

SS 2015

Diskrete Wahrscheinlichkeitstheorie

Susanne Albers

Fakultät für Informatik
TU München

<http://www14.in.tum.de/lehre/2015SS/dwt/>

Sommersemester 2015

Kapitel 0 Organisatorisches

- Vorlesungen:
 - Di 13:00–15:00 und Di 15:00–16:00 (MI HS1)
Pflichtvorlesung Bachelor IN, Bioinformatik
Modulnr.: IN0018
- Übung:
 - 2SWS Tutorübung: siehe Webseite zur Übung
 - Übungsleitung: Dennis Kraft, Richard Stotz
- Umfang:
 - 3V+2TÜ, 6 ECTS-Punkte
- Sprechstunde:
 - nach Vereinbarung

- Übungsaufgaben:
 - Ausgabe jeweils am Montag auf der Webseite der Vorlesung, ab 8:00 Uhr
 - Abgabe eine Woche später, jeweils Montag bis 12:00 Uhr, Briefkästen Westseite Untergeschoss FMI Magistrale
 - Vorbereitung in der Tutorübung
 - vorauss. 12 Übungsblätter, das letzte am 13. Juli 2015, jedes 20 Punkte
 - **Bonusregelung:** Werden bei den **ersten sechs und zweiten sechs Übungsblättern** jeweils mindestens 50% der insgesamt erreichbaren Punkte erzielt, so verbessert sich die Note einer *bestanden* Klausur um 1/3 Notenstufe.
- Klausur:
 - Klausur am **04. August 2015, 11–13 Uhr**
 - Wiederholungsklausur am **30. September 2015, 11–13 Uhr**
 - bei den Klausuren sind *keine* Hilfsmittel außer einem handbeschriebenen DIN-A4-Blatt zugelassen





- Vorkenntnisse:
 - Einführung in die Informatik I/II
 - Diskrete Strukturen
- Weiterführende Vorlesungen:
 - Effiziente Algorithmen und Datenstrukturen
 - Randomisierte Algorithmen
 - Online- und Approximationsalgorithmen
 - Komplexitätstheorie
 - ...
- Webseite:

<http://www14.in.tum.de/lehre/2015SS/dwt/>

1. Vorlesungsinhalt

- Diskrete Wahrscheinlichkeitsräume
 - Wahrscheinlichkeitsraum, Ereignis, Zufallsvariable
 - spezielle Verteilungen
 - Ungleichungen von Markov und Chebyshev
- Kontinuierliche Wahrscheinlichkeitsräume
 - Normalverteilung, Exponentialverteilung
 - Zentraler Grenzwertsatz
- Statistik
 - Schätzvariablen
 - Konfidenzintervalle
 - Testen von Hypothesen
- Stochastische Prozesse
 - Markovketten
 - Warteschlangen

2. Literatur

-  T. Schickinger, A. Steger:
Diskrete Strukturen - Band 2,
Springer Verlag, 2001
-  M. Greiner, G. Tinhofer:
Stochastik für Informatiker,
Carl Hanser Verlag, 1996
-  H. Gordon:
Discrete Probability,
Springer-Verlag, 1997
-  M. Mitzenmacher, E. Upfal:
Probability and Computing: Randomized Algorithms and Probabilistic Analysis,
Cambridge University Press, 2005

-  R. Motwani, P. Raghavan:
Randomized Algorithms,
Cambridge University Press, 1995
-  M. Hofri:
Probabilistic Analysis of Algorithms,
Springer Verlag, 1987
-  L. Fahrmeir, R. Künstler, I. Pigeot, G. Tutz:
Statistik - Der Weg zur Datenanalyse,
Springer-Verlag, 1997

3. Einleitung

Was bedeutet Zufall?

- Unkenntnis über den Ausgang eines durchgeführten Experiments
- Ein Experiment wird vielfach mit eventuell sich änderndem Ergebnis ausgeführt
- Ereignisse stehen in keinem kausalen Zusammenhang
- physikalischer Zufall (Rauschen, Kernzerfall)

Zufall in der *diskreten* Informatik

- Die Eingabe für einen bestimmten Algorithmus wird aus einer großen Menge möglicher Eingaben **zufällig** gewählt:

average case

- Kombination von Worst-Case- und Average-Case-Analyse, in der Eingaben gemäß einer Verteilung leicht pertubiert werden:

smoothed analysis

- Der Algorithmus verwendet Zufallsbits, um mit großer Wahrscheinlichkeit gewisse **Problemsituationen** zu vermeiden:

Randomisierung

Kapitel I Diskrete Wahrscheinlichkeitsräume

1. Grundlagen

Definition 1

- ① Ein **diskreter Wahrscheinlichkeitsraum** ist durch eine **Ergebnismenge** $\Omega = \{\omega_1, \omega_2, \dots\}$ von **Elementarereignissen** gegeben.
- ② Jedem Elementarereignis ω_i ist eine **(Elementar-)Wahrscheinlichkeit** $\Pr[\omega_i]$ zugeordnet, wobei wir fordern, dass $0 \leq \Pr[\omega_i] \leq 1$ und

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1.$$

- ③ Eine Menge $E \subseteq \Omega$ heißt Ereignis. Die Wahrscheinlichkeit $\Pr[E]$ eines Ereignisses ist durch

$$\Pr[E] := \sum_{\omega \in E} \Pr[\omega]$$

definiert.

Beispiel 2

Zwei faire Würfel (einer weiß, einer schwarz) werden geworfen. Wir sind an der Gesamtzahl der angezeigten Augen interessiert:

$$\Omega = \{ \begin{array}{l} (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), \\ (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \\ (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), \\ (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), \\ (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), \\ (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6) \end{array} \}$$

- ① Die Wahrscheinlichkeit $\Pr((i, j))$ eines jeden Elementarereignisses (i, j) ist $\frac{1}{36}$.
- ② Die Wahrscheinlichkeit $\Pr(E)$ des Ereignisses

$$E = \{\text{Die Gesamtzahl der Augen ist } 10\}$$

ist $\frac{1}{12}$.

Wir hätten aber auch sagen können:

$$\Omega = \{2, 3, 4, \dots, 10, 11, 12\}$$

Die Wahrscheinlichkeiten der Elementarereignisse sind dann aber nicht mehr gleich. Es ist z.B.

- ❶ $\Pr(2) = \frac{1}{36};$
- ❷ $\Pr(4) = \frac{1}{12};$
- ❸ $\Pr(7) = \frac{1}{6}.$

Beispiel 3

Eine faire Münze wird so lange geworfen, bis die gleiche Seite zweimal hintereinander fällt. Dann ist

$$\Omega = \{hh, tt, htt, thh, thtt, hthh, hthtt, ththh, \dots\}$$

Frage: Was sind die Wahrscheinlichkeiten der einzelnen Elementarereignisse?

\bar{E} heißt **komplementäres Ereignis** zu E .

Allgemein verwenden wir bei der Definition von Ereignissen alle bekannten Operatoren aus der Mengenlehre. Wenn also A und B Ereignisse sind, dann sind auch $A \cup B$, $A \cap B$, $A \setminus B$ etc. Ereignisse.

Zwei Ereignisse A und B heißen **disjunkt** oder auch **unvereinbar**, wenn $A \cap B = \emptyset$ gilt.

Definition 4

$$\begin{aligned}\text{relative H\u00e4ufigkeit von } E &:= \frac{\text{absolute H\u00e4ufigkeit von } E}{\text{Anzahl aller Beobachtungen}} \\ &= \frac{\text{Anzahl Eintreten von } E}{\text{Anzahl aller Beobachtungen}}.\end{aligned}$$

Definition 5

Ein Wahrscheinlichkeitsraum mit $\Omega = \{\omega_1, \dots, \omega_n\}$ heißt **endlicher Wahrscheinlichkeitsraum**.

Bei unendlichen Wahrscheinlichkeitsräumen werden wir gewöhnlich nur den Fall $\Omega = \mathbb{N}_0$ betrachten. Dies stellt keine große Einschränkung dar, da wir statt einer Ergebnismenge $\Omega = \{\omega_1, \omega_2, \dots\}$ auch \mathbb{N}_0 als Ergebnismenge verwenden können, indem wir ω_i mit $i - 1$ identifizieren. Wir sagen, dass durch die Angabe der Elementarwahrscheinlichkeiten ein **Wahrscheinlichkeitsraum auf Ω** definiert ist.

