

Kapitel T

Wahrscheinlichkeitsrechnung

*Statistical thinking will one day be as necessary
for efficient citizenship as the ability to read or write.*

H.G. Wells (1866–1946)

*Les questions les plus importantes de la vie ne sont en effet,
pour la plupart, que des problèmes de probabilité.*

Pierre-Simon de Laplace (1749–1827)

*So you're telling me it is a matter of probability and odds;
I was worried there was some chance involved.*

Vesper Lynd zu James Bond im Film *Casino Royale* (2006)

Inhalt dieses Kapitels T

- 1 Diskrete Wahrscheinlichkeitsräume
 - Zufall und Wahrscheinlichkeit: Grundbegriffe
 - Rechnen mit Ereignissen: Wahrscheinlichkeitsräume
- 2 Bedingte Wahrscheinlichkeit und Unabhängigkeit
 - Bedingte Wahrscheinlichkeiten, Formel von Bayes
 - Stochastische Unabhängigkeit von Ereignissen
- 3 Das Gesetz der großen Zahlen
 - Zufallsvariablen, Erwartung und Varianz
 - Die Ungleichungen von Chebychev
 - Unabhängigkeit und Gesetz der großen Zahlen
 - Korrelation, Unabhängigkeit und Kausalität
- 4 Fazit: Wahrscheinlichkeitsrechnung
 - Weitere Aufgaben und Anwendungsbeispiele
 - Sex: Was nützen dem Pfau seine Federn?
 - Tischkicker: Das Runde muss ins Eckige.
 - Google: Die zufällige Irrfahrt im Internet

Ziele der Wahrscheinlichkeitsrechnung


Unser Ziel sind **rationale Entscheidungen unter Unsicherheit**.

- **Experimente** unter Einfluss von Zufällen
- **Prozesse** mit unsicherem Ausgang

Wir suchen **nachvollziehbar begründete, quantitative Aussagen**.

- Ungenauigkeit von Messungen, Konfidenzintervalle.
- Ausfallwahrscheinlichkeit, Lebensdauer von Bauteilen
- Bewertung von Risiken (Unfälle, Schutz, Versicherung)
- Glücksspiele (Lotto, Roulette, Poker, Aktienbörse, Klausur)

That's life: Fast alles im Leben ist ein Wahrscheinlichkeitsexperiment!
Ingenieur:innen müssen Risiken abschätzen, minimieren, verkaufen bzw.
Chancen berechnen, maximieren, kaufen. Dazu dient die Stochastik!

 Literatur: H.-O. Georgii: *Stochastik*. De Gruyter 2009
E. Kreyszig: *Advanced Engineering Mathematics*. Wiley 2010
U. Krengel: *Einführung in die Wahrscheinlichkeitstheorie*. Vieweg 2005
W. Feller: *Probability Theory and Its Applications*. Wiley 1950–1971

Ziele der Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung beginnt im 17. Jh. mit Rechnungen zu Glücksspielen. T417
Heutzutage wird sie überall eingesetzt, von der Wettervorhersage bis zur Industrieproduktion und klinischen Tests neuer Medikamente. In ihrer Bedeutung für den Alltag kommt sie wohl gleich nach Grundrechenarten und Dreisatz. Ich möchte wetten, Wahrscheinlichkeitsrechnung wird von Otto Normalbürger (m/w/d) vielhundertmal häufiger benötigt und benutzt als die sagenhafte Mitternachtsformel! Das hängt natürlich davon ab, was Otto im Alltag oder Beruf so tut. . .

Viele Vorgänge haben unsicheren Ausgang; stochastische Argumente sind daher unausweichlich. Leider erstaunlich häufig werden aber selbst einfache Fragen falsch behandelt – manchmal mit katastrophalen Folgen. Damit es Ihnen nicht so, sondern besser ergeht, sollen Sie hier mit (einer homöopathischen Dosis) Wahrscheinlichkeitsrechnung geimpft werden. Zu Experten werden Sie dadurch noch nicht, aber die Grundregeln müssen Sie kennen und fehlerfrei anwenden können.

Statt *Wahrscheinlichkeitsrechnung* spricht man gleichbedeutend von *Wahrscheinlichkeitstheorie*: Auf Grundlage eines geeigneten Modells (eines Wahrscheinlichkeitsraums wie unten erklärt) erlaubt sie die Berechnung der Wahrscheinlichkeiten zu den betrachteten Ereignissen.

Ihr Gegenstück ist die *Statistik*, also die Analyse empirischer Daten (Messwerte, Beobachtungen). Diese versucht, möglichst verlässliche Daten zu erheben und daraus relevante Informationen zu extrahieren, insbesondere um ein geeignetes Modell zu finden, zu testen oder zu kalibrieren.

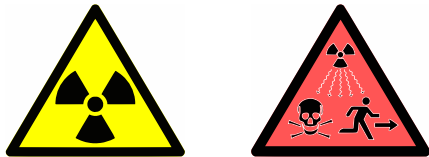
Wahrscheinlichkeitstheorie und Statistik ergänzen sich demnach. Beide fasst man unter dem Oberbegriff *Stochastik* zusammen, von altgriechisch $\sigma\tau\omicron\chi\acute{\alpha}\zeta\omicron\mu\alpha\iota$ [stochazomai] ‘zielen, vermuten, erraten’, auch $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\acute{o}\varsigma$ [stochastikos] ‘scharfsinnig’.

Typische Anwendung: Ausfallwahrscheinlichkeit

- Aufgabe:** Drei unabhängige Bauteile haben Ausfallwkten 0.3, 0.4, 0.5.
- (0) Mit welcher Wkt fällt mindestens ein Teil aus? (1) Genau 0, 1, 2, 3?
 - (2) Wenn nur genau ein Teil ausfällt, mit welcher Wkt ist es A , B , C ?
 - (3) Konstruieren Sie als Modell hierfür explizit einen WRaum (Ω, \mathbf{P}) .

Techniken: Rechenregeln für WRäume, bedingte Wkt, Bayes.

Lösung ab Seite T409: Die Wkten in (1) sind 0.21, 0.44, 0.29, 0.06.



Aus der Presse: „In 14 000 Reaktorjahren sind 4 Unfälle der höchsten Stufe aufgetreten. Demnach tritt bei 143 europäischen Reaktoren in 30 Jahren ein solcher Unfall mit Wahrscheinlichkeit von über 100% auf.“

Aufgabe: Was ist hieran falsch? Welche Rechnung wäre richtig?

