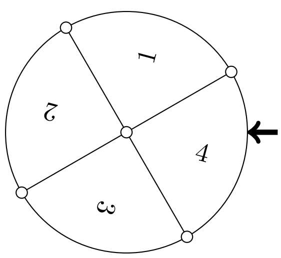
$\Omega = \{ZZ+, ZW+, WZ+, WW+, ZZ-, ZW-, WZ-, WW-\}$

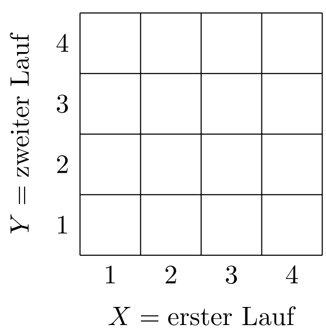
wobei $+$ hohe Luftfeuchtigkeit
$-$ niedrige Luftfeuchtigkeit

Die Wahl der Faktoren, die den Ausgang des Experiments beeinflussen, liegt in der Hand des Experimentators.

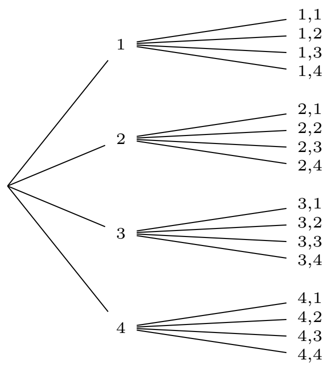
Beispiel 2.1 (diskreter Stichprobenraum)

Zwei Läufe mit einem Glücksrad:



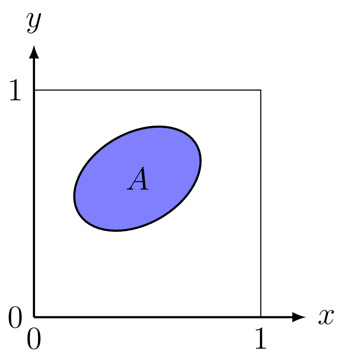


sequentielle Beschreibung mittels *Baumdiagramm*:



Bemerkung: Die Zwischenergebnisse beim ersten Lauf werden manchmal auch *Resultate* genannt, um sie von dem *Ergebnis* zu unterscheiden.

Beispiel 2.2 (stetiger Stichprobenraum)



$$\Omega = \{(x, y)\colon 0 \leq x, y, \leq 1\}$$

Ereignisse

- *Ereignis* (*event*): Teilmenge des Stichprobenraums.
- Jedem Ereignis wird eine Zahl (*Wahrscheinlichkeit*, *probability*) zugeordnet

Ein Ereignis $A \subset \Omega$ *trifft ein*, wenn das Ergebnis in A liegt.

Ist ω ein Ergebnis, so wird $\{\omega_0\}$ manchmal auch *Elementarereignis* genannt.

In überabzählbaren Stichprobenräumen gibt es „exotische“ Teilmengen, denen keine Wahrscheinlichkeit zugeordnet werden kann. Da sie „nur“ in der Theorie und nicht in den Anwendungen auftreten, schliessen wir sie für die folgenden Betrachtungen aus:

Σ bezeichnet die Menge aller $A \subset \Omega$, denen eine Wahrscheinlichkeit zugeordnet werden kann und wird *Sigma-Algebra* genannt.

Axiome der Wahrscheinlichkeitsrechnung

Eine Abbildung P, die jedem Ereignis $A \in \Sigma$ eine Zahl zuordnet und folgende Bedingungen erfüllt, wird *Wahrscheinlichkeitsfunktion* (*law of probability*) genannt.

(1) $\forall A \in \Sigma\colon P(A) \geq 0$

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

(3) $\forall A, B \in \Sigma$ disjunkt: $P(A \cup B) = P(A) + P(B)$

Für bestimmte Situationen muss (3) verschärft werden:

(3') $\forall A_1, A_2, \ldots \in \Sigma$ paarweise disjunkt: $P\left(\bigcup_{i}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Bemerkung

Die Wahl der Wahrscheinlichkeitsfunktion P wird durch die Axiome zwar eingeschränkt aber nicht eindeutig festgelegt.

Folgerung 1

Sei A ein Ereignis in einem Stichprobenraum Ω und $P\colon \Sigma \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Dann gilt:

$P(A) = 1 - P(\overline{A}) \quad (F_1)$

Beweis:

$1 \overset{(2)}{=} P(\Omega) = P(A \cup \overline{A}) \overset{(3)}{=} P(A) + P(\overline{A}) \quad \Rightarrow \quad P(A) = 1 - P(\overline{A})$

$\square$

Folgerung 2

Sei A ein Ereignis in einem Stichprobenraum Ω und $P\colon \Sigma \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Dann gilt:

$0 \leq P(A) \leq 1 \quad (F_2)$

Beweis: Wegen (1) muss nur noch $P(A) \leq 1$ gezeigt werden.

$P(A) \overset{(F_1)}{=} 1 - P(\overline{A}) \overset{(1)}{\leq} 1$

$\square$

Folgerung 3

Sei A ein Ereignis in einem Stichprobenraum Ω und $P\colon \Sigma \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Dann gilt:

$P(\emptyset) = 0 \quad (F_3)$

Beweis:

$P(\emptyset) \overset{(F_1)}{=} 1 - P(\overline{\emptyset}) = 1 - P(\Omega) \overset{(2)}{=} 1 - 1 = 0$

$\square$

Folgerung 4

Seien A und B Ereignisse in einem Stichprobenraum Ω mit $A \subset B$ und $P\colon \Sigma \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Dann gilt:

$P(A) \leq P(B) \quad (F_4)$

Beweis:

$P(B) = P(A \cup [B \setminus A]) \overset{(3)}{=} P(A) + P(B \setminus A)$

Wegen $P(B \setminus A) \overset{(1)}{\geq} 0$, muss $P(A) \leq P(B)$ gelten. $\square$

Folgerung 5

Seien A und B Ereignisse in einem Stichprobenraum Ω und $P\colon \Sigma \to \mathbb{R}$ eine Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Dann gilt der *Additionssatz*:

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

Beweis:

$$P(A \cup B) = P(A \cup [\overline{A} \cap B]) \overset{(3)}{=} P(A) + P(\overline{A} \cap B)$$

$$P(B) = P(\Omega \cap B) = P((A \cup \overline{A}) \cap B)$$

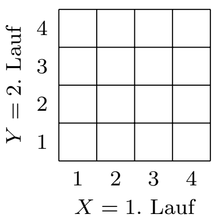
$$= P((A \cap B) \cup (\overline{A} \cap B)) = P(A \cap B) + P(\overline{A} \cap B)$$

$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \square$$

Beispiel 2.3 (**P auf endlichem Stichprobenraum**)

Zwei Läufe mit dem Glücksrad



sinnvoll: jedes Ereignis hat die Wahrscheinlichkeit $\dfrac{1}{16}$

(a) $P\big(\{(x,y)\colon (X,Y) = (1,1) \vee (X,Y) = (3,2)\}\big) = \dfrac{2}{16} = \dfrac{1}{8}$

(b) $P\big(\{(x,y)\colon X = 3\}\big) = \dfrac{4}{16} = \dfrac{1}{4}$

(c) $P\big(\{(x,y)\colon \operatorname{mod}(X+Y,2) = 1\}\big) = \dfrac{8}{16} = \dfrac{1}{2}$

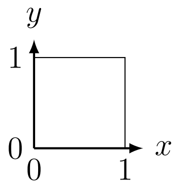
Diskrete Gleichverteilung (*Discrete uniform law*)

Wenn wir annehmen dürfen, dass alle Ereignisse gleichwahrscheinlich sind, gilt:

$$P(A) = \frac{\text{Anzahl Elemente von } A}{\text{Anzahl Elemente von } \Omega} = \frac{\text{günstige Fälle}}{\text{mögliche Fälle}}$$

- Experimente dieser Art heissen auch *Laplace-Experimente*
- Wahrscheinlichkeitsberechnungen $\rightarrow$ Zählen
- Anwendung: fairen Münze, Würfeln, Kartenspiele, ...
- kompliziertere Beispiele $\rightarrow$ *Kombinatorik*

Stetige Gleichverteilung (*Continuous uniform law*)



Wahrscheinlichkeitsfunktion: $P(A) =$ Flächeninhalt von A

(a) $P\big((X,Y) = (0.4, 0.7)\big) = 0$

(b) $P\big(X = 0.6\big) = 0$

(c) $P\big(X + Y \leq \frac{1}{2}\big) = \frac{1}{8}$

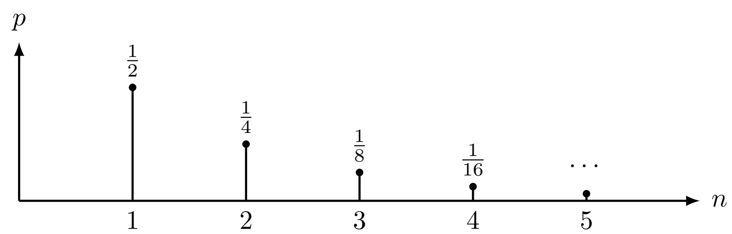
(d) $P\big(X^2 \leq Y\big) = 1 - \int_0^1 x^2 \, \mathrm{d}x = 1 - \left[\frac{1}{3}\right]_0^1 = \frac{2}{3}$

Diskreter unendlicher Wahrscheinlichkeitsraum

Wurf einer Münze, bis zum ersten Mal „Zahl“ erscheint:

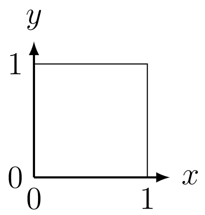
$\Omega = \{Z, WZ, WWZ, WWWZ, WWWWZ, \ldots\}$

$P(\text{Zahl beim } n\text{-ten Wurf}) = P(n) = 2^{-n}$



$$P(\{2,4,6,\ldots\}) \overset{(3')}{=} \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \ldots = \frac{1}{4} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{1}{4} \cdot \frac{4}{3} = \frac{1}{3}$$

Stetiger unendlicher Wahrscheinlichkeitsraum



Wahrscheinlichkeitsfunktion: $P(A) =$ Flächeninhalt von A

$$1 = P(\Omega) = P\left(\bigcup_{x,y}\{(x,y)\}\right) \overset{(3')}{=} \sum_{x,y} P\big(\{(x,y)\}\big) = \sum_{x,y} 0 = 0$$

Die Punkte eines Kontinuums (Strecke, Fläche, ...) lassen sich nicht durch eine Folge P_1, P_2, ... abzählen. Solche Mengen sind *überabzählbar* und (3') darf nicht angewendet werden.

„Rezept“ zur Lösung von Wahrscheinlichkeitsaufgaben

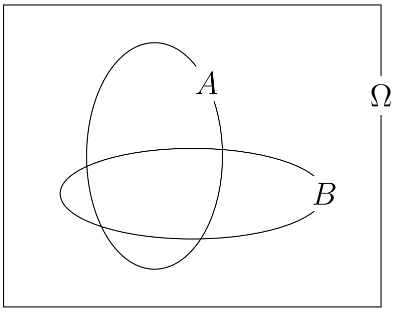
1. Beschreibung des Stichprobenraums, d. h. die Menge aller möglichen Ereignisse des gegebenen Experiments.
2. Wahl einer Wahrscheinlichkeitsfunktion. (Wie wahrscheinlich ist jedes Ereignis?)
3. Berechnung der (bedingten) Wahrscheinlichkeiten der interessierenden Ereignisse.

3 Bedingte Wahrscheinlichkeit

Worum geht es?

Wie verändern zusätzliche Informationen unsere Vorstellungen darüber, wie wahrscheinlich die Ereignisse eines Zufallsexperiments sind?

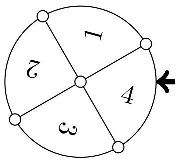
A sei ein Ereignis in Ω. Wie verändert sich unsere Einschätzung von $P(A)$ wenn wir erfahren, dass Ereignis B eingetreten ist?



Bedingte Wahrscheinlichkeit von A gegeben B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{wobei } P(B) \neq 0$$

Das Glücksrad mit vier gleich grossen Feldern, die mit den Zahlen 1, 2, 3 und 4 beschriftet sind, wird anderswo 2 Mal in Gang gesetzt.



A: Die Summe der beiden Zahlen beträgt 4.

$P(A) = P\big(\{(1,3),(2,2),(3,1)\}\big) = \frac{3}{16}$

Wir erfahren, dass ein Ereignis B eingetroffen sei. Wie verändert diese Informationen unsere Einschätzung bezüglich $P(A)$?

(a) B: Beide Zahlen sind ungerade

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(\{(1,3),(3,1)\})}{P(\{(1,1),(1,3),(3,1),(3,3)\})} \\ &= \frac{2}{4} = \frac{1}{2} > \frac{3}{16} \end{aligned}$$

(b) B: die erste Zahl ist gerade

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(\{(2,2)\})}{P(\{(2,1),\ldots,(2,4),(4,1),\ldots,(4,4)\})} \\ &= \frac{1}{8} < \frac{3}{16} \end{aligned}$$

Bedingte Wahrscheinlichkeiten sind Wahrscheinlichkeiten

Sei Ω ein Stichprobenraum mit einer Wahrscheinlichkeitsfunktion, welche die Axiome (1)–(3) erfüllt. Es sei ferner $B \subset \Omega$ ein fest gewähltes Ereignis mit $P(B) \neq 0$. Dann definiert für jedes $A \subset \Omega$ auch

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

eine Wahrscheinlichkeitsfunktion auf Ω. (Beweis: Übungen)

Bedingte Wahrscheinlichkeiten definieren also eine Wahrscheinlichkeitsfunktion relativ zum Ereignisraum B.