Beispiel 6

Wir beobachten die an einer Straße in Bayern vorbeifahrenden Autos. Dabei gelte:

- ① Es fahren doppelt so viele Autos von links nach rechts wie von rechts nach links.
- ② Von zehn Autos haben zwei die Farbe hellelfenbein, die übrigen eine andere Lackierung.

- Das Ereignis “*Wir beobachten ein von links nach rechts fahrendes Auto*” hat die Wahrscheinlichkeit $\frac{2}{3}$.
- Das Ereignis “*Das nächste Auto ist ein Taxi von rechts*” passiert mit Wahrscheinlichkeit

$$\frac{1}{3} \cdot \frac{1}{5}.$$

Beispiel 7 (Unendlicher Wahrscheinlichkeitsraum)

Wir betrachten eine Münze, die mit Wahrscheinlichkeit p Kopf zeigt und mit Wahrscheinlichkeit $q := 1 - p$ Zahl.

Wir führen Versuche aus, indem wir die Münze wiederholt solange werfen, bis *Zahl* fällt.

Das *Ergebnis* eines solchen Versuchs ist die Anzahl der durchgeführten Münzwürfe.

Damit ergibt sich hier als Ergebnismenge

$$\Omega = \mathbb{N} = \{1, 2, 3, \dots\} .$$

Beispiel 7 (Forts.)

Sei, für $i \in \mathbb{N}$, ω_i das Elementarereignis

$\omega_i \hat{=}$ Die Münze wird i -mal geworfen .

Dann gilt:

$$\Pr[\omega_i] = p^{i-1}q ,$$

und

$$\sum_{\omega \in \Omega} \Pr[\omega] = \sum_{i=1}^{\infty} p^{i-1}q = q \cdot \sum_{i=0}^{\infty} p^i = \frac{q}{1-p} = 1 .$$

(wie es sein soll!)

Lemma 8

Für Ereignisse A, B, A_1, A_2, \dots gilt:

- ① $\Pr[\emptyset] = 0, \Pr[\Omega] = 1.$
- ② $0 \leq \Pr[A] \leq 1.$
- ③ $\Pr[\bar{A}] = 1 - \Pr[A].$
- ④ Wenn $A \subseteq B$, so folgt $\Pr[A] \leq \Pr[B].$

Lemma 8 (Forts.)

- ⑤ (Additionssatz) Wenn die Ereignisse A_1, \dots, A_n paarweise disjunkt sind (also wenn für alle Paare $i \neq j$ gilt, dass $A_i \cap A_j = \emptyset$), so folgt

$$\Pr \left[\bigcup_{i=1}^n A_i \right] = \sum_{i=1}^n \Pr[A_i].$$

Für disjunkte Ereignisse A, B erhalten wir insbesondere

$$\Pr[A \cup B] = \Pr[A] + \Pr[B].$$

Für eine unendliche Menge von disjunkten Ereignissen A_1, A_2, \dots gilt analog

$$\Pr \left[\bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} \Pr[A_i].$$

Beweis:

Die Aussagen folgen unmittelbar aus Definition 1, den Eigenschaften der Addition und der Definition der Summe. □

Eigenschaft 5 in Lemma 8 gilt nur für disjunkte Ereignisse. Für den allgemeinen Fall erhalten wir folgenden

Satz 9 (Siebformel, Prinzip der Inklusion/Exklusion)

Für Ereignisse A_1, \dots, A_n ($n \geq 2$) gilt:

$$\begin{aligned} \Pr \left[\bigcup_{i=1}^n A_i \right] &= \sum_{i=1}^n \Pr[A_i] - \sum_{1 \leq i_1 < i_2 \leq n} \Pr[A_{i_1} \cap A_{i_2}] + \dots \\ &+ (-1)^{l-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} \Pr[A_{i_1} \cap \dots \cap A_{i_l}] + \dots \\ &+ (-1)^{n-1} \cdot \Pr[A_1 \cap \dots \cap A_n] . \end{aligned}$$

Satz 9 (Forts.)

Insbesondere gilt für zwei Ereignisse A und B

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] .$$

Für drei Ereignisse A_1 , A_2 und A_3 erhalten wir

$$\begin{aligned} \Pr[A_1 \cup A_2 \cup A_3] &= \Pr[A_1] + \Pr[A_2] + \Pr[A_3] \\ &\quad - \Pr[A_1 \cap A_2] - \Pr[A_1 \cap A_3] \\ &\quad - \Pr[A_2 \cap A_3] \\ &\quad + \Pr[A_1 \cap A_2 \cap A_3] . \end{aligned}$$

Beweis:

Wir betrachten zunächst den Fall $n = 2$. Dazu setzen wir $C := A \setminus B = A \setminus (A \cap B)$. Gemäß dieser Definition gilt, dass C und $A \cap B$ sowie C und B disjunkt sind. Deshalb können wir Eigenschaft 5 von Lemma 8 anwenden:

$$\Pr[A] = \Pr[C \cup (A \cap B)] = \Pr[C] + \Pr[A \cap B] .$$

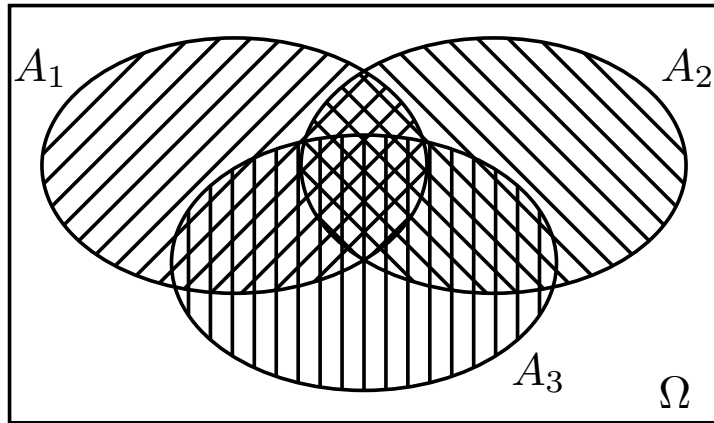
Wegen $A \cup B = C \cup B$ folgt daraus

$$\begin{aligned} \Pr[A \cup B] &= \Pr[C \cup B] = \Pr[C] + \Pr[B] = \\ &\Pr[A] - \Pr[A \cap B] + \Pr[B] \end{aligned}$$

und wir haben die Behauptung für $n = 2$ gezeigt.

Beweis (Forts.):

Der Fall $n = 3$:



Man beachte, dass durch die im Satz angegebene Summe jedes Flächenstück insgesamt genau einmal gezählt wird.

Beweis (Forts.):

Der allgemeine Fall kann nun durch Induktion über n gezeigt werden (was wir aber hier nicht ausführen!).

Satz 9 findet man manchmal auch unter der Bezeichnung *Satz von Poincaré-Sylvester*, nach dem Franzosen

Jules Henri Poincaré (1854–1912)

und dem Engländer

James Joseph Sylvester (1814–1897)

benannt.

Boolesche Ungleichung:

Die folgende Abschätzung ist nach **George Boole** (1815–1864) benannt:

Korollar 10

Für Ereignisse A_1, \dots, A_n gilt

$$\Pr \left[\bigcup_{i=1}^n A_i \right] \leq \sum_{i=1}^n \Pr[A_i] .$$

Analog gilt für eine unendliche Folge von Ereignissen A_1, A_2, \dots , dass

$$\Pr \left[\bigcup_{i=1}^{\infty} A_i \right] \leq \sum_{i=1}^{\infty} \Pr[A_i] .$$

Beweis:

Zunächst betrachten wir die linke Seite der Ungleichung für den endlichen Fall und erhalten

$$\Pr \left[\bigcup_{i=1}^n A_i \right] = \sum_{\omega \in \bigcup_{i=1}^n A_i} \Pr[\omega] .$$

Für die rechte Seite gilt

$$\sum_{i=1}^n \Pr[A_i] = \sum_{i=1}^n \sum_{\omega \in A_i} \Pr[\omega] .$$

Jedes Elementarereignis kommt links also genau einmal und rechts mindestens einmal vor. □

1.1 Wahl der Wahrscheinlichkeiten

Frage: Wie können Wahrscheinlichkeiten sinnvoll festgelegt werden?