Techniken: Ausfallwkt (exakt), Poisson-Verteilung (gute Näherung).

Lösung ab Seite U109: Die Wkt ist mit $\approx 70\%$ erschreckend hoch.

Typische Anwendung: Chancen beim Glücksspiel

Die Wahrscheinlichkeitsrechnung hat viele praktische Anwendungen! Ihre Ursprünge liegen im Glücksspiel. Eines der ersten Probleme war die gerechte Teilung bei vorzeitigem Spielabbruch. Konkretes Beispiel:



Zwei gleich starke Teams spielen Tischkicker bis 10.

Aufgabe: Wie stehen die Chancen bei 9 : 8? bei 4 : 7?

Lösung ab Seite T441:

(a) 75 : 25 (leicht) bzw.

(b) 14.5 : 85.5 (knifflig)

Aufgabe: Sie wiederholen 400 mal unabhängig ein Experiment mit Trefferwkt 20%. Mit welcher Wkt erhalten Sie 60 bis 90 Treffer?

Techniken: Erwartung und Varianz, Binomialverteilung (exakt), lokaler Grenzwertsatz (gute Näherung), Chebychev (Abschätzung).
Lösung ab Seite V317: Die gesuchte Wkt beträgt recht genau 90%.

Typische Anwendung: Gesetze der großen Zahlen

In der Praxis ist nichts exakt. Jede Messung X ist zufälligen Fehlern unterworfen, wiederholte Messwerte schwanken um den Erwartungswert μ mit Streuung $\sigma > 0$. Ingenieur:innen müssen sorgsam und sachgerecht mit Messfehlern umgehen: Konfidenzintervalle, Fehlerrechnung, etc.

Beispiel: Sie führen $n = 400$ unabhängige Messungen X_1, X_2, \dots, X_{400} durch und erhalten Messwerte x_1, x_2, \dots, x_{400} . Sie schätzen μ durch den

$$\text{Stichprobenmittelwert: } \hat{x} := \frac{1}{n} \sum_{k=1}^n x_k \quad \begin{array}{l} \text{Messung} \\ = \\ \text{(Beispiel)} \end{array} \quad 57.5$$

Die Streuung von X schätzen Sie (dank der Stichprobe) auf $\sigma \approx 10$. Wie genau ist \hat{x} ? Wo liegt vermutlich der wahre Erwartungswert μ ?

$$I_1 = [57, 58], \quad I_2 = [56.5, 58.5], \quad I_3 = [56, 59], \quad I_k = [57.5 \pm k \cdot 0.5]$$

Aufgabe: Welches Intervall I_k überdeckt den Erwartungswert μ mit Sicherheit $\alpha_k \geq 95\%$ bzw. $\alpha_k \geq 99\%$? **Lösung:** Es genügen I_2 bzw. I_3 !

Techniken: Erwartung und Varianz, Chebychev (Abschätzung, T337), genauer Normalverteilung und lokaler / zentraler Grenzwertsatz (W201).

Gewählte Themen und benötigte Techniken

Diese Aufgaben nenne ich stellvertretend für unsere vier Kapitel über Wahrscheinlichkeitsrechnung. Meine Auswahl der Themen (in Breite und Tiefe) ist naturgemäß ein Kompromiss: Für den einen ist es schon zuviel, für die andere noch lange nicht genug. Ihr Bedarf hängt ganz davon ab, womit Sie sich beschäftigen. Ich will und kann nur einen Einstieg bieten, damit Sie diese wichtigen Techniken je nach Bedarf vertiefen können.

Wenn Sie das alles schon aus der Schule kennen und beherrschen, dann können Sie die folgenden Kapitel als Wiederholung durcharbeiten. Vermutlich ist aber für die meisten noch viel Neues und Schönes dabei.

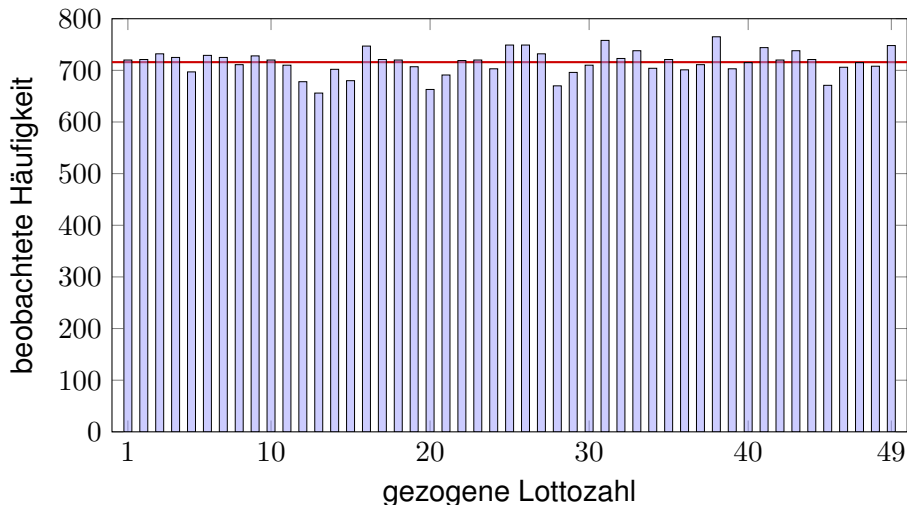
Welche mathematischen Grundlagen und Techniken benötigen wir?

Die Anfänge der Wahrscheinlichkeitsrechnung, wie hier dargelegt, erfordern vor allem begriffliche Präzision (Ereignisse und WMaße), sind aber technisch noch leicht (Reihen, Kombinatorik). Kontinuierliche Verteilungen, wie die Normalverteilung, erfordern zudem Integration.

Überall nützlich sind Grenzwerte und Näherungen, wie Stirling-Formel oder Taylor-Entwicklung. Den zentralen Grenzwertsatz W1D schließlich können wir mit der Fourier-Transformation schmerzfrei nachrechnen.