Falls Ω aus endlich vielen gleich wahrscheinlichen Ergebnissen besteht, gilt speziell:

$$P(A|B) = \frac{\text{Anzahl der Elemente von } A \cap B}{\text{Anzahl der Elemente von } B}$$

Beispiel 3.1

Eine faire Münze wird dreimal nacheinander geworfen.

Gesucht: $P(A|B)$ wobei

$A = \{\text{mehr Wappen als Zahl}\}$

$B = \{\text{erster Wurf zeigt Wappen}\}$

$\Omega = \{WWW, WWZ, WZW, ZWW, WZZ, ZWZ, ZZW, ZZZ\}$

$B = \{WWW, WWZ, WZW, ZWW\}$

$A \cap B = \{WWW, WWZ, WZW\}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3/8}{4/8} = \frac{3}{4}$$

Beispiel 3.2

In einer Population betrage die relative Häufigkeit, dass ein Individuum einen Risikofaktor für eine bestimmte Krankheit aufweist, 10% (*Prävalenz*).

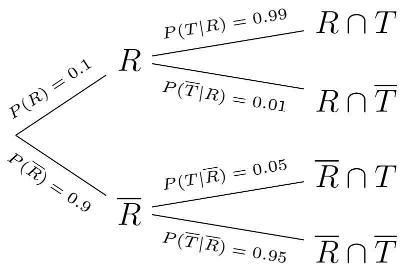
Nun lässt sich das Vorhandensein des Risikofaktors mit einem medizinischen Test untersuchen, der jedoch nicht perfekt ist.

Bei einem Individuum mit Risikofaktor zeigt der Test dies mit 99%-iger Sicherheit an (*Sensitivität*, *true positive rate*).

Bei einem Individuum ohne Risikofaktor zeigt der Test dies mit 95%-iter Sicherheit an. (*Spezifität*, *true false rate*).

$R = \{\text{Person hat Risikofaktor}\}$

$T = \{\text{Test zeigt Risikofaktor an}\}$



$P(R \cap T) = P(R) \cdot P(T|R) = 0.1 \cdot 0.99 = 0.099$

$$\begin{aligned} P(T) &= P(R \cap T) + P(\overline{R} \cup T) = P(R)P(T|R) + P(\overline{R})P(T|\overline{R}) \\ &= 0.1 \cdot 0.99 + 0.9 \cdot 0.05 = 0.144 \end{aligned}$$

$P(R|T) = \dfrac{P(R \cap T)}{P(T)} = \dfrac{0.099}{0.144} = 0.6875$ Ist das gut?

Multiplikationssatz

Umformung der Definition der bedingten Wahrscheinkeit

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

ergibt den *Multiplikationssatz*:

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

Um die Wahrscheinlichkeit von „A und B“ zu berechnen, muss die Wahrscheinlichkeit von A mit der Wahrscheinlichkeit von „B gegeben A“ multipliziert werden (oder umgekehrt).

Verallgemeinerung

Sind $A_1, A_2, A_3 \subset \Sigma$, so gilt die *Kettenregel*:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$

Analoge Formeln gelten für 4, 5, ... Ereignisse.

Beweis: $P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$

$$\overset{\text{def.}}{=} P(A_1) \cdot \frac{P(A_2 \cap A_1)}{P(A_1)} \cdot \frac{P(A_3 \cap A_1 \cap A_2)}{P(A_1 \cap A_2)}$$

$$= P(A_1 \cap A_2 \cap A_3) \quad \square$$

Beispiel 3.3

Ein gut gemischtes Blatt französischer Spielkarten bestehe aus jeweils 9 Karten der folgenden Farben:

Kreuz	Pik	Herz	Karo
$\clubsuit$	$\spadesuit$	$\heartsuit$	$\diamondsuit$

Es werden zufällig drei Karten ohne Zurücklegen gezogen. Wie gross ist die Wahrscheinlichkeit, dass keine Herz-Karte darunter ist?

Ereignis A_i: i-te Karte ist keine Herzkarte ($i = 1, 2, 3$)

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2)$$

$$= \frac{27}{36} \cdot \frac{26}{35} \cdot \frac{25}{34} = \frac{195}{476}$$

Zusammenfassung

Die Berechnung von Wahrscheinlichkeiten bei mehrstufigen Zufallexperimenten lässt sich wie folgt beschreiben:

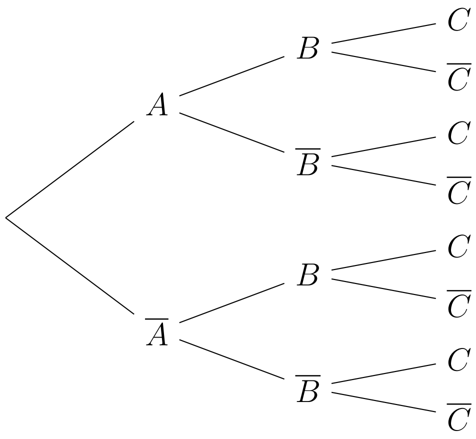
(1) Stelle das Baumdiagramm (oder einen Teil davon) so dar, dass das interessierende Ereignis E an einem Blatt auftritt.

(2) Ermittle die (bedingten) Wahrscheinlichkeiten entlang der Äste des Baums.

(3) Multipliziere die in (2) berechneten Wahrscheinlichkeiten von der Wurzel bis zum Blatt (erste Pfadregel).

Wenn das gesuchte Ereignis aus mehreren Endknoten des Baums besteht, sind die einzelnen Resultate zu addieren (2. Pfadregel).

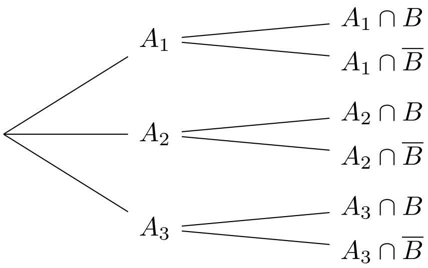
Dreistufiges Zufallsexperiment mit jeweils 2 Resultaten



$P(\{A\overline{B}C, \overline{A}BC\}) = P(A)P(\overline{B}|A)P(C|A \cap \overline{B}) + P(\overline{A})P(B|\overline{A})P(C|\overline{A} \cap B)$

4 Die Formel von Bayes

Satz von der totalen Wahrscheinlichkeit



$$
\begin{aligned}
P(B) &\overset{\text{ML}}{=} P(\Omega \cap B) \overset{\text{ML}}{=} P([A_1 \cup A_2 \cup A_3] \cap B) \\
&\overset{\text{DG}}{=} P([A_1 \cap B] \cup [A_2 \cap B] \cup [A_3 \cap B]) \\
&\overset{\text{KA}}{=} P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\
&\overset{\text{KR}}{=} P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)
\end{aligned}
$$

Beispiel 4.1

Ein Tennisspieler tritt in einem Turnier im ersten Match gegen einen von 15 Spielern an. Die folgende Tabelle besagt, wie viele Gegner sich in jeder der drei Stärkeklassen befinden und wie gross (erfahrungsgemäss) die mittlere Wahrscheinlichkeit ist, gegen einen Opponenten dieser Klasse zu gewinnen.

Stärkeklasse	Anzahl Spieler	Gewinnwahrscheinlichkeit
K_1	9	0.6
K_2	4	0.3
K_3	2	0.1

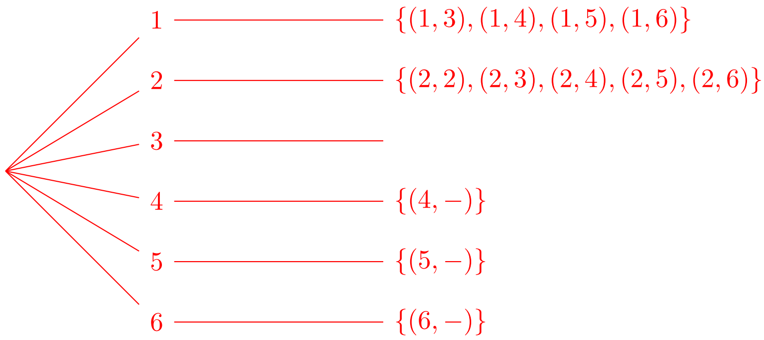
Wie gross ist die Wahrscheinlichkeit, gegen einen ausgelosten Gegner zu gewinnen?

Ereignis G: Der Spieler gewinnt das Match

$$
\begin{aligned}
P(G) &= P(K_1) \cdot P(G|K_1) + P(K_2) \cdot P(G|K_2) + P(K_3) \cdot P(G|K_3) \\
&= \frac{9}{15} \cdot \frac{6}{10} + \frac{4}{15} \cdot \frac{3}{10} + \frac{2}{15} \cdot \frac{1}{10} = \frac{68}{150} = \frac{34}{75} \approx 0.453
\end{aligned}
$$

Beispiel 4.2

Ein fairer Spielwürfel wird einmal geworfen. Wenn die Augenzahl 1 oder 2 ist, wirft man eine weiteres Mal. Wie gross ist die Wahrscheinlichkeit, dabei eine Augenzahl(summe) von 4 oder mehr zu erhalten?



Ereignis A: Augenzahl(summe) ist grösser oder gleich 4

$$P(A) = \frac{1}{6} \cdot \frac{4}{6} + \frac{1}{6} \cdot \frac{5}{6} + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1 = \frac{3}{4}$$

Formel von Bayes

Bilden $A_1, A_2, \ldots, A_n$ eine Partition von Ω mit $P(A_i) > 0$ für alle i, dann gilt für jedes Ereignis B mit $P(B) > 0$:

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)} \end{aligned}$$

Beweis

$$\begin{aligned} P(A_i|B) &\overset{\text{Def.}}{=} \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i) \cdot P(B|A_i)}{P(B)} \quad \text{(Kettenregel)} \\ &= \frac{P(A_i) \cdot P(B|A_i)}{P(A_1) \cdot P(B|A_1) + \cdots + P(A_n) \cdot P(B|A_n)} \\ &\quad \text{(Satz von der totalen W'keit)} \end{aligned}$$

□

Beispiel 4.3

Drei Anlagen produzieren das gleiche elektronische Bauteil in unterschiedlichen Mengen und in unterschiedlicher Qualität:

	Anlage 1	Anlage 2	Anlage 3
Menge (Stück/Tag)	50 000	10 000	40 000
davon defekt	2%	7%	3%

Ein ungeprüftes Bauteil gelangt versehentlich in den Verkauf und wird wird vom Kunden wegen eines Defekts zurückegeben. Mit welcher Wahrscheinlichkeit stammt es von Anlage 1?

A_i: Bauteil stammt von Anlage i ($i = 1, 2, 3$)

D: Bauteil ist defekt

$$\begin{aligned}
P(A_1|D) &= \frac{P(A_1 \cap D)}{P(D)} \\
&= \frac{P(A_1)P(D|A_1)}{P(A_1)P(D|A_1) + P(A_2)P(D|A_2) + P(A_3)P(D|A_3)} \\
&= \frac{0.5 \cdot 0.02}{0.5 \cdot 0.02 + 0.1 \cdot 0.07 + 0.4 \cdot 0.03} \\
&= 0.345
\end{aligned}$$

Inferenz

$$A_i \xrightarrow[P(B|A_i)]{\text{Ursache – Wirkung}} B$$

$$A_i \xleftarrow[P(A_i|B)]{\text{Schlussfolgerung (Inferenz)}} B$$

Die Formel von Bayes erlaubt in systematischer Weise, aus den Ergebnissen von Experimenten neue Aussagen zu gewinnen.

5 Unabhängigkeit

5.1 Unabhängigkeit von zwei Ereignissen

Unabhängige Ereignisse

Mittels der bedingten Wahrscheinlichkeit

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

lässt sich die Wahrscheinlichkeit eines Ereignisses A „relativieren“, wenn bekannt ist, dass das Ereignis B eingetreten ist.

Wir befassen uns jetzt mit dem Fall, dass diese Information über das Eintreffen von B *nicht* zu einer Neubewertung von $P(A)$ führt. In diesem Fall gilt nämlich:

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

Multiplizieren wir in der letzten Gleichung die Identität rechts mit $P(B)$, erhalten wir eine Charakterisierung der stochastischen Unabhängigkeit zweier Ereignisse A und B, die auch bei $P(B) = 0$ anwendbar ist:

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

Beispiel 5.1

Ein fairer Spielwürfel wird zweimal geworfen.

- A: beim ersten Wurf erscheint die Augenzahl 3
- B: beim zweiten Wurf erscheint die Augenzahl 5

Sind die Ereignisse A und B unabhängig?

$P(A) = P(\{(3,1),(3,2),\ldots,(3,6)\}) = 6/36$

$P(B) = P(\{(1,5),(2,5),\ldots,(6,5)\}) = 6/36$

$P(A \cap B) = P(\{(3,5)\}) = 1/36$

$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(A \cap B) \quad \Rightarrow \quad$ unabhängig

Beispiel 5.2

Ein fairer Spielwürfel wird zweimal geworfen.

- A: beim ersten Wurf erscheint die Augenzahl 3
- B: die Summe beider Augenzahlen beträgt 5

Sind die Ereignisse A und B unabhängig?