Prinzip von Laplace (Pierre Simon Laplace (1749–1827)): *Wenn nichts dagegen spricht, gehen wir davon aus, dass alle Elementarereignisse gleich wahrscheinlich sind.*
Also:

$$\Pr[E] = \frac{|E|}{|\Omega|}$$

1.2 Historische Anfänge der Wahrscheinlichkeitstheorie

Die ersten Hinweise auf mathematische Untersuchungen zu Problemen der Wahrscheinlichkeitstheorie finden sich in einem Briefwechsel zwischen den französischen Mathematikern

Pierre Fermat (1601–1665)

und

Blaise Pascal (1623–1662).

Pascal beschäftigte sich neben der Mathematik auch mit Fragestellungen aus dem Bereich der Physik und auch aus der Informatik! Sein Vater hatte als Steuerinspektor in Rouen umfangreiche Rechnungen durchzuführen und so wurde Pascal zum Bau einer mechanischen Rechenmaschine, der so genannten *Pascaline*, motiviert.

In dem Briefwechsel taucht bereits der Ansatz $\Pr[E] = |E|/|\Omega|$ zur Berechnung der Wahrscheinlichkeit von E auf. Auch den Begriff des Erwartungswerts kann man dort schon finden. Weder Fermat noch Pascal publizierten ihre Überlegungen zur Wahrscheinlichkeitstheorie. Der Niederländer

Christiaan Huygens (1629–1695)

entwickelte ebenfalls Methoden zum Arbeiten mit Wahrscheinlichkeiten aus. Er publizierte im Jahre 1657 auch eine kleine Arbeit mit dem Titel „De ratiociniis in ludo aleae“ (Über die Gesetzmäßigkeiten beim Würfelspiel).

2. Bedingte Wahrscheinlichkeiten

Beispiel 11

A und B spielen Poker (52 Karten, 5 Karten pro Spieler, keine getauschten Karten). A hält vier Asse und eine Herz Zwei in der Hand. B kann dieses Blatt nur überbieten, wenn er einen Straight Flush (fünf Karten *einer* Farbe in aufsteigender Reihenfolge hat. Die Wahrscheinlichkeit für das Ereignis $F :=$ „B hat einen Straight Flush“ beträgt

$$\Pr[F] = \frac{|F|}{|\Omega|} = \frac{3 \cdot 8 + 7}{\binom{52-5}{5}} = \frac{31}{1533939} = 2,02.. \cdot 10^{-5}.$$

Beispiel 11 (Forts.)

A hat die Karten allerdings gezinkt und weiß, dass B nur Kreuz in der Hand hält. Bezeichne nun Ω' den Wahrscheinlichkeitsraum aller Möglichkeiten für B und F' das Ereignis, dass B einen Straight Flush der Farbe Kreuz hat:

$$\Pr[F'] = \frac{|F'|}{|\Omega'|} = \frac{8}{\binom{12}{5}} = \frac{8}{792} \approx 0,01 !!$$

Für $\Pr[A|B]$ erforderliche Eigenschaften:

- 1 $\Pr[B|B] = 1$;
- 2 $\Pr[A|\Omega] = \Pr[A]$;
- 3 für festes B ist $\Pr[A|B]$ proportional zu $\Pr[A \cap B]$.

Definition 12

A und B seien Ereignisse mit $\Pr[B] > 0$. Die **bedingte Wahrscheinlichkeit** $\Pr[A|B]$ von A gegeben B ist definiert als

$$\Pr[A|B] := \frac{\Pr[A \cap B]}{\Pr[B]} .$$

Die bedingten Wahrscheinlichkeiten $\Pr[\cdot|B]$ bilden für ein beliebiges Ereignis $B \subseteq \Omega$ mit $\Pr[B] > 0$ einen neuen Wahrscheinlichkeitsraum über Ω .

Es ist leicht nachzurechnen, dass dadurch die Definition eines *diskreten Wahrscheinlichkeitsraums* erfüllt ist:

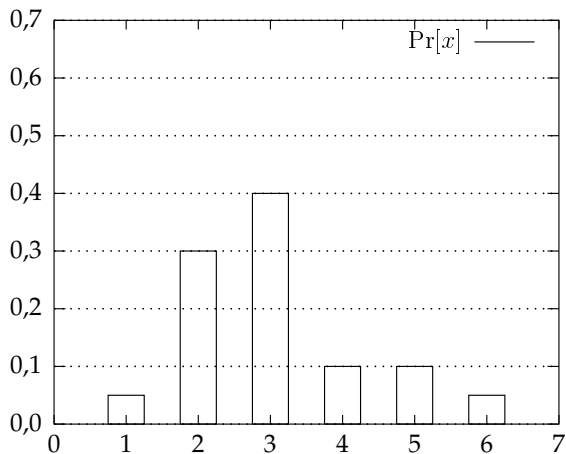
$$\sum_{\omega \in \Omega} \Pr[\omega|B] = \sum_{\omega \in \Omega} \frac{\Pr[\omega \cap B]}{\Pr[B]} = \sum_{\omega \in B} \frac{\Pr[\omega]}{\Pr[B]} = \frac{\Pr[B]}{\Pr[B]} = 1.$$

Damit gelten alle Rechenregeln für Wahrscheinlichkeiten auch für bedingte Wahrscheinlichkeiten. Beispielsweise:

$$\Pr[\emptyset|B] = 0 \text{ sowie } \Pr[\bar{A}|B] = 1 - \Pr[A|B].$$

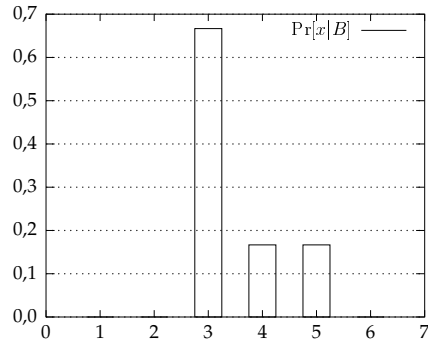
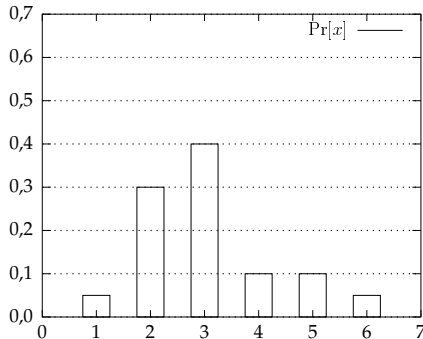
Beispiel 13 (Reskalierung bei bedingten Wahrscheinlichkeiten)

Betrachte folgenden **gezinkten** Würfel:



Beispiel 13 (Forts.)

Wir betrachten nun den durch $B := \{3, 4, 5\}$ gegebenen bedingten Wahrscheinlichkeitsraum:



Was genau war die Bedingung?

Beispiel 14 (Zweikinderproblem)

Wir nehmen an, dass bei der Geburt eines Kindes beide Geschlechter gleich wahrscheinlich sind. Wir wissen, dass eine bestimmte Familie zwei Kinder hat und eines davon ein Mädchen ist. Wie groß ist die Wahrscheinlichkeit, dass beide Kinder der Familie Mädchen sind?

Natürlich $\frac{1}{2}$.

Wirklich?

Beispiel 14 (Forts.)