Statistik der ersten 5011 Lottoziehungen

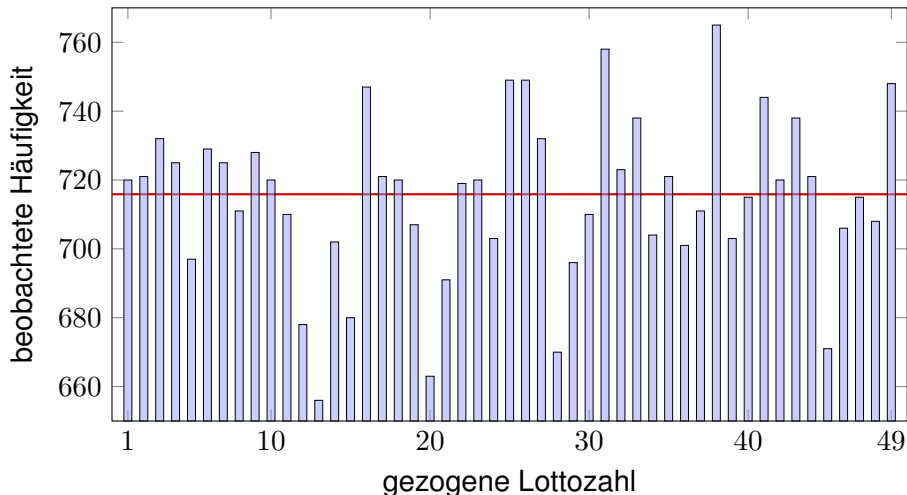
Ziehungen vom 09.10.1955 bis 09.11.2011, Sa+Mi, mit Zusatzzahl.



Modell und vereinfachende Annahme: Die Ziehungen unterliegen nur dem Zufall, und alle Zahlen $1, \dots, 49$ sind dabei gleich wahrscheinlich.

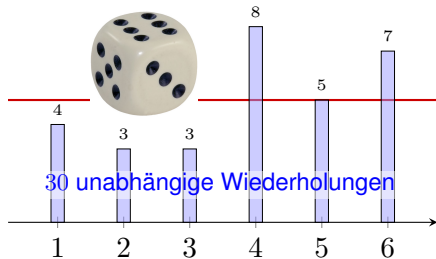
Statistik der ersten 5011 Lottoziehungen

Titanic (26.10.2011): „Experten warnen: Lottozahlen oft reiner Zufall!“

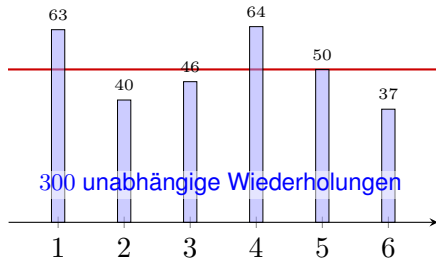


Alle Zahlen sollten gleich wahrscheinlich sein. Sind sie es wirklich?
Jedes Modell ist eine **Annahme** und sollte kritisch überprüft werden.

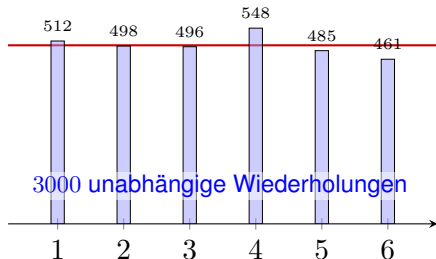
Statistik: empirische Häufigkeiten beim Würfeln



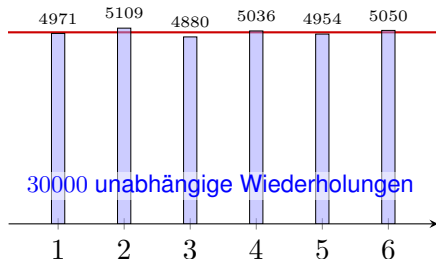
30 unabhängige Wiederholungen



300 unabhängige Wiederholungen



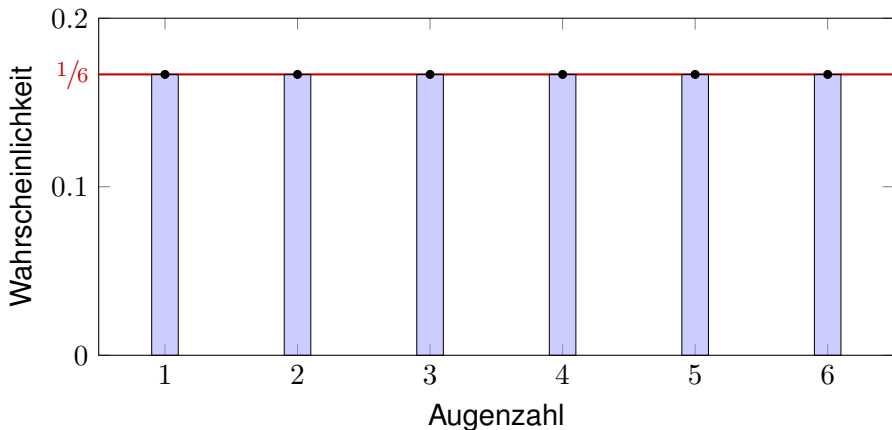
3000 unabhängige Wiederholungen



30000 unabhängige Wiederholungen

Wie sollen wir mit etwas rechnen, wie dem Zufall, das prinzipiell unberechenbar scheint? Das **empirische Gesetz der großen Zahlen** ist folgende Erfahrungstatsache: Bei häufiger unabhängiger Wiederholung nähern sich die relativen Häufigkeiten einem Grenzwert an.

Stochastik: Modell eines fairen Würfels



Diese **Wahrscheinlichkeiten** bilden das stochastische Modell.

Es handelt sich um eine *Abstraktion*, also *theoretische* Werte. Hiervon zu unterscheiden sind die *empirischen relativen Häufigkeiten*: Diese werden bei n -maliger Durchführungen des Experiments beobachtet. Nur für $n \rightarrow \infty$ nähern sie sich den Wahrscheinlichkeiten an.

😊 Satz T3J präzisiert und quantifiziert diese Aussage als **Gesetz der großen Zahlen**.

Stochastisches Modell eines fairen Würfels



Mögliche **Ergebnisse** sind die Zahlen 1, 2, 3, 4, 5, 6.

Annahme: Alle Ergebnisse sind gleich wahrscheinlich.

Das bedeutet anschaulich: Wenn Sie „sehr oft“ würfeln, dann treten die Ergebnisse 1, 2, 3, 4, 5, 6 gleich häufig auf, jedes also in $1/6$ aller Fälle.

Das ist zunächst eine Erfahrung, sodann eine empirische **Beobachtung**. Für das theoretische Modell hingegen ist es die grundlegende **Annahme**.