$P(A) = P(\{(3,1),(3,2),\dots,(3,6)\}) = 6/36 = 1/6$

$P(B) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = 4/36 = 1/9$

$P(A \cap B) = P(\{(3,2)\}) = 1/36$

$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{9} = \frac{1}{54} \neq \frac{1}{36} = P(A \cap B) \quad \Rightarrow \quad \text{abhängig}$

Beispiel 5.3

Ein fairer Spielwürfel wird zweimal geworfen.

- A: beim ersten Wurf erscheint die Augenzahl 3
- B: die Summe beider Augenzahlen beträgt 2

Sind die Ereignisse A und B unabhängig?

$P(A) = P(\{(3,1),(3,2),\dots,(3,6)\}) = \frac{6}{36} = \frac{1}{6}$

$P(B) = P(\{(1,1)\}) = 1/36$

$P(A \cap B) = P(\{\,\}) = 0$

$P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{216} \neq 0 = P(A \cap B) \quad \Rightarrow \quad \text{abhängig}$

5.2 Bedingte Unabhängigkeit von zwei Ereignissen

Bedingt unabhängige Ereignisse

Sind A und B zwei Ereignisse und ist C ein Ereignis, das bereits eingetroffen ist, so sind A und B unabhängig gegeben C:

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

Beispiel 5.4

Ein fairer Spielwürfel wird einmal geworfen

- $A = \{1, 2, 3, 5\}$
- $B = \{1, 2, 5, 6\}$
- $C = \{2, 3, 4, 6\}$

(a) Sind die Ereignisse A, B bedingt unabhängig gegeben C?

(b) Sind die Ereignisse A, B unabhängig?

(a) $P(A|C) = \dfrac{P(\{2,3\})}{P(\{2,3,4,6\})} = \dfrac{1}{2}$

$P(B|C) = \dfrac{P(\{2,6\})}{P(\{2,3,4,6\})} = \dfrac{1}{2}$

$P(A \cap B|C) = \dfrac{P(\{2\})}{P(\{2,3,4,6\})} = \dfrac{1}{4}$

$P(A|C) \cdot P(B|C) = \dfrac{1}{2} \cdot \dfrac{1}{2} = \dfrac{1}{4} = P(A \cap B|C)$

A und B sind bedingt unabhängig gegeben C.

(b) $P(A) = P(\{1,2,3,5\}) = \dfrac{2}{3}$

$P(B) = P(\{1,2,5,6\}) = \dfrac{2}{3}$

$P(A \cap B) = \{1,2,5\} = \dfrac{1}{2}$

$P(A) \cdot P(B) = \dfrac{2}{3} \cdot \dfrac{2}{3} = \dfrac{4}{9} \neq \dfrac{1}{2} = P(A \cap B)$

A und B sind abhängig.

Beispiel 5.5

Ein fairer Spielwürfel wird einmal geworfen.

- $A = \{1, 2, 5, 6\}$
- $B = \{1, 4, 6\}$
- $C = \{2, 3, 4, 5, 6\}$

(a) Sind die Ereignisse A, B bedingt unabhängig gegeben C?

(b) Sind die Ereignisse A, B unabhängig?

(a) $P(A|C) = \dfrac{P(\{2,5,6\})}{P(\{2,3,4,5,6\})} = \dfrac{3}{5}$

$P(B|C) = \dfrac{P(\{4,6\})}{P(\{2,3,4,5,6\})} = \dfrac{2}{5}$

$P(A \cap B|C) = \dfrac{P(\{6\})}{P(\{2,3,4,5,6\})} = \dfrac{1}{5}$

$P(A|C) \cdot P(B|C) = \dfrac{3}{5} \cdot \dfrac{2}{5} = \dfrac{6}{25} \neq \dfrac{1}{5} = P(A \cap B|C)$

A und B sind bedingt abhängig gegeben C.

(b) $P(A) = P(\{1,2,5,6\}) = \dfrac{2}{3}$

$P(B) = P(\{1,4,6\}) = \dfrac{1}{2}$

$P(A \cap B) = \{1,6\} = \dfrac{1}{3}$

$P(A) \cdot P(B) = \dfrac{2}{3} \cdot \dfrac{1}{2} = \dfrac{1}{3} = P(A \cap B)$

A und B sind unabhängig.

Die letzten beiden Beispiel zeigen, dass ...

- aus der bedingten Unabhängigkeit nicht zwingend die Unabhängigkeit folgt,
- aus der Unabhängigkeit nicht zwingend die bedingte Unabhängigkeit folgt.

5.3 Unabhängigkeit mehrerer Ereignisse

Damit drei Ereignisse A, B und C unabhängig sind, müssen die folgenden vier Bedingungen erfüllt sein:

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

(2) $P(B \cap C) = P(B)P(C)$

(3) $P(C \cap A) = P(C)P(A)$

(4) $P(A \cap B \cap C) = P(A)P(B)P(C)$

Wir werden Beispiele sehen, die zeigen, dass im Allgemeinen auf keine dieser Bedingungen verzichtet werden kann.

Beispiel 5.6

Eine faire Münze werden zweimal geworfen.

- A: Der erste Wurf ist „Wappen“
- B: Der zweite Wurf ist „Zahl“
- C: Die beiden Resultate fallen gleich aus.

$P(A) = P(\{WW\}, \{WZ\}) = 1/2$

$P(B) = P(\{WZ\}, \{ZZ\}) = 1/2$

$P(C) = P(\{WW\}, \{ZZ\}) = 1/2$

$P(A \cap B) = P(\{WZ\}) = 1/4$

$P(B \cap C) = P(\{ZZ\}) = 1/4$

$P(C \cap A) = P(\{WW\}) = 1/4$

$P(A \cap B \cap C) = P(\{\,\}) = 0$

offensichtlich sind (1)–(3) erfüllt aber (4) nicht.

Definition der Unabhängigkeit von n Ereignissen

Allgemein sind n Ereignisse A_1, A_2, ..., A_n unabhängig, wenn

$$P\left(\bigcap_{i\in S} A_i\right) = \prod_{i\in S} P(A_i) \quad \textit{für jede } \text{Teilmenge } S \subset \{1, 2, \dots, n\}.$$

Beispiel 5.7

Für 4 Ereignisse A, B, C, D muss man demnach bereits

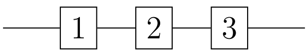
- für jeweils zwei Mengen: 6 Bedingungen
- für jeweils drei Mengen: 4 Bedingungen
- für alle vier Mengen: 1 Bedingung

überprüfen!

Zuverlässigkeit

Werden komplexe Systeme, die aus mehreren Komponenten bestehen, durch ein Wahrscheinlichkeitsmodell dargestellt, ist es manchmal sinnvoll anzunehmen, dass diese Teile *unabhängig* voneinander ausfallen. Dadurch kann der Rechenaufwand reduziert werden.

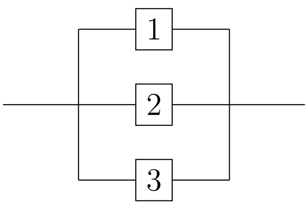
Serieschaltung von Komponenten



Das Gesamtsystem ist intakt, wenn *alle* Komponenten intakt sind.

$P(\text{Gesamtsystem funktioniert}) = p_1 p_2 p_3$

Parallelschaltung von Komponenten

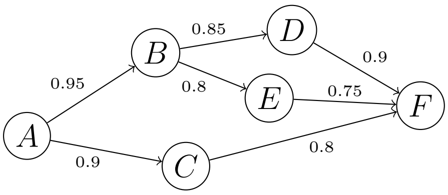


Das Gesamtsystem ist intakt, wenn *mindestens ein* Komponente intakt ist. Dies ist gleichbedeutend damit, dass *nicht alle* Komponenten gleichzeitig *versagen*.

$P(\text{Gesamtsystem funktioniert}) = 1 - (1 - p_1)(1 - p_2)(1 - p_3)$

Beispiel 5.8

Ein Computernetzwerk verbindet die Knoten A und F durch die Zwischenstationen B, C, D und E. Über jeder Verbindung steht die Wahrscheinlichkeit, dass sie zu einem bestimmten Zeitpunkt zuverlässig funktioniert. Wir nehmen an, dass sich Ausfälle der Router unabhängig voneinander ereignen.

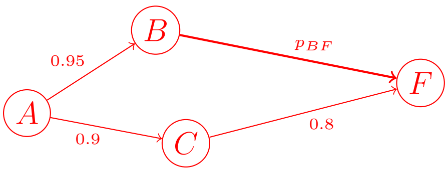


Mit welcher Wahrscheinlichkeit funktioniert die Verbindung von A nach F zuverlässig?

Der Weg von B nach F besteht aus zwei parallel verbundenen Subsystemen, die jeweils seriell aufgebaut sind:

$$\begin{aligned} P(BF) &= 1-(1-P(BD,DF))(1-P(BE,EF)) \\ &= 1-(1-p_{BD}p_{DF})(1-p_{BE}p_{EF}) \\ &= 1-(1-0.85\cdot 0.9)(1-0.8\cdot 0.75) = 0.906 \end{aligned}$$

Damit lässt sich das System reduzieren:



Der Weg von A nach F besteht ebenfalls aus zwei parallel geschalteten Subsystemen mit einer jeweils seriellen Struktur:

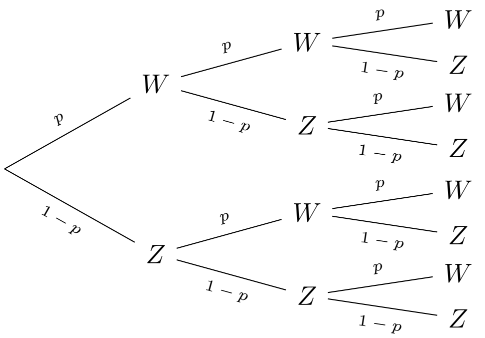
$$\begin{aligned} P(AF) &= 1-(1-P(AB,BF))(1-P(AC,CF)) \\ &= 1-(1-p_{AB}p_{BF})(1-p_{AC}p_{CF}) \\ &= 1-(1-0.95\cdot 0.906)(1-0.9\cdot 0.8) \\ &= 0.960996 \end{aligned}$$

5.4 Bernoulli-Versuche

Wir sprechen von einen *Bernoulli-Versuch*, wenn wir es mit einer Folge *identischer*, *unabhängiger* und *dichotomer* (zweiwertiger) Zufallsexperimente zu tun haben.

Beispiel 5.9

Eine Münze mit $P(Wappen) = p$ wird 3 Mal nacheinander geworfen. Berechne formal die Wahrscheinlichkeiten aller Elementarereignisse.



$P(WWW) = p^3$

$P(WWZ) = p^2(1-p)$

$P(WZW) = p^2(1-p)$

$P(WZZ) = p(1-p)^2$

$P(ZWW) = p^2(1-p)$

$P(ZWZ) = p(1-p)^2$

$P(ZZW) = p(1-p)^2$

$P(ZZZ) = (1-p)^3$

Die Wahrscheinlichkeit, bei einem n-fachen Münzwurf eine *bestimmte Folge* mit k Mal Wappen und $(n-k)$ Mal Zahl zu erhalten, beträgt:

$p^k(1-p)^{n-k}$ für $k \in \{0, 1, 2, \ldots, n\}$

Beispiel 5.10

$P(WWWZWWZZZ) = p^5(1-p)^4$

$P(ZZZZWWWWW) = p^5(1-p)^4$

In der Regel interessiert man sich bei einem Bernoulli-Versuch nicht für eine bestimmte Folge von Resultaten, sondern dafür, *wie oft* eines der beiden Ergebnisse eintritt:

$$B(n, p, k) = P(\text{Bei } n \text{ Münzwürfen erscheint genau } k \text{ mal Kopf})$$

Wir wissen bereits, wie man die Wahrscheinlichkeit eines einzelnen Ergebnisses berechnet. Um die Anzahl der Ergebnisse zu zählen, die genau k Mal Wappen und und $n-k$ Mal Kopf enthalten, benötigen wird den Binomialkoeffizienten, der im nächsten Kapitel genauer besprochen wird:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+2) \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 2 \cdot 1}$$

Die Formel rechts eignet sich besser zur manuellen Berechnung.

Die Wahrscheinlichkeitsfunktion eines Bernoulli-Versuchs

Die Wahrscheinlichkeit, bei einer Folge von n identischen, unabhängigen und dichotomen Experimenten mit der „Erfolgswahrscheinlichkeit“ p genau k Erfolge zu beobachten, beträgt:

$$B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Beispiel 5.11

Wie gross ist die Wahrscheinlichkeit, bei einer Folge von 10 Würfen mit einem idealen Spielwürfel genau 4 Mal eine Augenzahl kleiner oder gleich 2 zu erhalten?

k: Anzahl der Würfe mit Augenzahl ≤ 2

$$\begin{aligned} B(n=10, p=\frac{1}{3}, k=4) &= \binom{10}{4} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^6 \\ &= \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} \cdot \frac{1}{81} \cdot \frac{64}{729} \\ &= \frac{4480}{19683} \approx 0.2276 \end{aligned}$$

TI-84+: 2nd/DIST/binompdf(10,1/3,4)

Beispiel 5.12

Ein kleiner Internet Service Provider (ISP) hat $n = 100$ Privatkunden und $c = 30$ IPv4-Adressen zur Verfügung. Da jeder Rechner im Internet eindeutig identifizierbar sein muss, kann jede dieser IPv4-Adressen zu jedem Zeitpunkt nur einmal verwendet werden.

Die Wahrscheinlichkeit, dass ein Kunde zur „Busy Hour“ ins Internet will, sei $p = 0.2$ und soll unabhängig von den übrigen Benutzern sein.