Eigentlich gilt:

$$\Omega := \{mm, mj, jm, jj\}$$

und

$$M := \{mm, mj, jm\}.$$

Wir bedingen auf M , und damit gilt für $A := \{mm\}$:

$$\Pr[A|M] = \frac{\Pr[A \cap M]}{\Pr[M]} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Beispiel 15 (Ziegenproblem)

Sie nehmen an einer Spielshow im Fernsehen teil, bei der Sie eine von drei verschlossenen Türen auswählen sollen. Hinter einer Tür wartet der Preis, ein Auto, hinter den beiden anderen stehen Ziegen. Sie zeigen auf eine Tür, sagen wir Nummer eins. Sie bleibt vorerst geschlossen. Der Moderator weiß, hinter welcher Tür sich das Auto befindet; mit den Worten “Ich gebe Ihnen mal einen kleinen Hinweis” öffnet er eine andere Tür, zum Beispiel Nummer drei, und eine Ziege schaut heraus und meckert. Er fragt: “Bleiben Sie bei Nummer eins, oder wählen sie Nummer zwei? ”

Frage: Welche Strategie ist günstiger:

- S1 Der Spieler bleibt immer bei seiner ursprünglichen Wahl.
- S2 Der Spieler wechselt stets die ausgewählte Tür.

Beispiel (Forts.)

Wir betrachten hier eine Diskussion des Ziegenproblems mit Hilfe von bedingten Wahrscheinlichkeiten. Wir betrachten bei jeder Variante den Fall, dass der Spieler

- a) die “richtige”,*
- b) eine falsche Tür gewählt hat.*

Ersteres geschieht mit Wahrscheinlichkeit $\frac{1}{3}$, Letzteres mit Wahrscheinlichkeit $\frac{2}{3}$.

Mit der vom Moderator gegebenen Information ergeben sich für die beiden Strategien die folgenden Gewinnwahrscheinlichkeiten:

	S1	S2
a)	?	?
b)	?	?

Häufig verwendet man die Definition der **bedingten Wahrscheinlichkeit** in der Form

$$\Pr[A \cap B] = \Pr[B|A] \cdot \Pr[A] = \Pr[A|B] \cdot \Pr[B] . \quad (1)$$

Damit:

Satz 16 (Multiplikationssatz)

Seien die Ereignisse A_1, \dots, A_n gegeben. Falls $\Pr[A_1 \cap \dots \cap A_n] > 0$ ist, gilt

$$\begin{aligned} \Pr[A_1 \cap \dots \cap A_n] = \\ \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \Pr[A_3|A_1 \cap A_2] \cdot \dots \\ \dots \cdot \Pr[A_n|A_1 \cap \dots \cap A_{n-1}] . \end{aligned}$$

Beweis:

Zunächst halten wir fest, dass alle bedingten Wahrscheinlichkeiten wohldefiniert sind, da $\Pr[A_1] \geq \Pr[A_1 \cap A_2] \geq \dots \geq \Pr[A_1 \cap \dots \cap A_n] > 0$.

Die rechte Seite der Aussage im Satz können wir umschreiben zu

$$\frac{\Pr[A_1]}{1} \cdot \frac{\Pr[A_1 \cap A_2]}{\Pr[A_1]} \cdot \frac{\Pr[A_1 \cap A_2 \cap A_3]}{\Pr[A_1 \cap A_2]} \cdot \dots \cdot \frac{\Pr[A_1 \cap \dots \cap A_n]}{\Pr[A_1 \cap \dots \cap A_{n-1}]}.$$

Offensichtlich kürzen sich alle Terme bis auf $\Pr[A_1 \cap \dots \cap A_n]$.



Beispiel 17 (Geburtstagsproblem)

Wie groß ist die Wahrscheinlichkeit, dass in einer m -köpfigen Gruppe zwei Personen am selben Tag Geburtstag haben?

Umformulierung:

Man werfe m Bälle zufällig und gleich wahrscheinlich in n Körbe. Wie groß ist die Wahrscheinlichkeit, dass nach dem Experiment jeder Ball allein in seinem Korb liegt?

Für das Geburtstagsproblem: $n = 365$

Offensichtlich muss $m \leq n$ sein, damit überhaupt jeder Ball allein in einem Korb liegen kann.

Wir nehmen an, dass die Bälle nacheinander geworfen werden. A_i bezeichne das Ereignis „Ball i landet in einem noch leeren Korb“. Das gesuchte Ereignis „Alle Bälle liegen allein in einem Korb“ bezeichnen wir mit A . Nach Satz 16 können wir $\Pr[A]$ berechnen durch

$$\begin{aligned}\Pr[A] &= \Pr[\cap_{i=1}^m A_i] \\ &= \Pr[A_1] \cdot \Pr[A_2|A_1] \cdot \dots \cdot \Pr[A_m | \cap_{i=1}^{m-1} A_i].\end{aligned}$$

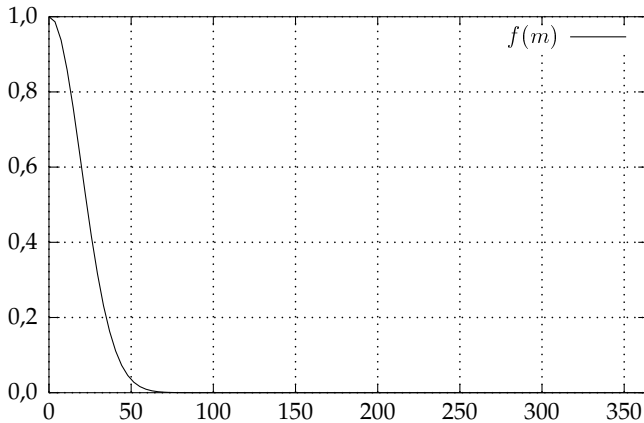
Unter der Bedingung, dass die ersten $j - 1$ Bälle jeweils in einem leeren Korb gelandet sind, bedeutet A_j , dass der j -te Ball in eine der $n - (j - 1)$ leeren Körbe fallen muss, die aus Symmetriegründen jeweils mit derselben Wahrscheinlichkeit gewählt werden.

Daraus folgt

$$\Pr[A_j | \cap_{i=1}^{j-1} A_i] = \frac{n - (j - 1)}{n} = 1 - \frac{j - 1}{n}.$$

Mit der Abschätzung $1 - x \leq e^{-x}$ und wegen $\Pr[A_1] = 1$ erhalten wir

$$\begin{aligned} \Pr[A] &= \prod_{j=1}^m \left(1 - \frac{j-1}{n}\right) \\ &\leq \prod_{j=2}^m e^{-(j-1)/n} = e^{-(1/n) \cdot \sum_{j=1}^{m-1} j} \\ &= e^{-m(m-1)/(2n)} =: f(m). \end{aligned}$$



Verlauf von $f(m)$ für $n = 365$

Ausgehend von der Darstellung der bedingten Wahrscheinlichkeit in Gleichung 1 zeigen wir:

Satz 18 (Satz von der totalen Wahrscheinlichkeit)

Die Ereignisse A_1, \dots, A_n seien paarweise disjunkt und es gelte $B \subseteq A_1 \cup \dots \cup A_n$. Dann folgt

$$\Pr[B] = \sum_{i=1}^n \Pr[B|A_i] \cdot \Pr[A_i] .$$

Analog gilt für paarweise disjunkte Ereignisse A_1, A_2, \dots mit $B \subseteq \bigcup_{i=1}^{\infty} A_i$, dass

$$\Pr[B] = \sum_{i=1}^{\infty} \Pr[B|A_i] \cdot \Pr[A_i] .$$

Beweis:

Wir zeigen zunächst den endlichen Fall. Wir halten fest, dass

$$B = (B \cap A_1) \cup \dots \cup (B \cap A_n) .$$

Da für beliebige i, j mit $i \neq j$ gilt, dass $A_i \cap A_j = \emptyset$, sind auch die Ereignisse $B \cap A_i$ und $B \cap A_j$ disjunkt. Wegen (1) folgt $\Pr[B \cap A_i] = \Pr[B|A_i] \cdot \Pr[A_i]$ (auch für den Fall, dass $\Pr[A_i] = 0$!). Wir wenden nun den Additionssatz (Lemma 8, Teil 5) an

$$\begin{aligned} \Pr[B] &= \Pr[B \cap A_1] + \dots + \Pr[B \cap A_n] = \\ &\Pr[B|A_1] \cdot \Pr[A_1] + \dots + \Pr[B|A_n] \cdot \Pr[A_n] \end{aligned}$$

und haben damit die Behauptung gezeigt. Da der Additionssatz auch für unendlich viele Ereignisse A_1, A_2, \dots gilt, kann dieser Beweis direkt auf den unendlichen Fall übertragen werden. □