Abstraktion des Experiments zu einem mathematischen Modell (Ω, p) :

- Die **Ergebnismenge** ist hier $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Die **Elementarwahrscheinlichkeiten** $p: \Omega \rightarrow [0, 1]$ sind

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

😊 In diesem Falle ist die Gleichverteilung plausibel aufgrund der Symmetrie des Würfels.

Dieses Modell besteht nicht aus empirischen Daten, sondern ist eine axiomatische Festlegung! Wir fixieren damit präzise unsere Annahmen und leiten alle weiteren Rechnungen hieraus ab.

Wie können Sie das interpretieren? oder messen? Was bedeutet Wahrscheinlichkeit $q \in [0, 1]$? Vereinfacht gesagt: Bei häufiger Wiederholung tritt das Ergebnis im Anteil q aller Fälle ein.

Wkt von Ergebnissen und Ereignissen

Aufgabe: Sie würfeln mit einem fairen Würfel. Wie wahrscheinlich ist das Ereignis A : gerade Zahl? B : mindestens 5? $A \cap B$: beides?

Lösung: (1) Das Ereignis $A = \{2, 4, 6\} \subseteq \Omega$ hat die Wahrscheinlichkeit

$$\mathbf{P}(A) = \mathbf{P}(\{2, 4, 6\}) = p(2) + p(4) + p(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

(2) Das Ereignis $B = \{5, 6\} \subseteq \Omega$ hat die Wahrscheinlichkeit

$$\mathbf{P}(B) = \mathbf{P}(\{5, 6\}) = p(5) + p(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

(3) Die Schnittmenge $A \cap B = \{6\}$ hat die Wahrscheinlichkeit

$$\mathbf{P}(A \cap B) = \mathbf{P}(\{6\}) = \frac{1}{6} = \mathbf{P}(A) \cdot \mathbf{P}(B) = \frac{1}{2} \cdot \frac{1}{3}.$$

Die Ereignisse A und B sind hier **stochastisch unabhängig**.

Das Ereignis „gerade Zahl“ entspricht der Teilmenge $A = \{2, 4, 6\} \subseteq \Omega$. Dieses Ereignis A tritt genau dann ein, wenn ein Ergebnis $\omega \in A$ gewürfelt wird. Interpretation der Wkt: Das Ereignis $A = \{2, 4, 6\}$ tritt in der Hälfte aller Fälle ein, das Ereignis $B = \{5, 6\}$ nur in einem Drittel aller Fälle. Beide zusammen, also das Ereignis $A \cap B = \{6\}$ tritt in einem Sechstel aller Fälle ein.

😊 Unsere Rechnungen nutzen ausschließlich das zuvor festgelegte Modell eines fairen Würfels. Ein gezinkter Würfel hat andere Wahrscheinlichkeiten, wie im nachfolgenden Beispiel illustriert.

Stochastisches Modell eines gezinkten Würfels

Man kann Würfel durch ungleiche Masse manipulieren (Bleieinlage).
Präzisionswürfel sind daher transparent, um Betrug zu erschweren.



Präzisionswürfel aus Kunststoff

Gleiche Wkten!



Handgesägte Würfel aus Holz

Ungleiche Wkten?

Wir nehmen an, die Wahrscheinlichkeiten sind nicht gleich, sondern:

$$p(1) = 0.15, \quad p(2) = p(3) = p(4) = p(5) = 0.16, \quad p(6) = 0.21$$

Dies ist ein anderes stochastisches Modell (Ω, p) als die zuvor betrachtete Gleichverteilung!
Es besagt anschaulich: Wenn Sie „sehr oft“ würfeln, dann tritt die Augenzahl 1 etwa in 15% aller Fälle auf, 2, 3, 4, 5 jeweils etwa in 16% aller Fälle, und 6 in etwa in 21% aller Fälle.

Wkt von Ergebnissen und Ereignissen

Aufgabe: Sie würfeln mit diesem gezinkten Würfel. Wie wahrscheinlich ist das Ereignis A : gerade Zahl? B : mindestens 5? $A \cap B$: beides?

Lösung: (1) Das Ereignis $A = \{2, 4, 6\} \subseteq \Omega$ hat die Wahrscheinlichkeit

$$\mathbf{P}(A) = \mathbf{P}(\{2, 4, 6\}) = p(2) + p(4) + p(6) = 0.16 + 0.16 + 0.21 = 0.53.$$

(2) Das Ereignis $B = \{5, 6\} \subseteq \Omega$ hat die Wahrscheinlichkeit

$$\mathbf{P}(B) = \mathbf{P}(\{5, 6\}) = p(5) + p(6) = 0.16 + 0.21 = 0.37.$$

(3) Die Schnittmenge $A \cap B = \{6\}$ hat die Wahrscheinlichkeit

$$\mathbf{P}(\{6\}) = 0.21 \neq \mathbf{P}(A) \cdot \mathbf{P}(B) = 0.1961.$$

Die Ereignisse A und B sind hier **stochastisch abhängig**.

Interpretation: Das Ereignis A tritt in 53% aller Fälle ein, und B tritt in 37% aller Fälle ein. Die Wahrscheinlichkeiten sind hier anders als beim fairen Würfel. Es handelt sich um zwei verschiedene Modelle, die unterschiedliche Vorhersagen machen. Welches Modell in einem konkreten Fall das „richtige“ oder das „realistischere“ ist, bleibt hierbei zunächst noch offen. Dazu muss man konkurrierende Modelle empirisch testen und das besser passende auswählen.

Wie beschreiben wir Zufallsereignisse?

Ziel: Geeignete Sprache zur Beschreibung von Zufallsereignissen.

- Möglichst klare und einfache, doch unmissverständliche Notation
- Bequem, flexibel und sicher anwendbar in konkreten Rechnungen

Hierfür bietet sich die universelle Sprache der Mengenlehre an!

Alle Ergebnisse fassen wir zur **Ergebnismenge** Ω zusammen.

Jedes Element $\omega \in \Omega$ nennen wir ein **Ergebnis** oder einen Ausgang.

Jede Teilmenge $A \subseteq \Omega$ nennen wir ein **Ereignis**. Wir vereinbaren dabei:

Das Ereignis A tritt genau dann ein, wenn ein Ergebnis $\omega \in A$ eintritt.

Beispiel Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\}$ besteht aus sechs Ergebnissen.