Wie gross ist die Wahrscheinlichkeit, dass mehr als 30 Kunden gleichzeitig aufs Internet zugreifen wollen und der ISP seine Dienstleistung nicht mehr vollständig erbringen kann?

A: ≥ 30 Kunden benötigen eine IPv4-Adresse

$$\begin{aligned} P(A) &= \sum_{k=31}^{100} B(100, 0.2, k) \\ &= 1 - \sum_{k=0}^{30} B(100, 0.2, k) \\ &= 1 - \sum_{k=0}^{30} \binom{100}{k} 0.2^k 0.8^{n-k} \\ &\approx 0.00606 \end{aligned}$$

TI-84+: 2nd/DIST/binomcdf(100,0.2,30)

5.5 Zusammenfassung

- Zwei Ereignisse A und B sind unabhängig, wenn

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

  Für $P(B) > 0$ ist dies gleichbedeutend mit $P(A|B) = P(A)$.

- Wenn A und B unabhängig sind, dann auch A und B^c.

- A und B sind bedingt unabhängig, gegeben C, wenn

  $P(C) > 0$ und $P(A \cap B|C) = P(A|C)P(B|C)$.

- Die bedingte Unabhängigkeit folgt im Allgemeinen nicht aus der Unabhängigkeit und umgekehrt.

- n Ereignisse $A_1, A_2, \ldots, A_n$ sind unabhängig, wenn

  $P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$ *für jede* Teilmenge $S \subset \{1, 2, \ldots, n\}$.

- Wahrscheinlichkeitsfunktion eines Bernoulli-Versuchs:

  $B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k}$

6 Kombinatorik

Motivation

Ist Ω endlich und sind alle Ergebnisse gleich wahrscheinlich, so können wir $P(A)$ durch *Abzählen* der „günstigen" Ergebnisse bestimmen:

$$p(A) = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega}$$

Die mathematische Disziplin der *Kombinatorik* befasst sich unter anderem damit, solche Abzählvorgänge systematisch zu erfassen und berechenbar zu machen.

6.1 Die Grundregeln

Die Produktregel der Kombinatorik

Gibt es in einem k-stufigen Abzählvorgang in jeder Stufe m_i Auswahlmöglichkeiten, so lässt sich die Gesamtzahl aller Möglichkeiten durch das Produkt

$$m_1 \cdot m_2 \cdot \ldots \cdot m_k$$

berechnen.

Beispiel 6.1

Eine Person hat

- 5 Paar verschiedene Schuhe
- 7 Paar verschiedene Socken
- 4 Paar verschiedene Hosen
- 8 verschiedene Sweatshirts

Auf wie viele verschiedene Arten kann sich die Person kleiden?

auf $5 \cdot 7 \cdot 4 \cdot 8 = 1120$ Arten

Die Summenregel

Lässt sich ein Abzählvorgang in k disjunkte Teilvorgänge mit jeweils m_i Auswahlmöglichkeiten zerlegen so lässt sich die Gesamtzahl aller Möglichkeiten durch die Summe

$$m_1 + m_2 + \cdots + m_k$$

berechnen.

Beispiel 6.2

Wie viele zwei- oder dreidimensionale Vektoren lassen sich mit den Komponenten $1, 2, 3$ bilden?

- zweidimensional: $(1,1)^{\mathrm{T}}$, $(1,2)^{\mathrm{T}}$, $\ldots$, $(3,3)^{\mathrm{T}}$

  Produktregel: $3 \cdot 3 = 9$ Vektoren

- dreidimensional: $(1,1,1)^{\mathrm{T}}$, $(1,1,2)^{\mathrm{T}}$, $\ldots$, $(3,3,3)^{\mathrm{T}}$

  Produktregel: $3 \cdot 3 \cdot 3 = 27$ Vektoren

Die Mengen der zwei- und dreidimensionalen Vektoren haben keine gemeinsamen Elemente; also keine Gefahr von Mehrfachzählungen:

Summenregel: $9 + 27 = 36$ Vektoren

Die Gleichheitsregel

Existiert eine bijektive, d. h. eindeutig umkehrbare, Abbildung zwischen zwei Mengen A und B, so ist die Anzahl Elemente von A gleich der Anzahl der Elemente von B.

Beispiel 6.3

Wie viele Teilmengen hat eine Menge A mit $n = 5$ Elementen?

$A = \{a_1, a_2, a_3, a_4, a_5\}$

Betrachte $B = \big\{(b_1, b_2, b_3, b_4, b_5) \mid b_i \in \{0, 1\}\big\}$

Die 5-Tupel in B entsprechen bijektiv den Teilmengen von A:

- $(0,1,1,0,1)$ entspricht $\{a_2, a_3, a_5\}$
- $(0,0,0,0,0)$ entspricht $\{\,\}$
- $(1,1,1,1,1)$ entspricht A

Somit: $|A| = |B| = 2^5 = 32$ Teilmengen (Produktregel)

6.2 Variationen

Ist $A = \{a_1, a_2, \ldots, a_n\}$ eine Menge und $k \in \mathbb{N}_0$, so wird ein k-Tupel mit Elementen aus A eine *Variation* oder *geordnete Stichprobe* genannt.

Abhängig davon, ob man Wiederholungen von Elementen zulässt oder nicht, spricht man von *Variationen mit Wiederholungen* bzw. von *Variationen ohne Wiederholungen*.

Variationen mit Wiederholungen

Die *Variationen mit Wiederholungen* der Länge k einer Menge $A = \{a_1, a_2, \dots, a_n\}$ ist die Menge aller k-Tupel mit Elementen aus A, wobei ein Element auch mehrfach vorkommen darf.

$(\square_1, \square_2, \dots, \square_{k-1}, \square_k)$

An jeder Position $\square_i$ kann ein beliebiges der n Elemente stehen.

Produktregel:

$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ Faktoren}} = n^k$ *Anzahl Variationen mit Wiederholungen*

Beispiel 6.4

$M = \{a, b, c, d\}$, $k = 3$

(a,a,a)	(b,a,a)	(c,a,a)	(d,a,a)
(a,a,b)	(b,a,b)	(c,a,b)	(d,a,b)
(a,a,c)	(b,a,c)	(c,a,c)	(d,a,c)
(a,a,d)	(b,a,d)	(c,a,d)	(d,a,d)
(a,b,a)	(b,b,a)	(c,b,a)	(d,b,a)
(a,b,b)	(b,b,b)	(c,b,b)	(d,b,b)
(a,b,c)	(b,b,c)	(c,b,c)	(d,b,c)
(a,b,d)	(b,b,d)	(c,b,d)	(d,b,d)
(a,c,a)	(b,c,a)	(c,c,a)	(d,c,a)
(a,c,b)	(b,c,b)	(c,c,b)	(d,c,b)
(a,c,c)	(b,c,c)	(c,c,c)	(d,c,c)
(a,c,d)	(b,c,d)	(c,c,d)	(d,c,d)
(a,d,a)	(b,d,a)	(c,d,a)	(d,d,a)
(a,d,b)	(b,d,b)	(c,d,b)	(d,d,b)
(a,d,c)	(b,d,c)	(c,d,c)	(d,d,c)
(a,d,d)	(b,d,d)	(c,d,d)	(d,d,d)

Beispiel 6.5

Wie viele Passwörter mit fünf Zeichen lassen sich aus den 26 Grossbuchstaben des lateinischen Alphabets bilden, wenn die Zeichen beliebig gewählt werden können?

$26 \cdot 26 \cdot 26 \cdot 26 \cdot 26 = 11\,881\,376$

Variationen ohne Wiederholungen

Die *Variationen ohne Wiederholungen* der Länge $k \leq n$ einer Menge $A = \{a_1, a_2, \dots, a_n\}$ ist die Menge aller k-Tupel mit Elementen aus A, wobei ein Element höchstens einmal vorkommen darf.

$(\square_1, \square_2, \dots, \square_{k-1}, \square_k)$

An jeder Position $\square_i$ kann eines der unverbrauchten $n - i + 1$ Elemente stehen.

Produktregel:

$\underbrace{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}_{k \text{ Faktoren}}$ *Anzahl Variationen ohne Wh.*

Alternative Darstellung:

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$$
$$= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \cdot (n-k) \cdot (n-k-1) \cdot \ldots \cdot 2 \cdot 1}{(n-k) \cdot (n-k-1) \cdot \ldots \cdot 2 \cdot 1}$$
$$= \frac{n!}{(n-k)!}$$

Spezialfall $n = k$:

$$n(n-1)(n-2) \cdot \cdots \cdot 2 \cdot 1 = n! \qquad \textit{(Permutation)}$$

Beispiel 6.6

$M = \{a, b, c, d\}$, $k = 3$

$\{a, b, c\}$ $\{a, b, d\}$ $\{a, c, d\}$ $\{b, c, d\}$)

Beispiel 6.7

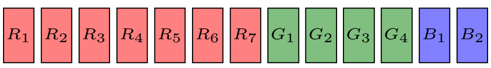
Wie viele Passwörter mit fünf Zeichen lassen sich aus den 26 Grossbuchstaben des lateinischen Alphabets bilden, wenn die Zeichen verschiedenen sein sollen?

$$26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 = \frac{26!}{21!} = 7\,893\,600$$

Beispiel 6.8

Auf wie viele Arten können 7 Romane, 4 Geschichtsbücher und 2 Biologie-Bücher in einem Regal angeordnet werden, wenn alle Bücher der gleichen Art nebeneinander stehen sollen?

Romane, Mathe, Lexika: $7! \cdot 4! \cdot 2!$



Anordnungen der 3 Themen: $3!$

Produktregel: $(7! \cdot 4! \cdot 2!) \cdot 3! = 1\,451\,520$

Bemerkung

Im englischen Sprachraum werden andere Ausdrücke verwendet:

(a) Variationen mit Wiederholungen: *Words*

(b) Variationen ohne Wiederholungen: k-*Permutations*,

Spezialfall $k = n$: *Permutations*

6.3 Kombinationen

Ist $A = \{a_1, a_2, \ldots, a_n\}$ eine Menge und $k \in \mathbb{N}_0$, so wird eine Teilmenge mit k Elementen aus A eine *Kombination* oder *ungeordnete Stichprobe* genannt.

Abhängig davon, ob man Wiederholungen von Elementen (sogenannte *Multimengen*) zulässt oder nicht, spricht man von *Kombinationen mit Wiederholungen* bzw. von *Kombinationen ohne Wiederholungen*.

Kombinationen ohne Wiederholungen

Die *Kombinationen ohne Wiederholungen* der Länge k einer Menge $A = \{a_1, a_2, \ldots, a_n\}$ mit $k \leq n$ ist die Menge aller k-elementigen Teilmengen von A

Ist $C(n,k)$ die gesuchte Anzahl, so können wir jede dieser Kombinationen auf durch $k!$ Arten zu Variationen $V(n,k)$ ohne Wiederholung machen.
Somit:

$$V(n,k) = \frac{n!}{(n-k)!} = C(n,k) \cdot k! \quad \Rightarrow \quad C(n,k) = \frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdot \ldots \cdot (n-k+1)}{k!}$$

(Kombinationen ohne Wiederholung)

Binomialkoeffizienten

Der *Binomialkoeffizient* $\binom{n}{k}$ ist der Koeffizient c_k im Term $c_k a^k b^{n-k}$ des ausgerechneten Binoms $(a+b)^n$ mit $0 \leq k \leq n$.

$$\begin{array}{lccccccccccc}
 & & & & & & & k=0 & & k=1 & & k=2 \\
n=0 & & & & & & 1 & & & & & \\
n=1 & & & & & 1 & & 1 & & & & \\
n=2 & & & & 1 & & 2 & & 1 & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & \\
 & 1 & & \cdots & & \cdots & & \cdots & & \cdots & & 1
\end{array}$$

Eigenschaften

- Symmetrie: $\binom{n}{k} = \binom{n}{n-k}$
- Additivität: $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$
- $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Beispiel 6.9

Auf wie viele Arten kann man aus einer Schulklasse mit 9 Schülerinnen und 8 Schülern eine Volleyballmannschaft mit

(a) 6 Spielenden

(b) 3 Spielerinnen und 3 Spielern

bilden?

(a) $\binom{17}{6} = 12\,376$

(b) $\binom{9}{3} \cdot \binom{8}{3} = 4704$

Multimengen

Da in einer Menge ein Element höchstens einmal aufgezählt werden darf, benötigen wir für den zweiten Fall das Konzept der Multimenge.

Eine *Multimenge* M ist ein Paar (A, f), bestehend aus einer Menge A und einer Abbildung $f\colon A \to \mathbb{N}_0$, die angibt, wie oft das Element $a \in A$ in der Multimenge M vorkommt.

Beispiel 6.10

Gegeben: $A = \{a, b, c, d\}$ und f mit $f(a) = 2,\ f(b) = 3,\ f(c) = 1,\ f(d) = 0$

Gesucht: Multimenge $M = (A, f)$

$M = \{a, a, b, b, b, c\}$

Da auch in Multimengen die Reihenfolge der Elemente keine Rolle spielt, wird (ohne Beschränkung der Allgemeinheit) eine feste Reihenfolge der Elemente aus A festgelegt (z. B. alphabetisch).

$$\underbrace{aa\ldots abb\ldots bcc\ldots c}_{k \text{ Positionen}}$$

Hat die Grundmenge A die Mächtigkeit n, werden zusätzlich $n-1$ „Trennsymbole“ eingefügt:

$$\underbrace{aa\ldots a|bb\ldots b|cc\ldots c}_{k+n-1 \text{ Positionen}}$$

Um alle Multimengen darzustellen, genügt es, die $n-1$ Positionen der Trennsymbole aus allen möglichen $k+n-1$ Plätzen auszuwählen. *Alternative:* Wähle die Positionen aller k Elemente aus allen $k+n-1$ Plätzen aus.