Mit Hilfe von Satz 18 erhalten wir leicht einen weiteren nützlichen Satz:

Satz 19 (Satz von Bayes)

Die Ereignisse A_1, \dots, A_n seien paarweise disjunkt, mit $\Pr[A_j] > 0$ für alle j . Ferner sei $B \subseteq A_1 \cup \dots \cup A_n$ ein Ereignis mit $\Pr[B] > 0$. Dann gilt für ein beliebiges $i = 1, \dots, n$

$$\Pr[A_i|B] = \frac{\Pr[A_i \cap B]}{\Pr[B]} = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{j=1}^n \Pr[B|A_j] \cdot \Pr[A_j]} .$$

Analog gilt für paarweise disjunkte Ereignisse A_1, A_2, \dots mit $B \subseteq \bigcup_{i=1}^{\infty} A_i$, dass

$$\Pr[A_i|B] = \frac{\Pr[A_i \cap B]}{\Pr[B]} = \frac{\Pr[B|A_i] \cdot \Pr[A_i]}{\sum_{j=1}^{\infty} \Pr[B|A_j] \cdot \Pr[A_j]} .$$

Mit dem Satz von Bayes dreht man gewissermaßen die Reihenfolge der Bedingung um. Gegeben die Wahrscheinlichkeit von B unter den Bedingungen A_i (sowie die Wahrscheinlichkeiten der A_i selbst), berechnet man die Wahrscheinlichkeit von A_i bedingt auf das Ereignis B .

Thomas Bayes (1702–1761) war ein bekannter Theologe und Mitglied der Royal Society. Als sein bedeutendstes Werk gilt sein Beitrag zur Wahrscheinlichkeitstheorie „Essay Towards Solving a Problem in the Doctrine of Chances“. Diese Arbeit wurde erst 1763 publiziert.

3. Unabhängigkeit

Bei einer bedingten Wahrscheinlichkeit $\Pr[A|B]$ kann der Fall auftreten, dass die Bedingung auf B , also das Vorwissen, dass B eintritt, keinen Einfluss auf die Wahrscheinlichkeit hat, mit der wir das Eintreten von A erwarten. Es gilt also $\Pr[A|B] = \Pr[A]$, und wir nennen dann die Ereignisse A und B **unabhängig**.

Beispiel 20 (Zweimaliges Würfeln)

$$\Omega := \{(i, j) \mid 1 \leq i, j \leq 6\} .$$

Alle Elementarereignisse erhalten nach dem Prinzip von Laplace die Wahrscheinlichkeit $\frac{1}{36}$.

Wir definieren die Ereignisse

$A :=$ Augenzahl im ersten Wurf ist gerade,

$B :=$ Augenzahl im zweiten Wurf ist gerade,

$C :=$ Summe der Augenzahlen beider Würfe beträgt 7.

Es gilt $\Pr[A] = \Pr[B] = \frac{1}{2}$ und $\Pr[C] = \frac{1}{6}$. Wie groß ist $\Pr[B|A]$?

Beispiel 20 (Forts.)

Nach unserer Intuition beeinflusst der Ausgang des ersten Wurfs den zweiten Wurf nicht. Daher gewinnen wir durch das Eintreten von A keine Information in Bezug auf das Ereignis B hinzu:

$$B \cap A = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}.$$

Daraus folgt

$$\Pr[B|A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{\frac{9}{36}}{\frac{1}{2}} = \frac{1}{2} = \Pr[B] .$$

Das Eintreffen des Ereignisses B hat mit dem Ereignis A „nichts zu tun“.

Definition 21

Die Ereignisse A und B heißen **unabhängig**, wenn gilt

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] .$$

Falls $\Pr[B] \neq 0$, so können wir diese Definition zu

$$\Pr[A] = \frac{\Pr[A \cap B]}{\Pr[B]} = \Pr[A|B]$$

umschreiben.

Beispiel 20 (Zweimaliges Würfeln, Forts.)

Zur Erinnerung:

A := Augenzahl im ersten Wurf ist gerade,

B := Augenzahl im zweiten Wurf ist gerade,

C := Summe der Augenzahlen beider Würfe beträgt 7.

Bei den Ereignissen A und B ist die Unabhängigkeit klar, da offensichtlich kein kausaler Zusammenhang zwischen den Ereignissen besteht. Wie steht es mit A und C ?

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

und damit

$$\Pr[A \cap C] = \frac{3}{36} = \frac{1}{2} \cdot \frac{1}{6} = \Pr[A] \cdot \Pr[C] \text{ bzw. } \Pr[C|A] = \Pr[C] .$$

Beispiel 20 (Forts.)

Also sind auch A und C (und analog B und C) unabhängig.

Bemerkung: Im Beispiel ist $A \cap C \neq \emptyset$.

Es gilt sogar allgemein für zwei unabhängige Ereignisse A und B mit $\Pr[A], \Pr[B] > 0$, dass sie gar nicht disjunkt sein können, da ansonsten

$$0 = \Pr[\emptyset] = \Pr[A \cap B] \neq \Pr[A] \cdot \Pr[B] .$$

Beispiel 20 (Zweimaliges Würfeln (Forts.))

Zur Erinnerung:

$A :=$ Augenzahl im ersten Wurf ist gerade,

$B :=$ Augenzahl im zweiten Wurf ist gerade,

$C :=$ Summe der Augenzahlen beider Würfe beträgt 7.

Wir betrachten das Ereignis $A \cap B \cap C$. Wenn $A \cap B$ eintritt, so sind beide gewürfelten Augenzahlen gerade und somit ergibt auch die Summe davon eine gerade Zahl. Daraus folgt $\Pr[A \cap B \cap C] = 0$ bzw. $\Pr[C|A \cap B] = 0 \neq \Pr[C]$. Das Ereignis $A \cap B$ liefert uns also Information über das Ereignis C .

Definition 22

Die paarweise verschiedenen Ereignisse A_1, \dots, A_n heißen **unabhängig**, wenn für alle Teilmengen $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ mit $i_1 < i_2 < \dots < i_k$ gilt, dass

$$\Pr[A_{i_1} \cap \dots \cap A_{i_k}] = \Pr[A_{i_1}] \cdot \dots \cdot \Pr[A_{i_k}]. \quad (2)$$

Eine unendliche Familie von paarweise verschiedenen Ereignissen A_i mit $i \in \mathbb{N}$ heißt unabhängig, wenn (2) für jede endliche Teilmenge $I \subseteq \mathbb{N}$ erfüllt ist.

Lemma 23

Die (paarweise verschiedenen) Ereignisse A_1, \dots, A_n sind genau dann unabhängig, wenn für alle $(s_1, \dots, s_n) \in \{0, 1\}^n$ gilt, dass

$$\Pr[A_1^{s_1} \cap \dots \cap A_n^{s_n}] = \Pr[A_1^{s_1}] \cdot \dots \cdot \Pr[A_n^{s_n}], \quad (3)$$

wobei $A_i^0 = \bar{A}_i$ und $A_i^1 = A_i$.

Beweis:

Zunächst zeigen wir, dass aus (2) die Bedingung (3) folgt. Wir beweisen dies durch Induktion über die Anzahl der Nullen in s_1, \dots, s_n . Wenn $s_1 = \dots = s_n = 1$ gilt, so ist nichts zu zeigen. Andernfalls gelte ohne Einschränkung $s_1 = 0$. Aus dem Additionssatz folgt dann

$$\begin{aligned}\Pr[\bar{A}_1 \cap A_2^{s_2} \cap \dots \cap A_n^{s_n}] &= \Pr[A_2^{s_2} \cap \dots \cap A_n^{s_n}] \\ &\quad - \Pr[A_1 \cap A_2^{s_2} \cap \dots \cap A_n^{s_n}].\end{aligned}$$

Darauf können wir die Induktionsannahme anwenden und erhalten

$$\begin{aligned}\Pr[\bar{A}_1 \cap A_2^{s_2} \cap \dots \cap A_n^{s_n}] &= \Pr[A_2^{s_2}] \cdot \dots \cdot \Pr[A_n^{s_n}] - \Pr[A_1] \cdot \Pr[A_2^{s_2}] \cdot \dots \cdot \Pr[A_n^{s_n}] \\ &= (1 - \Pr[A_1]) \cdot \Pr[A_2^{s_2}] \cdot \dots \cdot \Pr[A_n^{s_n}],\end{aligned}$$

woraus die Behauptung wegen $1 - \Pr[A_1] = \Pr[\bar{A}_1]$ folgt.