Das Ereignis „gerade Zahl“ entspricht der Teilmenge $A = \{2, 4, 6\}$.

Das Ereignis „mindestens 5“ entspricht der Teilmenge $B = \{5, 6\}$.

(Die Zusammenfassung als Menge ist eine Oder-Verknüpfung.)

„Ergebnis“ und „Ereignis“ klingen ähnlich, bedeuten aber verschiedenes:

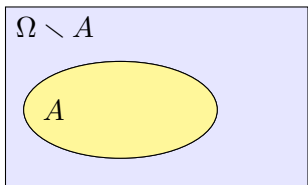
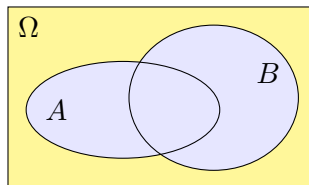
Unterscheiden Sie beides sorgfältig! Dann weiß jede:r, was gemeint ist.

Wir unterscheiden insbesondere das Ergebnis $\omega \in \Omega$ als ein Element von Ω und das Elementarereignis $\{\omega\} \subseteq \Omega$ als eine Teilmenge von Ω .

Grundmenge und Komplemente

Wir bezeichnen mit Ω unsere **Gesamtmenge**, \emptyset ist die **leere Menge**.

Wir schreiben $A \subseteq B$, wenn jedes Element von A auch in B liegt.



Die Teilmenge aller Elemente mit einer **Eigenschaft** P schreiben wir

$$\Omega_P = \{ \omega \in \Omega \mid P(\omega) \}.$$

Beispiel: Für $\Omega = \{1, 2, \dots, 6\}$ gilt $\{ \omega \in \Omega \mid \omega \text{ gerade} \} = \{2, 4, 6\}$.

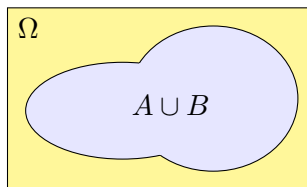
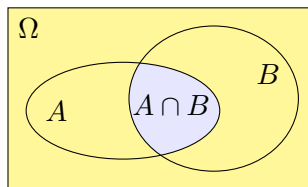
Das **Komplement** von A in Ω ist „ Ω ohne A “ oder „nicht A “:

$$\Omega \setminus A := \bar{A} = \{ \omega \in \Omega \mid \omega \notin A \}.$$

Es besteht aus allen Elementen von Ω , die nicht in A liegen.

Gebräuchliche Schreibweisen: $\Omega \setminus A = \complement_{\Omega} A = \complement A = A^c = \bar{A} = \dots$

Schnitt und Vereinigung



Die **Schnittmenge** „ A und B “ ist

$$A \cap B = \{ \omega \in \Omega \mid \omega \in A \text{ und } \omega \in B \}.$$

Sie besteht aus den Elementen, die sowohl in A als auch in B liegen.

Die **Vereinigung** „ A oder B “ ist

$$A \cup B = \{ \omega \in \Omega \mid \omega \in A \text{ oder } \omega \in B \}.$$

Sie besteht aus den Elementen, die in A oder B liegen (oder in beiden).

Zwei Mengen A, B heißen **disjunkt**, wenn $A \cap B = \emptyset$ gilt.

In diesem Fall erhalten wir die **disjunkte Vereinigung**:

$$A \sqcup B := A \cup B \quad \text{mit} \quad A \cap B = \emptyset.$$

Schnitt und Vereinigung

Die Vereinigung einer endlichen Familie von Mengen:

$$\bigcup_{k=0}^n A_k = A_0 \cup A_1 \cup A_2 \cup \dots \cup A_n$$

Vereinigung einer (abzählbar) unendlichen Familie:

$$\bigcup_{k=0}^{\infty} A_k = \{ \omega \mid \omega \in A_k \text{ für mind. ein } k \in \mathbb{N} \}$$

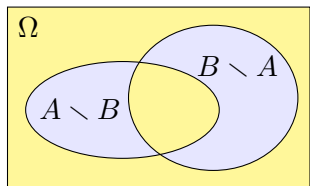
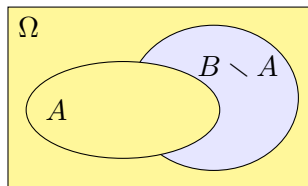
Die Schnittmenge einer endlichen Familie von Mengen:

$$\bigcap_{k=0}^n A_k = A_0 \cap A_1 \cap A_2 \cap \dots \cap A_n$$

Schnittmenge einer (abzählbar) unendlichen Familie:

$$\bigcap_{k=0}^{\infty} A_k = \{ \omega \mid \omega \in A_k \text{ für alle } k \in \mathbb{N} \}$$

Restmenge und symmetrische Differenz



Die **Restmenge** „ B ohne A “ ist

$$B \setminus A := \{ \omega \in B \mid \omega \notin A \}.$$

Sie besteht aus allen Elementen, die zwar in B nicht aber in A liegen.

Die **symmetrische Differenz** „entweder A oder B “ ist

$$A \Delta B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A).$$

Sie besteht aus allen Elementen, die entweder in A oder in B liegen, aber nicht in beiden zugleich! Wir nennen dies auch **exklusives Oder**.

Rechenregeln für Vereinigung \cup und Schnitt \cap Satz T1A: Rechenregeln für Vereinigung \cup und Schnitt \cap

Folgende nützliche Rechenregeln gelten für alle $A, B, C \subseteq \Omega$:

Neutralität, Absorbtion, Idempotenz, Komplemente:

$$A \cup \emptyset = A, \quad A \cup \Omega = \Omega, \quad A \cup A = A, \quad A \cup \bar{A} = \Omega$$

$$A \cap \Omega = A, \quad A \cap \emptyset = \emptyset, \quad A \cap A = A, \quad A \cap \bar{A} = \emptyset$$

Kommutativität:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

Assoziativität:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$$

Distributivität:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Übung: Illustrieren und begründen Sie sorgfältig jede dieser Regeln.

Rechenregeln für Komplement und Restmenge

Satz T1B: Regeln von DeMorgan für Komplemente

Das Komplement einer Menge A in Ω ist definiert durch

$$\bar{A} = \Omega \setminus A := \{\omega \in \Omega \mid \omega \notin A\}.$$

Es gilt $\bar{\Omega} = \emptyset$ und $\bar{\emptyset} = \Omega$ sowie $\overline{\bar{A}} = A$. Für die Restmenge gilt:

$$B \setminus A = B \cap \bar{A}$$

Für je zwei Mengen $A, B \subseteq \Omega$ gilt

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Allgemein: Für alle $A, B, C \subseteq \Omega$ gilt

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

Übung: Illustrieren und begründen Sie sorgfältig jede dieser Regeln.