Kombinationen mit Wiederholungen

Damit lässt sich die folgende Aufgabe lösen:

Gegeben: $A = \{a_1, a_2, \ldots, a_n\}$ und $k \in \mathbb{Z}$ mit $0 \leq k$

Gesucht: Anzahl Multimengen $B \subset A$ mit $|B| = k$

Kombinationen mit Wiederholungen:

$$\binom{k+n-1}{n-1} = \binom{k+n-1}{k}$$

Beispiel 6.11

Gegeben: ausreichend rote, weisse, orange und gelbe Rosen

Gesucht: auf wie viele Arten kann man (ohne farbliche Einschränkungen) 10 Rosen zu einem Strauss binden?

Lösung:

$$\binom{10+(4-1)}{10} = \binom{13}{10} = \binom{13}{3} = \frac{13 \cdot 12 \cdot 11}{3 \cdot 2 \cdot 1} = 286$$

Beispiel 6.12

Gesucht: auf wie viele Arten kann man 10 Kekse an 3 Kinder verteilen, wenn jedes Kind mindestens zwei Kekse bekommen soll?

Lösung:

Gibt man jedem Kind die obligatorischen 2 Kekse, dann sind noch 4 Kekse zur freien Verteilung vorhanden.

$$\binom{4+(3-1)}{3-1} = \binom{6}{2} = \frac{6 \cdot 5}{2 \cdot 1} = 15 \text{ Arten}$$

Der Multinomialkoeffizient

Auf wie viele Arten können n verschiedene Objekte restlos auf r verschiedene Behälter verteilt werden, wenn der i-te Behälter am Ende genau k_i Objekte enthalten soll.

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-k_2-\ldots-k_{r-1}}{k_r}$$
$$=\frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdot\ldots\cdot\frac{(n-k_1-\ldots-k_{r-1})!}{k_r!(n-k_1-\ldots-k_r)!}$$
$$=\frac{n!}{k_1!k_2!\ldots k_r!}\overset{\text{Def.}}{=}\binom{n}{k_1,k_2,\ldots,k_r}$$

wobei $(n-k_1-k_2-\ldots-k_r)=0$ verwendet wurde.

Beispiel 6.13

Auf wie viele Arten kann man 36 Spielkarten an 4 Spieler verteilen, wenn jeder Spieler gleich viele Karten bekommen soll?

Lösung mit Multinomialkoeffizient:

$$\frac{36!}{9!\cdot 9!\cdot 9!\cdot 9!}$$

Lösung mit mehreren Binomialkoeffizienten:

$$\binom{36}{9}\cdot\binom{27}{9}\cdot\binom{18}{9}\cdot\binom{9}{9}=\frac{36!}{9!\cdot 27!}\cdot\frac{27!}{9!\cdot 18!}\cdot\frac{18!}{9!\cdot 9!}\cdot\frac{9!}{9!\cdot 0!}$$
$$=\frac{36!}{9!\cdot 9!\cdot 9!\cdot 9!}\quad\text{(OK)}$$

Beispiel 6.14

Wie viele verschiedene „Wörter“ lassen sich unter Verwendung aller Buchstaben des Wortes „TATOO“ bilden?

Lösung 1: Wären alle Zeichen verschieden ($\mathrm{T_1AT_2O_1O_2}$), gäbe es 5! Möglichkeiten.

Da sich $\mathrm{T_1}$ und $\mathrm{T_2}$ sowie $\mathrm{O_1}$ und $\mathrm{O_2}$ jeweils auf 2! Arten (unerkannt) vertauschen lassen, müssen wir das obige Resultat entsprechend korrigieren:

$$\frac{5!}{2!\cdot 2!}\quad\text{oder „vollständig“}\quad\frac{5!}{2!\cdot 2!\cdot 1!}$$

Lösung 2: Um die unsere ursprüngliche kombinatorische Deutung des Multinomialkoeffizienten zu verwenden, müssen wir das in der Aufgabe gestellte Problem damit identifizieren.

Die Zeichenpositionen $\{1, 2, 3, 4, 5\}$ entsprechen den verschiedenen Objekten.

Für jedes unterschiedliche Zeichen denken wir uns eine Box mit der Häufigkeit des Zeichens als Fassungsvermögen:

$\overset{\text{T}}{\square\square}\quad \overset{\text{A}}{\square}\quad \overset{\text{O}}{\square\square}$

Jede Verteilung der Positionen auf die Boxen entspricht genau einem „Wort“, wobei die Reihenfolge in einer Box irrelevant ist.

$\boxed{12}\ \boxed{3}\ \boxed{45} \leftrightarrow$ TTAOO

$\boxed{24}\ \boxed{5}\ \boxed{13} \leftrightarrow$ OTOTA

...

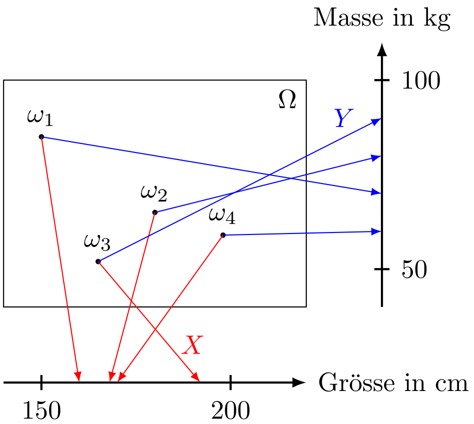
Daher erhalten wir auch bei dieser Betrachtungsweise dasselbe Resultat:

$$\frac{5!}{2! \cdot 2! \cdot 1!} = 30$$

7 Diskrete Zufallsvariablen

7.1 Einleitung

Motivation



Zufallsvariable

Eine *Zufallsvariable* X ist eine Funktion, die jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet. Der Begriff *Zufallsgrösse* wird synonym verwendet.

Eine Zufallsvariable ist keine Variable sondern eine Funktion!

Diskrete und stetige Zufallsvariable

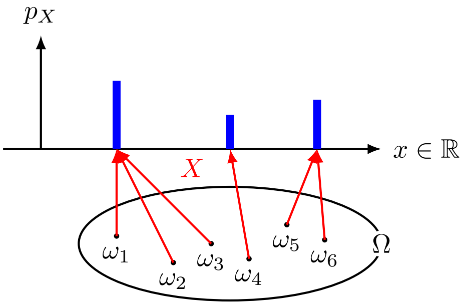
- Ist der Wertebereich von X abzählbar, so wird die Zufallsvariable *diskret* genannt.
- Ist der Wertebereich von X überabzählbar, so wird die Zufallsvariable *stetig* genannt.

Beachte

- Die Zufallsvariable X ist eine *Funktion*, die jedem Ergebnis des Stichprobenraums eine reelle Zahl zuordnet und wird mit einem *Grossbuchstaben* bezeichnet.

- Der Wert einer Zufallsvariable $x = X(\omega)$ für ein bestimmtes Element $\omega \in \Omega$ ist eine *Zahl* und wird mit einem *Kleinbuchstaben* bezeichnet.

Die Wahrscheinlichkeitsfunktion einer Zufallsvariablen



Wahrscheinlichkeitsfunktion (Probability Mass Function, PMF):

$p_X(x) = P\big(\{\omega \in \Omega\colon X(\omega) = x\}\big) = P\big(X = x\big)$

Ist klar, um welche Zufallsvariable es geht, schreibt man $p(x)$ statt $p_X(x)$.

Berechnung von $p(x)$: Für jeden möglichen Wert x von X ...

- bestimme alle Ergebnisse $\omega \in \Omega$ mit $X(\omega) = x$,
- addiere alle ihre Wahrscheinlichkeiten, um $p(x)$ zu erhalten.
- Falls es kein $\omega \in \Omega$ mit $X(\omega) = x$ gibt, setze $p(x) = 0$.

Eigenschaften der Wahrscheinlichkeitsfunktion $p(x)$:

- $p(x) \geq 0$ für alle $x \in \mathbb{R}$
- $\sum\limits_{x \in \mathbb{R}} p(x) = 1$

Beispiel 7.1

Wurf mit zwei Spielwürfeln; X: Summe der Augenzahlen

- $P(X = 5) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = 4 \cdot \frac{1}{36} = \frac{1}{9}$
- $P(X < 4) = P(\{(1,1),(1,2),(2,1)\}) = 3 \cdot \frac{1}{36} = \frac{1}{12}$
- $P(X = 7.5) = P(\{\,\}) = 0$

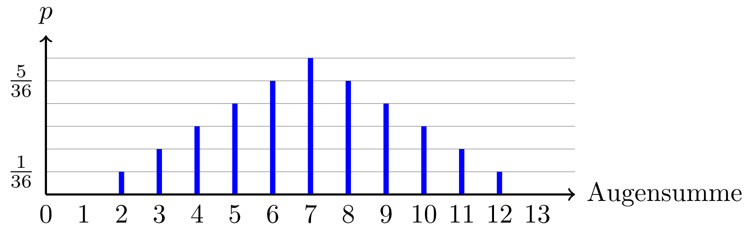
Die Verteilungsfunktion einer Zufallsvariable

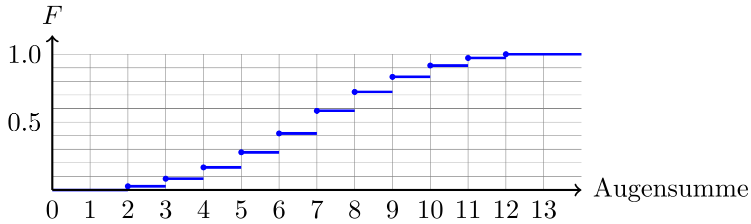
Eine weitere Möglichkeit, die Verteilung der Wahrscheinlichkeiten einer Zufallsvariable darzustellen, ist ihre *Verteilungsfunktion*:

$F_X(x) = P\big(\{\omega \in \Omega\colon X(\omega) \leq x\}\big) = P\big(X \leq x\big)$

Beispiel 7.2

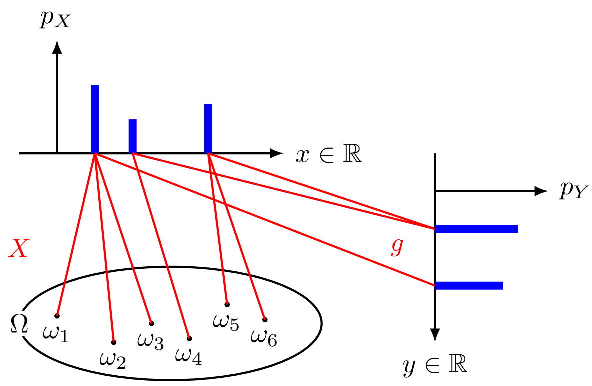
Wurf mit zwei Spielwürfeln; X: Summe der Augenzahlen





Funktionen von Zufallsvariablen

Ist X eine Zufallsvariable auf Ω und $g\colon \mathbb{R} \to \mathbb{R}$ eine rellewertige Funktion, so definiert die Verkettung $Y = g(X)$ eine weitere Zufallsvariable Y auf Ω.



Beispiel 7.3

x	-1	1	3	5
$p_X(x)$	0.4	0.3	0.1	0.2

$Y = g(X) = X^2 - 4X + 1$

x	-1	1	3	5
$y = g(x)$	6	-2	-2	6

y	-2	6
$p_Y(y)$	0.4	0.6

7.2 Erwartungswert und Varianz einer Zufallsvariable

Der Erwartungswert

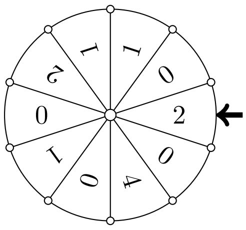
Der *Erwartungswert* $E(X)$ einer Zufallsvariable X mit der Wahrscheinlichkeitsfunktion p wird wie folgt definiert:

$$E(X) = \sum_x x\, p(x)$$

Der Erwartungswert ist das mit den Wahrscheinlichkeiten gewichtete Mittel aller Werte der Zufallsvariable. Statt $E(X)$ schreibt man oft auch kürzer μ.

Beispiel 7.4

Ein faires Glücksrad wird einmal bewegt. X: aufgedruckte Zahl



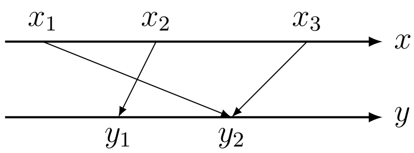
x	0	1	2	4
$p(x)$	0.4	0.3	0.2	0.1

$E(X) = 0 \cdot 0.4 + 1 \cdot 0.3 + 2 \cdot 0.2 + 4 \cdot 0.1 = 1.1$

Erwartungswert für Funktionen von Zufallsvariablen

Gegeben: Zufallsvariable X und Funktion $f\colon \mathbb{R} \to \mathbb{R}$

Gesucht: $E(Y)$ mit $Y = g(X)$



$$E(g(X)) = E(Y) = \sum_y y p_Y(y) = y_1 p_Y(y_1) + y_2 p_Y(y_2)$$

$$= y_1 p_X(x_2) + y_2 \big[p_X(x_1) + p_X(x_3)\big]$$

$$= y_1 p_X(x_2) + y_2 p_X(x_1) + y_2 p_X(x_3)$$

$$= g(x_2) p_X(x_2) + g(x_1) p_X(x_1) + g(x_3) p_X(x_3) = \sum_x g(x) p_X(x)$$

Beispiel 7.5

Berechne aus der Zufallsgrösse X mit

x	-1	1	3	5
$p_X(x)$	0.4	0.3	0.1	0.2

und $Y = g(X) = X^2 - 4X + 1$ den Erwartungswert $E(Y)$.

$$
\begin{aligned}
E(Y) = E(g(X)) &= \sum_x g(x)\, p_X(x) \\
&= g(-1) \cdot 0.4 + g(1) \cdot 0.3 + g(3) \cdot 0.1 + g(5) \cdot 0.2 \\
&= 6 \cdot 0.4 - 2 \cdot 0.3 - 2 \cdot 0.1 + 6 \cdot 0.2 = 2.8
\end{aligned}
$$

cf. Beispiel 7.3

Die schlechte Nachricht

Im Allgmeinen ist

$$\sum_x g(x) p_X(x) \neq f\left(\sum_x x p_X(x)\right)$$

was gleichbedeutend mit

$$E(g(X)) \neq g(E(X))$$

ist.