Beweis (Forts.):

Für die Gegenrichtung zeigen wir nur, dass aus (3) $\Pr[A_1 \cap A_2] = \Pr[A_1] \cdot \Pr[A_2]$ folgt. Es gilt wegen des Satzes von der totalen Wahrscheinlichkeit, dass

$$\begin{aligned}\Pr[A_1 \cap A_2] &= \sum_{s_3, \dots, s_n \in \{0,1\}} \Pr[A_1 \cap A_2 \cap A_3^{s_3} \cap \dots \cap A_n^{s_n}] \\ &= \sum_{s_3, \dots, s_n \in \{0,1\}} \Pr[A_1] \cdot \Pr[A_2] \cdot \Pr[A_3^{s_3}] \cdot \dots \cdot \Pr[A_n^{s_n}] \\ &= \Pr[A_1] \cdot \Pr[A_2] \cdot \sum_{s_3=0,1} \Pr[A_3^{s_3}] \cdot \dots \cdot \sum_{s_n=0,1} \Pr[A_n^{s_n}] \\ &= \Pr[A_1] \cdot \Pr[A_2],\end{aligned}$$

und es folgt die Behauptung. □

Aus der Darstellung in Lemma 23 folgt die wichtige Beobachtung, dass für zwei unabhängige Ereignisse A und B auch die Ereignisse \bar{A} und B (und analog auch A und \bar{B} bzw. \bar{A} und \bar{B}) unabhängig sind!

Ebenso folgt:

Lemma 24

Seien A , B und C unabhängige Ereignisse. Dann sind auch $A \cap B$ und C bzw. $A \cup B$ und C unabhängig.

Beweis:

Die Unabhängigkeit von $A \cap B$ und C folgt unmittelbar aus Definition 22.

Aus

$$\begin{aligned}\Pr[(A \cup B) \cap C] &= \Pr[(A \cap C) \cup (B \cap C)] \\ &= \Pr[A \cap C] + \Pr[B \cap C] - \Pr[A \cap B \cap C] \\ &= \Pr[C] \cdot (\Pr[A] + \Pr[B] - \Pr[A \cap B]) \\ &= \Pr[A \cup B] \cdot \Pr[C]\end{aligned}$$

folgt die Unabhängigkeit von $A \cup B$ und C .



4. Zufallsvariablen

4.1 Grundlagen

Anstatt der Ereignisse selbst sind wir oft an „Auswirkungen“ oder „Merkmale“ der (Elementar)Ereignisse interessiert.

Definition 25

Sei ein Wahrscheinlichkeitsraum auf der Ergebnismenge Ω gegeben. Eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

heißt (numerische) Zufallsvariable.

Eine Zufallsvariable X über einer endlichen oder abzählbar unendlichen Ergebnismenge Ω heißt diskret.

Bei diskreten Zufallsvariablen ist der Wertebereich

$$W_X := X(\Omega) = \{x \in \mathbb{R}; \exists \omega \in \Omega \text{ mit } X(\omega) = x\}$$

ebenfalls wieder endlich (bzw. abzählbar unendlich).

Beispiel 26

Wir werfen eine ideale Münze drei Mal. Als Ergebnismenge erhalten wir $\Omega := \{H, T\}^3$. Die Zufallsvariable Y bezeichne die Gesamtanzahl der Würfe mit Ergebnis „Head“.

Beispielsweise gilt also $Y(HTH) = 2$ und $Y(HHH) = 3$. Y hat den Wertebereich $W_Y = \{0, 1, 2, 3\}$.

Für $W_X = \{x_1, \dots, x_n\}$ bzw. $W_X = \{x_1, x_2, \dots\}$ betrachten wir (für ein beliebiges $1 \leq i \leq n$ bzw. $x_i \in \mathbb{N}$) das Ereignis

$$A_i := \{\omega \in \Omega; X(\omega) = x_i\} = X^{-1}(x_i).$$

Bemerkung: Anstelle von $\Pr[X^{-1}(x_i)]$ verwendet man häufig auch die Schreibweise $\Pr[„X = x_i“]$. Analog setzt man

$$\begin{aligned}\Pr[„X \leq x_i“] &= \sum_{x \in W_X : x \leq x_i} \Pr[„X = x“] \\ &= \Pr[\{\omega \in \Omega; X(\omega) \leq x_i\}].\end{aligned}$$

Oft lässt man auch die Anführungszeichen weg.

Definition 27

- Die Funktion

$$f_X : \mathbb{R} \ni x \mapsto \Pr[X = x] \in [0, 1] \quad (4)$$

nennt man (diskrete) Dichte(funktion) der Zufallsvariablen X .

- Die Funktion

$$F_X : \mathbb{R} \ni x \mapsto \Pr[X \leq x] = \sum_{x' \in W_X : x' \leq x} \Pr[X = x'] \in [0, 1] \quad (5)$$

heißt Verteilung(sfunktion) der Zufallsvariablen X .

Beispiel 28

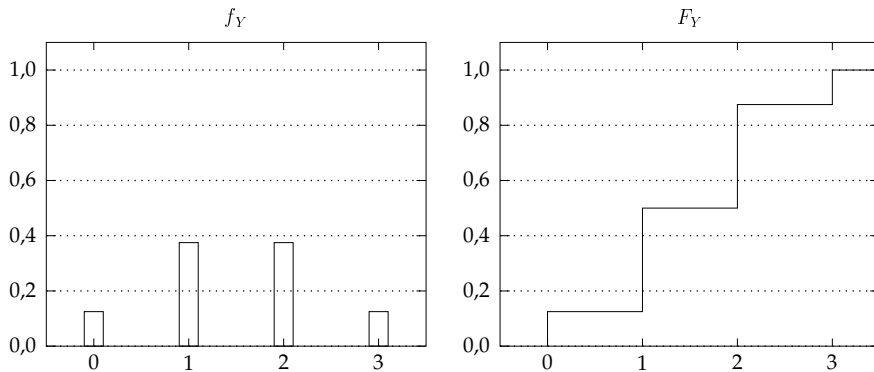
Für die Zufallsvariable Y erhalten wir

$$\Pr[Y = 0] = \Pr[TTT] = \frac{1}{8},$$

$$\Pr[Y = 1] = \Pr[HTT] + \Pr[THT] + \Pr[TTH] = \frac{3}{8},$$

$$\Pr[Y = 2] = \Pr[HHT] + \Pr[HTH] + \Pr[THH] = \frac{3}{8},$$

$$\Pr[Y = 3] = \Pr[HHH] = \frac{1}{8}.$$



Dichte und Verteilung von Y

Bemerkung: Man kann statt Ω auch den zugrunde liegenden Wahrscheinlichkeitsraum über \mathcal{W}_X betrachten.

4.2 Erwartungswert und Varianz

Definition 29

Zu einer Zufallsvariablen X definieren wir den **Erwartungswert** $\mathbb{E}[X]$ durch

$$\mathbb{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x] = \sum_{x \in W_X} x \cdot f_X(x) ,$$

sofern $\sum_{x \in W_X} |x| \cdot \Pr[X = x]$ konvergiert.