Einfache Beispiele zu DeMorgan

Wie zuvor sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ sowie $A = \{2, 4, 6\}$ und $B = \{5, 6\}$.

Aufgabe: Führen Sie an diesem Beispiel die DeMorgan-Regeln aus.

Lösung: Wir bestimmen Schnitt, Vereinigung und ihre Komplemente.

Das Komplement der Vereinigung ist der Schnitt der Komplemente:

$$\overline{A \cup B} = \overline{\{2, 4, 5, 6\}} = \{1, 3\}$$

$$\overline{A} \cap \overline{B} = \{1, 3, 5\} \cap \{1, 2, 3, 4\} = \{1, 3\}$$

Das Komplement des Schnittes ist die Vereinigung der Komplemente:

$$\overline{A \cap B} = \overline{\{6\}} = \{1, 2, 3, 4, 5\}$$

$$\overline{A} \cup \overline{B} = \{1, 3, 5\} \cup \{1, 2, 3, 4\} = \{1, 2, 3, 4, 5\}$$

😊 Diese Rechenregeln sind sehr einfach, sehr häufig und sehr nützlich. Sie schützen vor ärgerlichen Rechenfehlern und Trugschlüssen!

⚠ Wir sehen insbesondere $\overline{A \cup B} \neq \overline{A} \cup \overline{B}$ und $\overline{A \cap B} \neq \overline{A} \cap \overline{B}$. Nutzen Sie die mathematische Notation mit Sinn und Verstand!

Rechnen mit Ereignissen

Alle Ergebnisse fassen wir zur **Ergebnismenge** Ω zusammen. Jedes Element $\omega \in \Omega$ heißt **Ergebnis**. Jede Teilmenge $A \subseteq \Omega$ heißt **Ereignis**.

Menge	Bedeutung als Zufallsereignis
Ω	Das sichere Ereignis: Ω tritt immer ein, $\mathbf{P}(\Omega) = 1$.
\emptyset	Das unmögliche Ereignis: \emptyset tritt nie ein, $\mathbf{P}(\emptyset) = 0$.
$A \subseteq \Omega$	Das Ereignis A tritt ein bei jedem Ergebnis $\omega \in A$.
$\Omega \setminus A$	Komplement: $\bar{A} = \Omega \setminus A$ tritt ein, wenn A nicht eintritt.
$A \subseteq B$	Teilmenge: Immer wenn A eintritt, dann auch B .
$B \setminus A$	Restmenge: Das Ereignis B tritt ein, aber nicht A .
$A \cap B$	Die Ereignisse A und B treten beide ein.
$A \cup B$	Das Ereignis A oder B tritt ein (evtl. auch beide).
$A \triangle B$	Entweder A oder B tritt ein (aber nicht beide).
$A \sqcup B$	Das Ereignis A oder B tritt ein, wobei $A \cap B = \emptyset$.

Unser Ziel sind nachvollziehbar begründete, quantitative Aussagen.



Diese Notation erlaubt präzises Rechnen mit Zufallsereignissen.

Anwendungsbeispiel

Sie würfeln mit unserem gezinkten Würfel, $p: \{1, 2, 3, 4, 5, 6\} \rightarrow \mathbb{R}$ mit

$$p(1) = 0.15, \quad p(2) = p(3) = p(4) = p(5) = 0.16, \quad p(6) = 0.21$$

Aufgabe: Mit welcher Wahrscheinlichkeit ergibt unser gezinkter Würfel entweder eine gerade Zahl oder mindestens 5? (aber nicht beides!)

Lösung: Das Ereignis „gerade Zahl“ ist die Teilmenge $A = \{2, 4, 6\}$. Entsprechend ist das Ereignis „mindestens 5“ die Teilmenge $B = \{5, 6\}$. „Entweder A oder B “ entspricht der symmetrischen Differenz:

$$A \Delta B = (A \cup B) \setminus (A \cap B) = \{2, 4, 5, 6\} \setminus \{6\} = \{2, 4, 5\}.$$

Die Wkt berechnen wir durch Aufsummieren der Elementarwkten:

$$\mathbf{P}(A \Delta B) = p(2) + p(4) + p(5) = 0.16 + 0.16 + 0.16 = 0.48$$

Interpretation: Dieses Ereignis tritt in 48% aller Fälle ein.

😊 Bequeme Notation und präzises Rechnen mit Zufallsereignissen. Übersetzung von Prosa zur Formel und zurück zur Interpretation.

Definition T1c: diskrete Wahrscheinlichkeitsverteilung

Eine **diskrete Wahrscheinlichkeitsverteilung** ist eine Abbildung

$$p : \Omega \rightarrow [0, 1] : \omega \mapsto p(\omega) \quad \text{mit} \quad \sum_{\omega \in \Omega} p(\omega) = 1.$$

Siehe Definition A2H; hierbei hat p höchstens abzählbaren Träger.
Für jedes **Ereignis** $A \subseteq \Omega$ summieren sich die Elementarwkten:

$$\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1] : A \mapsto \mathbf{P}(A) = \sum_{\omega \in A} p(\omega).$$

Ein solches Paar (Ω, \mathbf{P}) heißt **diskreter Wahrscheinlichkeitsraum**.

Jedem Ergebnis $\omega \in \Omega$ wird seine Elementarwahrscheinlichkeit $p(\omega) \in [0, 1]$ zugeordnet.

Beispiele: Für den fairen Würfel gilt $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$,
für unseren gezinkten $p(1) = 0.15$, $p(2) = p(3) = p(4) = p(5) = 0.16$, $p(6) = 0.21$.

Weitere typische Beispiele: Wir betrachten später die Verteilungen $p : \mathbb{N} \rightarrow [0, 1]$ mit
Elementarwkten $p(k) = \binom{n}{k} t^k (1-t)^{n-k}$ oder $p(k) = e^{-\lambda} \lambda^k / k!$ oder $p(k) = q^k (1-q)$.