Die gute Nachricht

Für Funktionen der Form $g(X) = aX + b$ (affine Funktionen) gilt:

$$E(aX + b) = aE(X) + b$$

Beweis:

$$
\begin{aligned}
E(aX + b) &= \sum_x (ax + b) p_X(x) \\
&= a \sum_x x p_X(x) + b \sum_x p_X(x) = aE(X) + b
\end{aligned}
$$

Die Varianz einer Zufallsvariable

Ist X eine Zufallsvariable, so ist nach dem oben Gesagten auch

$$g(X) = \left[X - E(X)\right]^2$$

eine Zufallsvariable. Dieser Term stellt den quadratischen Unterschied der Werte von X von ihrem Erwartungswert („Mittel“) dar. Bestimmt man von dieser neuen Zufallsvariable den Erwartungswert, so erhält man die *Varianz* von X:

$$\mathrm{Var}(X) = E\left(\left[X - E(X)\right]^2\right) = \sum_x \left[x - E(X)\right]^2 p_X(x)$$

Die Varianz ist ein Mass für die Streuung der Werte von X um ihren Erwartungswert.

Die Standardabweichung

Das Quadrat in der Definition der Varianz ist nötig, damit sich die positiven und negativen Abweichungen der Werte der Zufallsvariable vom Erwartungswert nicht gegenseitig auslöschen. Dies führt dazu, dass die Masseinheit der Varianz das Quadrat der Masseinheit von X ist.

Möchte man ein mit der ursprünglichen Einheit verträgliches Mass für die Streuung einer Zufallsvariable, so zieht man die Quadratwurzel aus der Varianz und nennt diese Grösse *Standardabweichung* von X:

$$\sigma_X(X) = \sqrt{\operatorname{Var}(X)}$$

Beispiel 7.10

Berechne Varianz und Standardabweichung der folgenden Zufallsvariablen:

x	-2	0	3	5
$p_X(X)$	0.2	0.4	0.3	0.1

$E(X) = -2 \cdot 0.2 + 0 \cdot 0.4 + 3 \cdot 0.3 + 5 \cdot 0.1 = 1$

$$\begin{aligned}\operatorname{Var}(X) &= E\left(\left[X - E(X)\right]^2\right)\\ &= (-2-1)^2 \cdot 0.2 + (0-1)^2 \cdot 0.4 + 4 \cdot 0.3 + 16 \cdot 0.1\\ &= 5\end{aligned}$$

$\sigma(X) = \sqrt{5}$

Rechenregeln für die Varianz (1)

Auch wenn

$$g(X) = \left[X - E(X)\right]^2$$

in der Berechnung von

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

nicht linear ist, kann die Summendarstellung von f zur einfacheren Berechnung von $\operatorname{Var}(X)$ verwendet werden:

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

Rechenregeln für die Varianz (2)

Für eine affine Transformation $Y = aX + b$ der Zufallsvariablen X gilt:

$$
\begin{aligned}
&\operatorname{Var}(aX + b) \\
&= E\big((aX+b)^2\big) - \big(E(aX+b)\big)^2 \\
&= E\big(a^2X^2 + 2abX + b^2\big) - \big(aE(X) + b\big)^2 \\
&= a^2E(X^2) + 2abE(X) + b^2 - \big(a^2E(X)^2 + 2abE(X) + b^2\big) \\
&= a^2E(X^2) - a^2E(X)^2 \\
&= a^2\big(E(X^2) - E(X)^2\big) \\
&= a^2 \operatorname{Var}(X)
\end{aligned}
$$

7.3 Häufig verwendete Verteilungen

Bernoulli-Verteilung

Ein *Bernoulli-Experiment* ist ein Experiment mit genau zwei Ausgängen, d. h. $\Omega = \{\omega_1, \omega_2\}$ und $P(\omega_1) = p$ und $P(\omega_2) = 1 - p$.

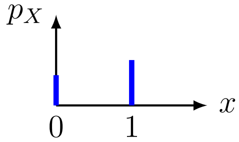
Ordnet man den Ereignissen wie folgt die Werte 1 und 0 zu:

$$
\begin{aligned}
X(\omega_1) &= 1 \quad \text{„Erfolg“} \\
X(\omega_2) &= 0 \quad \text{„Misserfolg“}
\end{aligned}
$$

erhält man die Wahrscheinlichkeitsfunktion

$$
\begin{aligned}
p_X(1) &= P(X = 1) = p \\
p_X(0) &= P(X = 0) = 1 - p
\end{aligned}
$$

die *Bernoulli-(Wahrscheinlichkeits)Verteilung* genannt wird.



Erwartungswert und Varianz einer bernoulliverteilten Zufallsvariable:

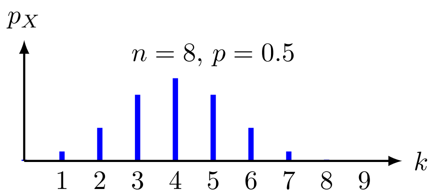
$$E(X) = \sum_x x \cdot p_X(x) = 1 \cdot p + 0 \cdot (1-p) = p$$

$$
\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - E(X)^2 \\
&= \underbrace{\left[1^2 \cdot p + 0^2 \cdot (1-p)\right]}_{E(X^2)} - p^2 = p - p^2 = p(1-p)
\end{aligned}
$$

Die Binomialverteilung

k: Anzahl der Erfolge bei n unabhängigen Wiederholungen eines Bernoulli-Experiments mit der Erfolgswahrscheinlichkeit p.

$$P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{für } k = 0, 1, 2, \dots, n \\ 0 & \text{sonst} \end{cases}$$



Erwartungswert und Varianz:

Um bereits berechnete Resultate wiederverwenden zu können, stellen wir X als Summe bernoulliverteilter Zufallsvariablen X_i mit der Erfolgswahrscheinlickeit p dar, wobei die X_i entweder den Wert 1 (Erfolg) oder 0 (Misserfolg) annehmen:

$$X = X_1 + X_2 + \ldots + X_n$$

Dann gilt wegen der Linearität des Erwartungswerts:

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

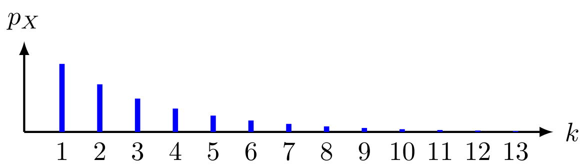
und aufgrund der Unabhängigkeit von X_1, X_2, …, X_n

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(X_1 + X_2 + \cdots + X_n) \\ &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = np(1-p) \end{aligned}$$

Die geometrische Verteilung

Ein Bernoulli-Experiment wird so lange wiederholt, bis der erste Erfolg eintritt. (X: Nummer des Versuchs mit dem ersten Erfolg)

$$P(X = k) = \begin{cases} p(1-p)^{k-1} & \text{für } k = 1, 2, 3, \dots \\ 0 & \text{sonst} \end{cases}$$



$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \cdot \frac{1}{1 - (1-p)} = 1$$

Erwartungswert: Um

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

zu berechnen, differenziere die für $|x| < 1$ gültige Identität

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad \Rightarrow \quad \frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1}$$

Setze $x = 1-p$ und lasse den verschwindenen Summanden weg:

$$\frac{1}{p^2} = \sum_{k=1}^{\infty} k(1-p)^{k-1} \quad \Rightarrow \quad \frac{1}{p} = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = E(X)$$

Varianz: Um

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1}$$

zu bestimmen, differenziere die für $|x| < 1$ gültige Identität:

$$\frac{x}{(1-x)^2} = \sum_{k=0}^{\infty} kx^k \Rightarrow \frac{1+x}{(1-x)^3} = \sum_{k=0}^{\infty} k^2 x^{k-1}$$

Setze $x = 1-p$ und lasse den verschwindenden Summanden weg:

$$\frac{2-p}{p^3} = \sum_{k=1}^{\infty} k^2(1-p)^{k-1} \Rightarrow \frac{2-p}{p^2} = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = E(X^2)$$

Insgesamt: $\mathrm{Var}(X) = E(X^2) - E(X)^2 = \dfrac{2-p}{p^2} - \dfrac{1}{p^2} = \dfrac{1-p}{p^2}$

Die hypergeometrische Verteilung

In einer Grundgesamtheit vom Umfange m haben r Elemente eine bestimmte Eigenschaft. Aus dieser Menge werden ohne Zurücklegen n Elemente gezogen. k ist die Anzahl der Elemente in der Stichprobe, welche die oben erwähnte bestimmte Eigenschaft haben.

$$P(X = k) = \frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}}$$

mit $n < m$, $r < m$ und $\max(0, n-m+r) \leq k \leq \min(n, r)$

Erwartungswert:

$$E(X) = n \cdot \frac{r}{m}$$

Varianz:

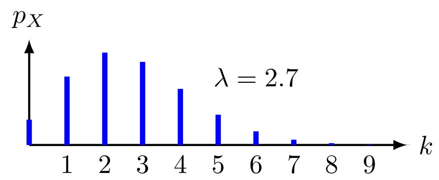
$$\mathrm{Var}(X) = n \cdot \frac{r}{m} \cdot \left(1 - \frac{r}{m}\right) \cdot \frac{m-n}{m-1}$$

[ohne Beweis]

Die Poisson-Verteilung

k: Anzahl (unabhängiger) Ereignisse, die während eines festen Zeitintervalls beobachtet werden können, wenn im Mittel λ Ereignisse pro Zeitintervall eintreten.

$$P(X=k) = \begin{cases} \mathrm{e}^{-\lambda}\dfrac{\lambda^k}{k!} & \text{für } k = 0,1,2,3,\ldots \\ 0 & \text{sonst} \end{cases}$$



$$\sum_{k=0}^{\infty} p_X(k) = \mathrm{e}^{-\lambda}\left[\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots\right] = \mathrm{e}^{-\lambda}\cdot\mathrm{e}^{\lambda} = 1$$

$$E(X) = \sum_{k=0}^{\infty} k\,\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} = \lambda\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda\mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \overset{k-1\to k}{=} \lambda\mathrm{e}^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2\,\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!}$$

$$= \mathrm{e}^{-\lambda}\left(\sum_{k=1}^{\infty}(k-1)\frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\right)$$

$$= \mathrm{e}^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} + \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}$$

$$= \mathrm{e}^{-\lambda}\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \mathrm{e}^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

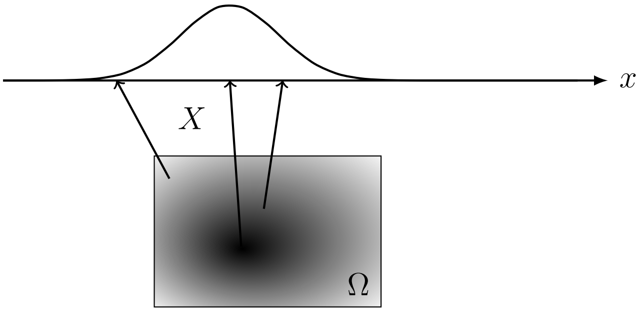
$$= \mathrm{e}^{-\lambda}\lambda^2\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} + \mathrm{e}^{-\lambda}\lambda\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\lambda^2\mathrm{e}^{\lambda} + \mathrm{e}^{-\lambda}\lambda\mathrm{e}^{\lambda}$$

$$= \lambda^2 + \lambda$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

8 Stetige Zufallsvariablen

Versuch einer Visualisierung



Es ist unmöglich, jedem $x \in \mathbb{R}$ eine Wahrscheinlichkeit $P(\{X = x\})$ zuzuordnen, so dass gilt:

$$\sum_{x \in \mathbb{R}} p_X(X) = \sum_{x \in \mathbb{R}} P(X = x) = 1$$

Dichtefunktionen

Eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ heisst *Dichtefunktion*, wenn sie folgende Eigenschaften hat.

- Für alle $x \in \mathbb{R}$ gilt $f(x) \geq 0$ (*Nicht-Negativität*)
- $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ (*Normiertheit*)

Eine stetige Zufallsvariable X hat die Dichtefunktion f_X, wenn für alle $a < b$ gilt:

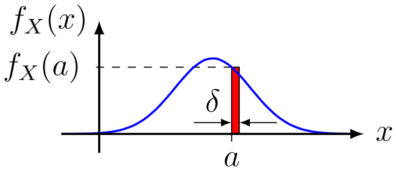
$$P(a \leq X \leq b) = \int_a^b f_X(x)\,\mathrm{d}x$$

Eigenschaften von Wahrscheinlichkeitsdichtefunktionen

- Im Gegensatz zu diskreten Zufallsvariablen kann eine Wahrscheinlichkeitsdichte f_X auch Werte annehmen, die grösser als 1 sind.