Beispiel 30

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{i=0}^3 i \cdot \Pr[Y = i] \\ &= 1 \cdot \Pr[Y = 1] + 2 \cdot \Pr[Y = 2] + 3 \cdot \Pr[Y = 3] \\ &= 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2} . \end{aligned}$$

Beispiel 31

Eine Münze wird so lange geworfen, bis sie zum ersten Mal „Head“ zeigt. Sei k die Anzahl der durchgeführten Würfe. Wenn k ungerade ist, zahlt der Spieler an die Bank k Euro. Andernfalls (k gerade) zahlt die Bank k Euro an den Spieler.

$$G := \begin{cases} k & \text{falls } k \text{ ungerade,} \\ -k & \text{falls } k \text{ gerade.} \end{cases}$$

Wie schon gesehen, gilt dann

$$\Pr[\text{„Anzahl Würfe} = k\text{“}] = (1/2)^k .$$

Damit erhalten wir

$$\mathbb{E}[G] = \sum_{k=1}^{\infty} (-1)^{k-1} \cdot k \cdot \left(\frac{1}{2}\right)^k .$$

Da

$$\sum_{k=1}^{\infty} |(-1)^{k-1} \cdot k| \cdot \left(\frac{1}{2}\right)^k \leq \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k ,$$

existiert der Erwartungswert $\mathbb{E}[G]$.

Es gilt

$$\begin{aligned}\mathbb{E}[G] &= \sum_{j=1}^{\infty} \left[(2j-1) \cdot \left(\frac{1}{2}\right)^{2j-1} - 2j \cdot \left(\frac{1}{2}\right)^{2j} \right] \\ &= \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{2j-1} \cdot [(2j-1) - j] \\ &= \frac{1}{2} \cdot \sum_{j=1}^{\infty} (j-1) \cdot \left(\frac{1}{4}\right)^{j-1} = \frac{1}{2} \cdot \frac{\frac{1}{4}}{\left(1 - \frac{1}{4}\right)^2} = \frac{2}{9} .\end{aligned}$$

Wird jedoch, um das Risiko zu steigern, der zu zahlende Betrag von k Euro jeweils auf 2^k Euro erhöht, also

$$G' := \begin{cases} 2^k & \text{falls } k \text{ ungerade,} \\ -2^k & \text{falls } k \text{ gerade,} \end{cases}$$

dann existiert $\mathbb{E}[G']$ nicht, da

$$\begin{aligned} \mathbb{E}[G'] &= \sum_{k=1}^{\infty} (-1)^{k-1} \cdot 2^k \cdot \left(\frac{1}{2}\right)^k \\ &= \sum_{k=1}^{\infty} (-1)^{k-1} = +1 - 1 + 1 - 1 + - \dots \end{aligned}$$

Berechnung des Erwartungswerts:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] = \sum_{x \in W_X} x \cdot f_X(x) \\ &= \sum_{x \in W_X} x \sum_{\omega \in \Omega: X(\omega) = x} \Pr[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega] .\end{aligned}$$

Bei unendlichen Wahrscheinlichkeitsräumen ist dabei analog zur Definition des Erwartungswerts erforderlich, dass $\sum_{\omega \in \Omega} |X(\omega)| \cdot \Pr[\omega]$ konvergiert (**absolute** Konvergenz).

Satz 32 (Monotonie des Erwartungswerts)

Seien X und Y Zufallsvariablen über dem Wahrscheinlichkeitsraum Ω mit $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$. Dann gilt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Beweis:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr[\omega] \leq \sum_{\omega \in \Omega} Y(\omega) \cdot \Pr[\omega] = \mathbb{E}[Y].$$



Aus Satz 32 folgt insbesondere, dass $a \leq \mathbb{E}[X] \leq b$ gilt, wenn für die Zufallsvariable X die Eigenschaft $a \leq X(\omega) \leq b$ für alle $\omega \in \Omega$ erfüllt ist.

4.2.1 Rechenregeln für den Erwartungswert

Oft betrachtet man eine Zufallsvariable X nicht direkt, sondern wendet noch eine Funktion darauf an:

$$Y := f(X) = f \circ X,$$

wobei $f : \mathcal{D} \rightarrow \mathbb{R}$ eine beliebige Funktion sei mit $W_X \subseteq \mathcal{D} \subseteq \mathbb{R}$.

Beobachtung: $f(X)$ ist wieder eine Zufallsvariable.

Aus

$$\Pr[Y = y] = \Pr[\{\omega \mid f(X(\omega)) = y\}] = \sum_{x: f(x)=y} \Pr[X = x]$$

folgt

$$\begin{aligned}\mathbb{E}[f(X)] &= \mathbb{E}[Y] = \sum_{y \in W_Y} y \cdot \Pr[Y = y] \\ &= \sum_{y \in W_Y} y \cdot \sum_{x: f(x)=y} \Pr[X = x] = \sum_{x \in W_X} f(x) \cdot \Pr[X = x] \\ &= \sum_{\omega \in \Omega} f(X(\omega)) \cdot \Pr[\omega].\end{aligned}$$

Satz 33 (Linearität des Erwartungswerts, einfache Version)

Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b.$$

Beweis:

$$\begin{aligned}\mathbb{E}[a \cdot X + b] &= \sum_{x \in W_X} (a \cdot x + b) \cdot \Pr[X = x] \\ &= a \cdot \sum_{x \in W_X} x \cdot \Pr[X = x] + b \cdot \sum_{x \in W_X} \Pr[X = x] \\ &= a \cdot \mathbb{E}[X] + b.\end{aligned}$$



Satz 34

Sei X eine Zufallsvariable mit $W_X \subseteq \mathbb{N}_0$. Dann gilt

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$

Beweis:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \Pr[X = i] = \sum_{i=0}^{\infty} \sum_{j=1}^i \Pr[X = i] \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \Pr[X = i] = \sum_{j=1}^{\infty} \Pr[X \geq j].\end{aligned}$$



Definition 35

Sei X eine Zufallsvariable und A ein Ereignis mit $\Pr[A] > 0$. Die **bedingte Zufallsvariable** $X|A$ besitzt die Dichte

$$f_{X|A}(x) := \Pr[X = x \mid A] = \frac{\Pr[„X = x“ \cap A]}{\Pr[A]}.$$

Die Definition von $f_{X|A}$ ist zulässig, da

$$\sum_{x \in W_X} f_{X|A}(x) = \sum_{x \in W_X} \frac{\Pr[„X = x“ \cap A]}{\Pr[A]} = \frac{\Pr[A]}{\Pr[A]} = 1.$$

Der Erwartungswert $\mathbb{E}[X|A]$ der Zufallsvariablen $X|A$ berechnet sich entsprechend:

$$\mathbb{E}[X|A] = \sum_{x \in W_X} x \cdot f_{X|A}(x).$$

Satz 36

Sei X eine Zufallsvariable. Für paarweise disjunkte Ereignisse A_1, \dots, A_n mit $A_1 \cup \dots \cup A_n = \Omega$ und $\Pr[A_1], \dots, \Pr[A_n] > 0$ gilt

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X|A_i] \cdot \Pr[A_i].$$

Für paarweise disjunkte Ereignisse A_1, A_2, \dots mit $\bigcup_{i=1}^{\infty} A_i = \Omega$ und $\Pr[A_1], \Pr[A_2], \dots > 0$ gilt analog

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{E}[X|A_i] \cdot \Pr[A_i],$$

sofern die Erwartungswerte auf der rechten Seite alle existieren und die Summe $\sum_{i=1}^{\infty} |\mathbb{E}[X|A_i]| \cdot \Pr[A_i]$ konvergiert.

Beweis:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in W_X} x \cdot \Pr[X = x] = \sum_{x \in W_X} x \cdot \sum_{i=1}^n \Pr[X = x|A_i] \cdot \Pr[A_i] \\ &= \sum_{i=1}^n \Pr[A_i] \sum_{x \in W_X} x \cdot \Pr[X = x|A_i] = \sum_{i=1}^n \Pr[A_i] \cdot \mathbb{E}[X|A_i].\end{aligned}$$

Der Beweis für den unendlichen Fall verläuft analog.



Beispiel 37

Wir werfen eine Münze so lange, bis zum ersten Mal „Kopf“ erscheint. Dies geschehe in jedem Wurf unabhängig mit Wahrscheinlichkeit p . Wir definieren dazu die Zufallsvariable $X :=$ „Anzahl der Würfe“. Wir haben bereits gesehen, dass

$$\Pr[X = k] = p(1 - p)^{k-1}$$

und damit

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot p(1 - p)^{k-1} = p \cdot \frac{1}{(1 - (1 - p))^2} = \frac{1}{p}.$$

Beispiel 37

Andere Berechnungsmethode: (gestützt auf Satz 36)

Definiere das Ereignis

$$K_1 := \text{„Im ersten Wurf fällt Kopf“}.$$

Offensichtlich gilt $\mathbb{E}[X|K_1] = 1$.