Es handelt sich hierbei jeweils um die Binomialverteilung $B(n, t)$ mit $n \in \mathbb{N}$ und $t \in [0, 1]$,
die Poisson-Verteilung $P(\lambda)$ mit $\lambda \geq 0$, und die geometrische Verteilung $G(q)$ mit $q \in [0, 1]$.
Diese erfüllen $p(k) \geq 0$ und $\sum_{\omega \in \Omega} p(\omega) = 1$, wie gefordert. Versuchen Sie es als Übung!

Wahrscheinlichkeit von Ereignissen

Es gilt die **Normierung** $\mathbf{P}(\Omega) = 1$ sowie $\mathbf{P}(\emptyset) = 0$. Außerdem gilt

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

Bei disjunkter Vereinigung addieren sich die Wkten: Aus $A \cap B = \emptyset$ folgt

$$\mathbf{P}(A \sqcup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

Per Induktion folgt aus paarweiser die **endliche Additivität**:

$$\mathbf{P}(A_1 \sqcup A_2 \sqcup \dots \sqcup A_n) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots + \mathbf{P}(A_n)$$

Bei diskreten Wahrscheinlichkeiten gilt sogar **allgemeine Additivität**:

$$\mathbf{P}\left(\bigsqcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbf{P}(A_i).$$

Insbesondere gilt $\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\{\omega\})$ für jede Teilmenge $A \subseteq \Omega$.

Hier bedeutet $\bigsqcup_i A_i = \bigcup_i A_i$ mit $A_i \cap A_j = \emptyset$ für $i \neq j$. Die Indexmenge I ist beliebig. Siehe Definition A2H. Bei endlichem Wert ist der Träger abzählbar, wir haben also eine Reihe. Für kontinuierliche Wahrscheinlichkeitsräume nutzen wir später Integrale statt Summen. Zudem beschränken wir uns dann (aus unumgänglichen technischen Gründen) auf höchstens abzählbare Vereinigungen. Einstweilen ist alles noch diskret und somit besonders einfach.

Diskrete Wahrscheinlichkeitsräume

Auch die Zuordnung $A \mapsto \mathbf{P}(A)$ wollen wir als Abbildung auffassen. Wir fassen alle Ereignisse $A \subseteq \Omega$ zur **Ereignismenge** \mathcal{A} zusammen. Zur gegebenen **Ergebnismenge** Ω ist dies die **Potenzmenge** U207

$$\mathcal{A} = \mathfrak{P}(\Omega) := \{ A \mid A \subseteq \Omega \}.$$

Definition T1D: diskreter Wahrscheinlichkeitsraum

Ein **diskretes WMaß** ist eine Abbildung $\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1]$ für die gilt:

- 1 Normierung: Es gilt $\mathbf{P}(\Omega) = 1$ für die Gesamtmenge Ω .
- 2 Additivität: Es gilt $\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\{\omega\})$ für jedes $A \subseteq \Omega$.

Ein **diskreter Wahrscheinlichkeitsraum** (Ω, \mathbf{P}) besteht aus

- einer (nicht-leeren) Menge Ω von möglichen Ergebnissen und
- einem diskreten WMaß $\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1]$ wie oben erklärt.

Diese einfache Definition fasst unsere bisherigen Überlegungen zusammen. Sie ist ausreichend flexibel für viele Anwendungen und legt einen präzisen und bequemen Sprachgebrauch fest. Bei diskreten WRäumen ist \mathbf{P} bestimmt durch die Elementarwkten $p(\omega) := \mathbf{P}(\{\omega\})$ für $\omega \in \Omega$. Neben diskreten werden wir später auch kontinuierliche Wahrscheinlichkeitsmaße kennenlernen.

Rechenregeln für Wahrscheinlichkeitsmaße

0 Unmögliches Ereignis:

$$\mathbf{P}(\emptyset) = 0$$

1 Sicheres Ereignis:

$$\mathbf{P}(\Omega) = 1$$

2 Komplement:

$$\mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A)$$

3 Monotonie:

$$A \subseteq B \implies \mathbf{P}(A) \leq \mathbf{P}(B)$$

4 Restmenge:

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

5 Vereinigung:

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

6 Disjunkte Vereinigung:

$$\mathbf{P}(A \sqcup B) = \mathbf{P}(A) + \mathbf{P}(B) \quad \text{dank} \quad A \cap B = \emptyset$$

Definition T1E: endlicher WRaum und Gleichverteilung

Ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) heißt **endlich**, wenn die Ergebnismenge Ω endlich ist, und **laplacesch**, wenn zudem alle Ergebnisse $\omega \in \Omega$ gleich wahrscheinlich sind:

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

Für dieses **Laplace–Wahrscheinlichkeitsmaß** auf Ω gilt daher

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl aller Ergebnisse in } \Omega} = \frac{\text{günstige Ergebnisse}}{\text{mögliche Ergebnisse}}.$$

Dies nennen wir auch die (diskrete, endliche) **Gleichverteilung** auf Ω .

Beispiel: Der Wurf eines fairen Würfels ist ein Laplace–Experiment auf $\Omega = \{1, 2, \dots, 6\}$, der Wurf unseres gezinkten Würfels hingegen nicht. Roulette ist ein Laplace–Experiment auf der Menge $\Omega = \{0, 1, 2, \dots, 36\}$; hingegen sind hierbei $\{\text{rot, schwarz, zéro}\}$ nicht gleichwahrscheinlich.

Zweimaliger Wurf eines fairen Würfels

Aufgabe: Sie würfeln zweimal. Beschreiben Sie dies als WRaum.

Lösung: Als Ergebnismenge Ω betrachten wir hierzu

$$\Omega = \left\{ \begin{array}{cccccc} (1, 1), & (1, 2), & (1, 3), & (1, 4), & (1, 5), & (1, 6), \\ (2, 1), & (2, 2), & (2, 3), & (2, 4), & (2, 5), & (2, 6), \\ (3, 1), & (3, 2), & (3, 3), & (3, 4), & (3, 5), & (3, 6), \\ (4, 1), & (4, 2), & (4, 3), & (4, 4), & (4, 5), & (4, 6), \\ (5, 1), & (5, 2), & (5, 3), & (5, 4), & (5, 5), & (5, 6), \\ (6, 1), & (6, 2), & (6, 3), & (6, 4), & (6, 5), & (6, 6) \end{array} \right\}.$$

Wir nehmen an, beide Würfe seien gleichverteilt und unabhängig.
Es handelt sich demnach bei (Ω, \mathbf{P}) um ein Laplace-Experiment!