- Für ein Intervall $[a, a + \delta]$ von sehr kleiner Länge δ gilt:

$$P(a \leq X \leq a + \delta) = \int_a^{a+\delta} f_X(x)\,\mathrm{dx} \approx f_X(a) \cdot \delta$$



Erwartungswert und Varianz

Für eine stetige Zufallsvariable X mit Dichte f_X definiert man analog zum diskreten Fall:

- $E(X) = \displaystyle\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$
- $\operatorname{Var}(X) = \displaystyle\int_{-\infty}^{\infty} \big(x - E(X)\big)^2 f_X(x)\,\mathrm{d}x$

Rechenregeln

Analog zum diskreten Fall lassen sich folgende Regeln herleiten:

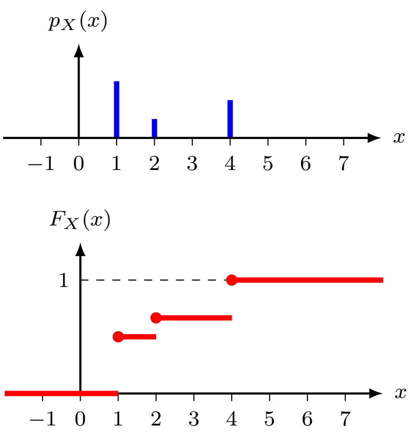
- $E(g(X)) = \displaystyle\int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$
- $E(aX + b) = aE(X) + b$
- $\operatorname{Var}(X) = E(X^2) - E(X)^2$
- $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$

Die kumulative Verteilungsfunktion

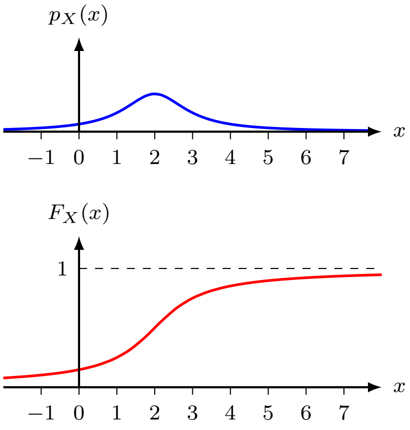
Sowohl für diskrete wie auch für stetige Zufallsvariablen X lässt sich die *(kumulative) Verteilungsfunktion* (*cumulative distribution function, CDF*) $F_X\colon \mathbb{R} \to \mathbb{R}$ definieren:

- $F_X(x) = P(X \leq x) = \displaystyle\sum_{t \leq x} p_X(t)$
- $F_X(x) = P(X \leq x) = \displaystyle\int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$

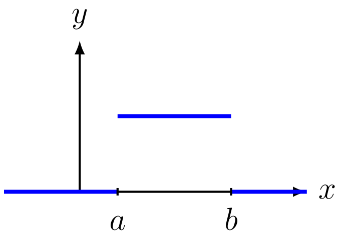
Verteilungsfunktion für eine diskrete Zufallsvariable



Verteilungsfunktion für eine stetige Zufallsvariable

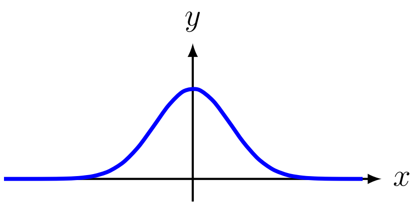


Beispiel 8.1: uniforme stetige Verteilung



- $f_X(x) = 1/(b-a)$ für alle $x \in (a, b)$.
- $E(X) = (a + b)/2$ (Symmetrie)
- $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$ (→ Übungen)

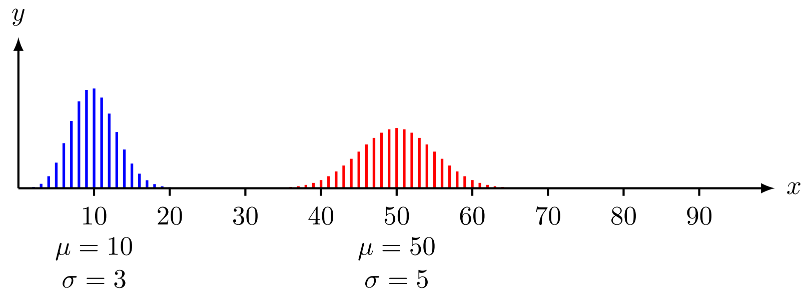
Beispiel 8.2: Standardnormalverteilung



- $f_X(x) = \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$
- $E(X) = 0$ (Symmetrie)
- $\mathrm{Var}(X) = 1$ (→ Übungen)
- $\varphi_{0;1}(x) = f_X(x)$

Die Näherungsformel von De Moivre-Laplace

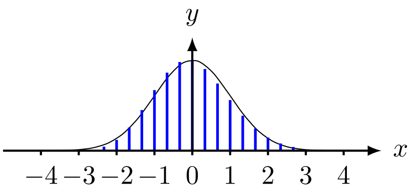
Stabdiagramme für $B(x, 100, 0.1)$ und $B(x, 100, 0.5)$:



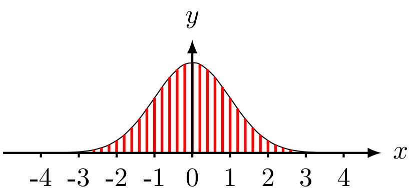
Normalisieren der Stabdiagramme: ($\mu = np$, $\sigma = \sqrt{np(1-p)}$)

- $x \to \dfrac{x-\mu}{\sigma}$
- $y \to \sigma\, y$

normalisiertes Stabdiagramm für $B(x, 100, 0.1)$:



normalisiertes Stabdiagramm für $B(x, 100, 0.5)$:



Für eine binominalverteilte Zufallsvariable X mit den Parametern p und n gilt bei grossen Werten von n:

$$P(x_1 \le X \le x_2) \approx \int_{z_1}^{z_2} \varphi_{0;1}(z)\,\mathrm{d}z$$

mit $z_1 = \dfrac{x_1 - np}{\sqrt{np(1-p)}}$ und $z_2 = \dfrac{x_2 - np}{\sqrt{np(1-p)}}$

Diese Näherung liefert für $np(1-q) > 9$ „brauchbare“ Werte.

Anstatt die Binomialverteilung an die Standardnormalverteilung anzupassen, kann man die Standardnormalverteilung an die Binomialverteilung anpassen ($\mu = np$, $\sigma = \sqrt{np(1-p)}$):

$$P(x_1 \le X \le x_2) \approx \int_{x_1}^{x_2} \varphi_{\mu;\sigma}(x)\,\mathrm{d}x$$

wobei $\varphi_{\mu;\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$

9 Hypothesentests

9.1 Einleitung

Motivation

Am Anfang steht eine Theorie über eine bestimmte Eigenschaft einer Population (Grundgesamtheit).

- Ein neues Medikament soll die Dauer des Heilungsprozesses einer Krankheit verkürzen
- Klassische Musik beeinflusst die Intelligenz von Kindern.

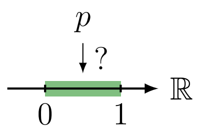
Wie kann nun eine solche Theorie durch stichprobenartig erhobene Daten (Umfragen, Experimente) bestätigt oder verworfen werden?

Beispiel 9.1

Ist eine gegebene Münze fair?

Sei p die Wahrscheinlichkeit für „Zahl".

Wo liegt der Parameter p?

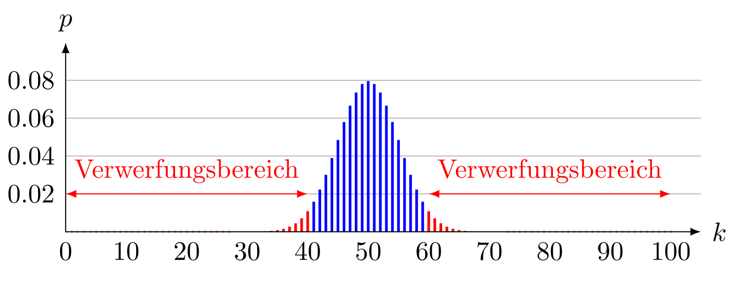


Nullhypothese (H_0): $p = 0.5$

Alternativhypothese (H_1): $p \neq 0.5$ ungerichtet (zweiseitig)
$p < 0.5$ gerichtet (linksseitig)
$p > 0.5$ gerichtet (rechtsseitig)

Die Wahl der Alternativhypothese hängt vom Vorwissen über das Objekt ab. Im Zweifelsfall sollte man zweiseitig testen.

Teststatistik: Wahrscheinlichkeitsverteilung von „Anzahl Zahl" (k) bei $n = 100$ (unabhängigen) Münzwürfen



Je mehr das Stichprobenresultat vom Zentrum der Verteilung abweicht, umso mehr sind wir bereit H_0 zu verwerfen und H_1 (vorläufig) anzunehmen.

Die Wahrscheinlichkeit, dass die Teststatistik einer Stichprobe im Verwerfungsbereich liegt, wird als *Fehler 1. Art* oder *Signifikanzniveau (α)* genannt. Sollten wir nämlich trotz wahrer Nullhypothese ein äusserst „extremes" Resultat in Richtung von H_1 erhalten, so beziffert α gerade die Wahrscheinlichkeit, mit der wir H_0 fälschlicherweise verwerfen würden. Insgesamt:

	H_0 beibehalten	H_0 verwerfen
H_0 gilt	richtig entschieden	Fehler 1. Art (α)
H_0 gilt nicht	Fehler 2. Art (β)	richtig entschieden

Wenn H_0 falsch ist, wissen wir nicht, welches p von H_1 gilt. Daher ist es ohne zusätzliche Annahmen über H_1 auch nicht möglich, den β-Fehler zu bestimmen. Klar ist nur, dass wir mit der Verkleinerung des α-Fehlers die Wahrscheinlichkeit für einen β-Fehler erhöhen würden (und umgekehrt).

Die Alternativhypothese

Laut Fremdwörterduden ist eine Hypothese eine

„zunächst unbewiesene Annahme von Gesetzlichkeiten oder Tatsachen, mit dem Ziel, sie durch Beweise zu verifzieren oder zu falsifizieren.“

Hypothesen, die im Widerspruch zu anderen Theorien stehen oder oder diese ergänzen werden *Alternativhypothesen* genannt (Bortz & Schuster, S. 97).

Gibt die Alternativhypothese die Richtung eines Unterschieds vor, spricht man von einer *gerichteten Hypothese* (den Heilungsprozess *verkürzen*); andernfalls von einer *ungerichteten Hypothese* (die Intelligenz *beeinflussen*).

Die Alternativhypothese wird in der Statistik mit H_1 bezeichnet.

Die Nullhypothese

Die *Nullhypothese* ist die Gegenspielerin der Alternativhypothese. Sie besagt, dass der in der Alternativhypothese postulierte (geforderte) Unterschied nicht vorhanden sei.

Die Nullhypothese bildet die statistische Grundlage eines Hypothesentests, von der aus (indirekt) die Gültigkeit der Alternativhypothese beurteilt wird.

Die Nullhypothese wird in der Statistik mit H_0 bezeichnet.

Der Hypothesentest

Wie soll zwischen den beiden konkurrierenden Hypothesen entschieden werden, wenn nur eine Zufallsstichprobe aus der Grundgesamtheit verfügbar ist?

Üblicherweise geht man so vor, dass untersucht wird, ob die aus der Stichprobe gewonnenen Daten mit der Nullhypothese *verträglich* sind. Sind sie es nicht so wird die Nullhypothese zugunsten der Alternativhypothese *verworfen.* Andernfalls wird die Nullhypothese *beibehalten.*

Das Konzept könnte man etwas salopp auch so zusammenfassen: *Das Bisherige gilt; das Neue muss sich bewähren.*

Um diese Entscheidung zu treffen, wird aus den empirischen Daten eine *Prüfgrösse* (Synonym: *Teststatistik*) berechnet. Beispiele dafür sind:

- z-Wert
- t-Wert
- χ^2-Wert
- Rangsumme
- ...

Diese Prüfgrösse hängt vom verwendeten Test ab, der wiederum auf den Voraussetzungen basiert, unter denen die Daten zustande gekommen sind:

- Zufallsstichprobe
- Normalverteilung
- „kleiner“ Stichprobenumfang
- Stichproben mit gleicher Varianz
- ...

Auf der Grundlage der Nullhypothese und den Voraussetzungen des jeweiligen Tests lässt sich die Verteilung der Prüfgrösse rechnerisch (oder durch Simulation) bestimmen.

Die Bestimmung von Testverteilungen ist eine im Allgemeinen schwierige Arbeit. Sie wurde von den Statistikern in den letzten Jahrzehnten bereits geleistet und das Resultat steht uns heute in Form von Tabellen oder Computerprogrammen zur Verfügung.

Damit wird es möglich, den aus der Stichprobe berechneten Wert der Prüfgrösse mit der theoretischen (erwarteten) Verteilung der Prüfgrösse in Beziehung zu setzen:

- Hat der berechnete Wert der Prüfgrösse eine „grosse“ Wahrscheinlichkeit, so ist die Stichprobe mit der Nullhypothese vereinbar.
- Hat der berechnete Wert der Prüfgrösse eine „kleine“ Wahrscheinlichkeit, so ist die Stichprobe nicht mit der Nullhypothese vereinbar.