Nehmen wir nun an, dass im ersten Wurf *nicht* „Kopf“ gefallen ist. Wir starten das Experiment neu.

Beispiel 37

Sei X' die Anzahl der Würfe bis zum ersten Auftreten von „Kopf“ im neu gestarteten Experiment. Wegen der Gleichheit der Experimente gilt $\mathbb{E}[X'] = \mathbb{E}[X]$. Damit schließen wir

$$\mathbb{E}[X|\bar{K}_1] = 1 + \mathbb{E}[X'] = 1 + \mathbb{E}[X]$$

und erhalten mit Satz 36:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X|K_1] \cdot \Pr[K_1] + \mathbb{E}[X|\bar{K}_1] \cdot \Pr[\bar{K}_1] \\ &= 1 \cdot p + (1 + \mathbb{E}[X]) \cdot (1 - p).\end{aligned}$$

Daraus ergibt sich wiederum $\mathbb{E}[X] = 1/p$.

4.2.2 Varianz

Wir betrachten die beiden folgenden Zufallsexperimente:

- ① Wir würfeln (mit einem fairen Würfel), bei **gerader** Augenzahl erhalten wir 1 Euro, bei **ungerader** Augenzahl müssen wir 1 Euro bezahlen.
- ② Wir würfeln (mit einem fairen Würfel), bei 6 Augen erhalten wir 5 Euro, ansonsten müssen wir 1 Euro bezahlen.

Beobachtung:

In beiden Fällen ist der erwartete Gewinn $= 0$.

Dennoch sind die „Schwankungen“ im ersten Fall geringer als im zweiten.

Eine nahe liegende Lösung wäre,

$$\mathbb{E}[|X - \mu|]$$

zu berechnen, wobei $\mu = \mathbb{E}[X]$ sei. Dies scheitert jedoch meist an der „unhandlichen“ Betragsfunktion. Aus diesem Grund betrachtet man stattdessen $\mathbb{E}[(X - \mu)^2]$, also die quadratische Abweichung vom Erwartungswert.

Definition 38

Für eine Zufallsvariable X mit $\mu = \mathbb{E}[X]$ definieren wir die *Varianz* $\text{Var}[X]$ durch

$$\text{Var}[X] := \mathbb{E}[(X - \mu)^2] = \sum_{x \in W_X} (x - \mu)^2 \cdot \Pr[X = x].$$

Die Größe $\sigma := \sqrt{\text{Var}[X]}$ heißt *Standardabweichung* von X .

Satz 39

Für eine beliebige Zufallsvariable X gilt

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Beweis:

Sei $\mu := \mathbb{E}[X]$. Nach Definition gilt

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2 - 2\mu \cdot X + \mu^2] \\&= \sum_{x \in W_X} (x^2 - 2\mu \cdot x + \mu^2) \cdot \Pr[X = x] \\&= \sum_{x \in W_X} x^2 \cdot \Pr[X = x] - \sum_{x \in W_X} 2\mu \cdot x \cdot \Pr[X = x] + \sum_{x \in W_X} \mu^2 \cdot \Pr[X = x] \\&= \mathbb{E}[X^2] - 2\mu \cdot \mathbb{E}[X] + \mu^2 \\&= \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

Beispiel 40

- ① Wir würfeln (mit einem fairen Würfel), bei **gerader** Augenzahl erhalten wir 1 Euro, bei **ungerader** Augenzahl müssen wir 1 Euro bezahlen. Es ist

$$\mu = 0 \text{ und } \text{Var}[X] = \frac{1}{2} \cdot 1^2 + \frac{1}{2} \cdot (-1)^2 = 1.$$

- ② Wir würfeln (mit einem fairen Würfel), bei 6 Augen erhalten wir 5 Euro, ansonsten müssen wir 1 Euro bezahlen.
Es ist

$$\mu = 0 \text{ und } \text{Var}[X] = \frac{1}{6} \cdot 5^2 + \frac{5}{6} \cdot (-1)^2 = 5.$$

Satz 41

Für eine beliebige Zufallsvariable X und $a, b \in \mathbb{R}$ gilt

$$\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X].$$

Beweis:

Aus der in Satz 33 gezeigten Linearität des Erwartungswerts folgt $\mathbb{E}[Y + b] = \mathbb{E}[Y] + b$. Zusammen mit der Definition der Varianz ergibt sich damit sofort

$$\text{Var}[Y + b] = \mathbb{E}[(Y + b - \mathbb{E}[Y + b])^2] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \text{Var}[Y].$$

Weiter folgt mit Satz 39:

$$\text{Var}[a \cdot X] = \mathbb{E}[(aX)^2] - \mathbb{E}[aX]^2 = a^2\mathbb{E}[X^2] - (a\mathbb{E}[X])^2 = a^2 \cdot \text{Var}[X],$$

und daraus zusammen die Behauptung. □

Der Erwartungswert und die Varianz gehören zu den so genannten **Momenten** einer Zufallsvariablen:

Definition 42

Für eine Zufallsvariable X nennen wir $\mathbb{E}[X^k]$ das **k -te Moment** und $\mathbb{E}[(X - \mathbb{E}[X])^k]$ das **k -te zentrale Moment**.

Der Erwartungswert ist also identisch zum ersten Moment, während die Varianz dem zweiten zentralen Moment entspricht.

4.3 Mehrere Zufallsvariablen

Beispiel 43

Aus einem Skatblatt mit 32 Karten ziehen wir zufällig eine Hand von zehn Karten sowie einen Skat von zwei Karten. Unter den Karten gibt es vier Buben. Die Zufallsvariable X zählt die Anzahl der Buben in der Hand, während Y die Anzahl der Buben im Skat angibt. Die Werte von X und Y hängen offensichtlich stark voneinander ab. Beispielsweise muss $Y = 0$ sein, wenn $X = 4$ gilt.

Wie kann man mit mehreren Zufallsvariablen über demselben Wahrscheinlichkeitsraum rechnen, auch wenn sie, wie im obigen Beispiel, sehr voneinander abhängig sind? Wir untersuchen Wahrscheinlichkeiten der Art

$$\Pr[X = x, Y = y] = \Pr[\{\omega; X(\omega) = x, Y(\omega) = y\}].$$

Beispiel 44

Wenn wir nur die Zufallsvariable X betrachten, so gilt für $0 \leq x \leq 4$

$$\Pr[X = x] = \frac{\binom{4}{x} \binom{28}{10-x}}{\binom{32}{10}}.$$

Allgemein nennt man Zufallsvariablen mit der Dichte

$$\Pr[X = x] = \frac{\binom{b}{x} \binom{a}{r-x}}{\binom{a+b}{r}}$$

hypergeometrisch verteilt. Durch diese Dichte wird ein Experiment modelliert, bei dem r Elemente ohne Zurücklegen aus einer Grundmenge der Mächtigkeit $a + b$ mit b besonders ausgezeichneten Elementen gezogen werden.

Beispiel 44 (Forts.)

Die Zufallsvariable Y ist für sich gesehen ebenfalls hypergeometrisch verteilt mit $b = 4$, $a = 28$ und $r = 2$.

Für X und Y zusammen gilt jedoch z.B.

$$\Pr[X = 4, Y = 1] = 0,$$

und allgemein

$$\Pr[X = x, Y = y] = \frac{\binom{4}{x} \binom{28}{10-x} \binom{4-x}{y} \binom{28-(10-x)}{2-y}}{\binom{32}{10} \binom{22}{2}}.$$

Bemerkung: Die Schreibweise $\Pr[X = x, Y = y]$ stellt eine Abkürzung von $\Pr[„X = x \wedge Y = y“]$ dar. Ein anderes Beispiel ist

$$\Pr[X \leq x, Y \leq y_1, \sqrt{Y} = y_2].$$