Das Signifikanzniveau

Es muss also ein Grenze bzw. ein kritischer Wert festgelegt werden, bei der die Entscheidung von „noch mit H_0 vereinbar“ zu „nicht mehr mit H_0 vereinbar“ wechselt.

Diese Grenze wird *Signifikanzniveau* genannt und mit α bezeichnet. Übliche Werte für α sind 0.05 und 0.01.

Der p-Wert

Unter der Voraussetzung der Nullhypothese ist der p-Wert gleich der Wahrscheinlichkeit, dass die Prüfgrösse den beobachteten oder einen extremeren Wert in Richtung der Alternativhypothese annimmt.

- Ein tiefer p-Wert besagt, dass es unwahrscheinlich ist, dass ein solches Ergebnis auf der Basis der Nullhypothese zustande kommt. Das spricht gegen H_0 und damit für H_1.
- Ein hoher p-Wert besagt, dass es wahrscheinlich ist, dass ein solches Ergebnis auf der Basis der Nullhypothese zustande kommt. Das spricht für H_0 und damit gegen H_1.

Dieser p-Wert wird von Statistikprogrammen ausgegeben und muss dann mit dem Signifikanzniveau verglichen werden, um eine Entscheidung zu fällen.

Der Begriff p-Wert ist etwas unglücklich gewählt, da es Verfahren gibt, in denen eine Wahrscheinlichkeit (Proportion) getestet wird, die ebenfalls mit p bezeichnet wird.

Zusammenfassung

Bortz und Schuster definieren einen *statistischen Test* wie folgt:

> „Ein statistischer Test ist eine Regel, die es erlaubt, für jedes Stichprobenergebnis eine Entscheidung zwischen der Null- und der Alternativhypothese zu treffen.“

Ablauf eines Hypothesentests

1. Die Hypothesen (H_0 und H_1) aufstellen und die Richtung der Alternativhypothese vorgeben (im Zweifelsfall zweiseitig).
2. Das Signifikanzniveau α festlegen.
3. Das Testverfahren aufgrund der Voraussetzungen wählen.
4. Die Prüfgrösse berechnen:
   - *computergestützt:* Den p-Wert mit α vergleichen
   - *klassisch:* Den Wert der Prüfgrösse mit den kritischen Werten aus Tabellen zu gegebenem α vergleichen
5. Die Entscheidung fällen:
   - falls p-Wert $< \alpha$: H_0 verwerfen
   - falls p-Wert $\geq \alpha$: H_0 beibehalten

Fehlerarten

Bei einem Hypothesentest ist nicht auszuschliessen, dass die durch den Test gefällte Entscheidung falsch ist, zumal sie in der Regel auf einer einzigen Stichprobe beruht.

Durch die zwei Entscheidungsmöglichkeiten sind folgende Situationen denkbar

	H_0 beibehalten	H_0 verwerfen
H_0 gilt	richtig entschieden	Fehler 1. Art
H_0 gilt nicht	Fehler 2. Art	richtig entschieden

Der Fehler 1. Art

Bei der oben beschriebenen Art des Testens kann die der Fehler 1. Art gerade durch das Signifikanzniveau quantifiziert werden. Warum?

Stellt man bei einem Signifikanzniveau von $\alpha = 0.05$ fest, dass die Wahrscheinlichkeit für die vorliegende (oder eine noch weiter von H_0 entfernte) Stichprobe den p-Wert 0.04 hat, so entscheidet man sich *gegen* H_0. Sollte H_0 dennoch wahr sein, so irrt man sich höchstens in 5 von 100 Fällen.

Der Fehler 2. Art

Die Grösse des Fehlers 2. Art kann ohne zusätzliche Annahmen nicht wie der Fehler 1. Art abgeschätzt werden.

Es sollte aber klar sein, dass man durch Verkleinern von α zwar den Fehler 1. Art „vermindert"; dadurch aber die Hürde zum Akzeptieren von H_1 höher ansetzt und damit die Möglichkeit eines Fehlers 2. Art vergrössert.

9.2 Der Einstichproben-t-Test

Die Idee

Der Einstichproben-t-Test überprüft, ob sich der Mittelwert μ der Grundgesamtheit, aus der die Stichprobe gezogen wird, von einem vorgegebenen Sollwert μ_0 unterscheidet.

	H_0	H_1
zweiseitig:	$\mu = \mu_0$	$\mu \neq \mu_0$
rechtsseitig:	$\mu = \mu_0$	$\mu > \mu_0$
linksseitig:	$\mu = \mu_0$	$\mu < \mu_0$

Voraussetzungen

- Die Werte der Stichprobe müssen durch eine Zufallsstichprobe ermittelt werden.
- Das Merkmal ist in der Grundgesamtheit normalverteilt.

Der Umfang n der Stichprobe wird vom t-Test nicht vorgegeben. Allerdings sind grössere Werte von n wünschenswert.

Die Prüfgrösse

$$t = \sqrt{n} \cdot \frac{\overline{x} - \mu_0}{s}$$

mit $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ und $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}$

Beispiel

In einem Fliessgewässer wurde vor 5 Jahren eine mittlere Nitratkonzentration von 1.7 mg/l gemessen.

Eine vergleichbare aktuelle Messreihe, die in Abständen von 14 Tagen durchgeführt wurde, ergab folgende Werte in mg/l:

1.6, 1.9, 1.5, 1.8, 2.1, 1.8, 2.0, 1.6

Untersuche, ob sich die Nitratkonzentration gegenüber dem Wert vor 5 Jahren signifikant erhöht hat ($\alpha = 0.05$).

Lösung

μ: aktuelle mittlere (unbekannte) Nitratkonzentration

$\mu_0 = 1.7\,\text{mg/l}$

H_0: $\mu = \mu_0$

H_1: $\mu > \mu_0$ (rechtsseitiger Test)

$n = 8$, $\overline{x} = 1.7875$, $s = 0.2100$

t-Wert: $\sqrt{8} \cdot \frac{1.7875 - 1.7}{0.2100} = 1.1784$

p-Wert $= 0.1386$

Entscheidung: H_0 beibehalten

Der Wert der Nitratkonzentration hat sich nicht signifikant erhöht.

9.3 Der Binomialtest ($\approx$ z-Test für eine Proportion)

Die Testidee

Der Binomialtest für eine Proportion überprüft ob die Häufigkeitsverteilung einer dichotomen (binären) Zufallsvariablen einer vorgegebenen Verteilung entspricht.

Oder etwas konkreter ausgedrückt: Entspricht die beobachtete relative Auftretenshäufigkeit p der theoretisch erwarteten Auftretenswahrscheinlichkeit p_0?

	H_0	H_1
zweiseitig:	$p = p_0$	$p \neq p_0$
rechtsseitig:	$p = p_0$	$p > p_0$
linksseitig:	$p = p_0$	$p < p_0$

Voraussetzungen

- Die Zufallsvariable ist dichotom (zweiwertig).
- Die Werte der Stichprobe müssen durch eine Zufallsstichprobe ermittelt werden.

Wenn die verwendete Software die „exakte“ Binomialverteilung durch die Normalverteilung approximiert, sollte die entsprechende Bedingung $n \cdot p_0(1-p_0) > 9$ erfüllt sein. Dies muss auch bei der Verwendung des TI-84 Plus getan werden.

Die Prüfgrösse

n: Umfang der Stichprobe

x: Anzahl der Erfolge in der Stichprobe

linksseitig: $P(X \leq x) = \sum_{i=0}^{x} p_0^i (1-p_0)^{n-i}$

zweiseitig: $P(X \leq x) + P(X \geq n-x) = 1 - \sum_{i=x+1}^{n-x-1} p_0^i (1-p_0)^{n-i}$

rechtsseitig: $P(X \geq x) = \sum_{i=x}^{n} p_0^i (1-p_0)^{n-i}$

Beispiel

Bei einem Multiple-Choice-Test gibt es zu jeder der 24 Fragen vier Antwortmöglichkeiten, von denen jeweils genau eine richtig ist.

Ein Schüler löste bei diesem Test 11 der 24 Fragen richtig. Untersuche mit einem Test für eine Proportion, ob der Schüler gänzlich unvorbereitet an den Test gekommen ist ($\alpha = 0.05$).

Lösung

$p = P$(eine Frage richtig zu beantworten) (unbekannt)

$p_0 = 0.25$ Wahrscheinlichkeit, die Antwort richtig zu raten

$n = 24$, $x = 11$

Gilt $np_0(1-p_0) > 9$? $\quad 24 \cdot 0.25 \cdot 0.75 = 4.5 \leq 9$

Nein; aber wir führen den Test trotzdem durch!

H_0: $p = p_0$

H_1: $p > p_0$ (rechtsseitiger Test)

z-Wert: 2.1939

p-Wert $= 0.014$

Entscheidung: H_0 verwerfen

Die Wahrscheinlichkeit, eine Frage richtig zu beantworten, ist signifikant höher als beim Raten. (Mehr aber auch nicht!)

Bemerkung: bei der Verwendung der exakten binomischen Teststatistik beträgt der p-Wert 0.02134, was immer noch signifikant wäre.

9.4 bDer z-Test für zwei Proportionen

Die Idee

Dieser Test überprüft, ob zwei dichotome Zufallsvariablen dieselbe Verteilung haben. Dazu wird die Approximation durch die Normalverteilung verwendet.

	H_0	H_1
zweiseitig:	$p_1 = p_2$	$p_1 \neq p_0$
rechtsseitig:	$p_1 = p_2$	$p_1 > p_0$
linksseitig:	$p_1 = p_2$	$p_1 < p_0$

Voraussetzungen

- Beide Zufallsvariablen sind dichotom (zweiwertig).
- Die Werte der Stichprobe müssen durch eine Zufallsstichprobe ermittelt werden.

Die Prüfgrösse

Berechne

- $d = \dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}$
- $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$
- $\mathrm{Var}_{\hat{p}}(d) = \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right) \cdot \hat{p}(1 - \hat{p})$

und damit $z = \dfrac{d}{\sqrt{\mathrm{Var}_{\hat{p}}(d)}}$

Beispiel

Im Jahr 2016 konnte Roger Federer auf eine 18-jährige Karriere zurückblicken und war dabei während 302 Wochen die Nummer 1 in der Weltrangliste des Herrentennis.

Die Karriere von Pete Sampras dauerte 14 Jahre. Während dieser Zeit war er während 286 Wochen die Nummer 1.

Gibt es einen signifikanten Unterschied zwischen den beiden Spielern in Bezug auf den Anteil der Nummer 1-Position während ihrer jeweiligen Karriere?

Lösung

Federer: $n_1 = 302$; $n_2 = 18 \cdot 52 = 936$

Sampras: $n_2 = 286$; $n_2 = 14 \cdot 52 = 728$

H_0: $p_1 = p_2$

H_1: $p_1 \neq p_2$ (zweiseitiger Test)

$z = -2.972$

p-Wert $= 0.002957$

Entscheidung: H_0 verwerfen

$\hat{p}_1 = 0.3226$ Anteil der Wochen als Nr. 1 von Federer

$\hat{p}_2 = 0.3928$ Anteil der Wochen als Nr. 1 von Sampras

Bis zum Jahr 2016 hatte Pete Sampras, bezogen auf seine Karrieredauer, einen signifikant höheren Nummer 1-Wochenanteil.

9.5 Der Chiquadrat-Anpassungstest

Die Testidee

Der Pearson Chi-Quadrat-Test wird angewendet, um zu prüfen, ob sich eine empirisch beobachtete Verteilung einer kategorialen Variable von einer bestimmten theoretisch erwarteten Verteilung unterscheidet. Die erwartete Verteilung kann dabei beliebig sein.

Dieser Test wird auch als „Chi-Quadrat-Anpassungstest“ (engl. „Goodness of fit test“), „Chi-Quadrat-Homogenitätstest“ oder „Einstichproben-Chi-Quadrat-Test“ bezeichnet.

Voraussetzungen

- Die Variable ist kategorial
- Die erwartete Häufigkeit in jeder Kategorie muss mindestens 1 betragen.
- Bei höchstens 20% der Kategorien darf die erwartete Häufigkeit unter 5 liegen.

Die Prüfgrösse

$b_1, b_2, \ldots, b_n$: die beobachteten Häufigkeiten

$e_1, e_2, \ldots, e_n$: die erwarteten Häufigkeiten

$$\chi^2 = \sum_{i=1}^{n} \frac{(b_i - e_i)^2}{e_i}$$

df = Anzahl Freiheitsgrade (degrees of freedom)

Für mehr als zwei Kategorien sind links- und rechtsseitige Tests nicht mehr sinnvoll.

Beispiel

Kreuzt man Pflanzen, die sich in zwei chromosomal unabhängigen Merkmalen reinerbig unterscheiden, in einem dominant-rezessiven Erbgang, so erhält man in der zweiten Filialgeneration vier Phänotypen im Verhältnis 9 : 3 : 3 : 1 (3. Mendelsche Regel).

Überprüfe mit dem χ^2-Anpassungstest, ob die in Mendels Züchtungsversuch experimentell bestimmten Häufigkeiten

	B	b
A	315	108
a	101	32

mit den von ihm postulierten Verhältnissen übereinstimmen ($\alpha = 0.05$).

Lösung

	AB	Ab	aB	ab
beobachtet	315	108	101	32
erwartet	312.75	104.25	104.25	34.75

$\chi^2 = 0.4700$

df $= 3$

p-Wert $= 0.9254$

Entscheidung: H_0 beibehalten

Die experimentell bestimmten Zahlen unterscheiden sich nicht signifikant von den postulierten Verhältnissen.