Aufgabe: Welche Wkt hat die Augensumme 2, 3, 4, ..., 12?

Lösung: Wir untersuchen das Ereignis $A_k =$ „Die Augensumme ist k .“

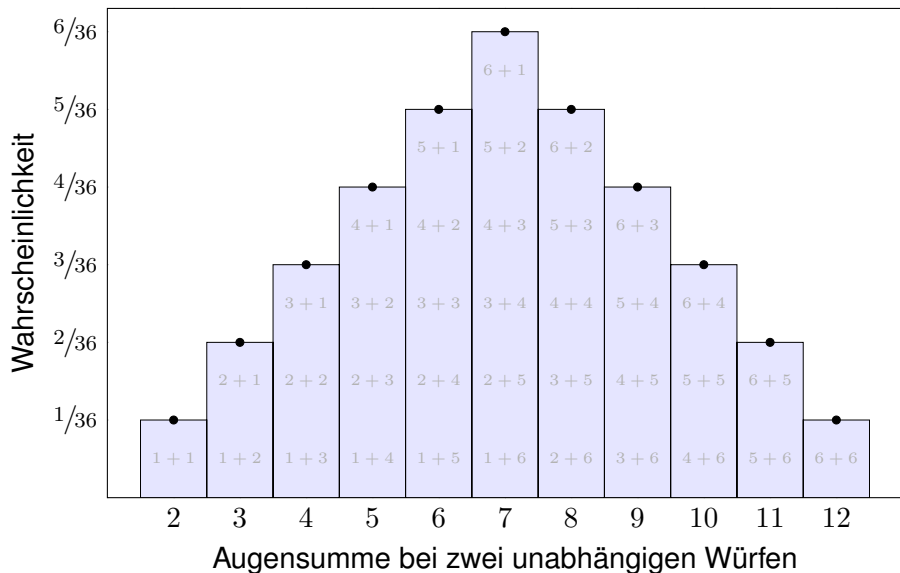
$$A_2 = \{(1, 1)\}, \quad \mathbf{P}(A_2) = |A_2|/|\Omega| = 1/36$$

$$A_3 = \{(1, 2), (2, 1)\}, \quad \mathbf{P}(A_3) = |A_3|/|\Omega| = 2/36$$

$$A_4 = \{(1, 3), (2, 2), (3, 1)\}, \quad \mathbf{P}(A_4) = |A_4|/|\Omega| = 3/36$$

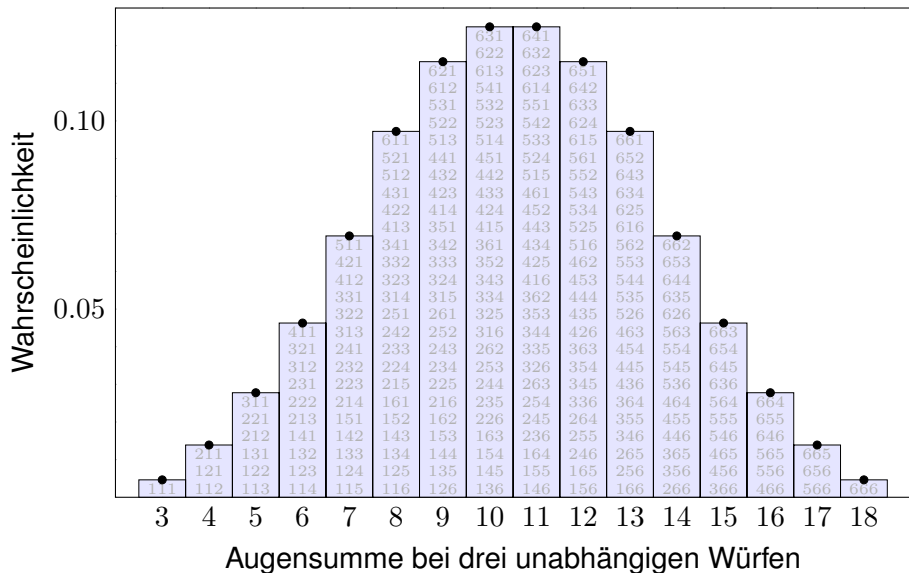
Zweimaliger Wurf eines fairen Würfels

Alle Ergebnisse $(i, j) \in \{1, 2, 3, 4, 5, 6\}^2$ sind gleichwahrscheinlich, die zugehörigen Augensummen $s = i + j$ hingegen nicht!



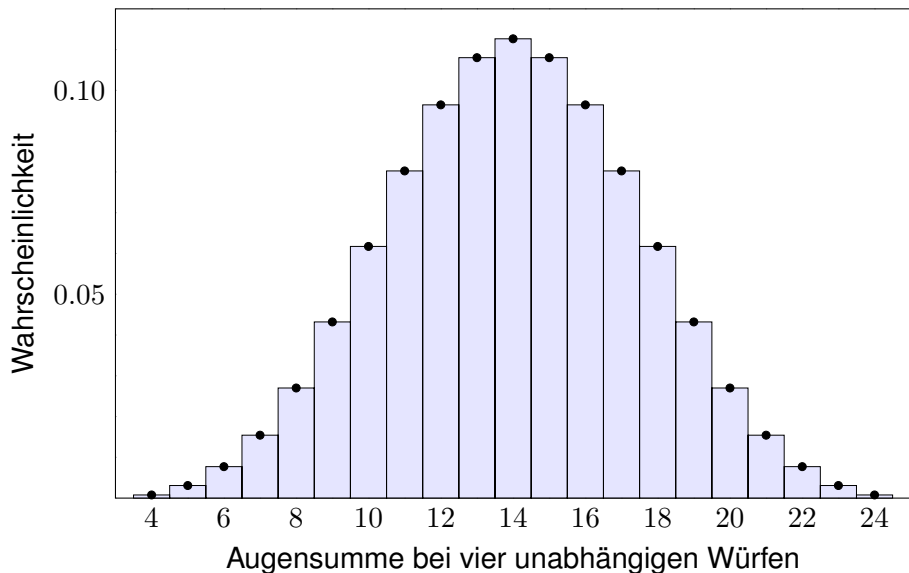
Dreimaliger Wurf eines fairen Würfels

Bei dreimaligem Würfeln sind alle Ergebnisse $(i, j, k) \in \{1, 2, 3, 4, 5, 6\}^3$ gleichwahrscheinlich, die Augensummen $s = i + j + k$ hingegen nicht!



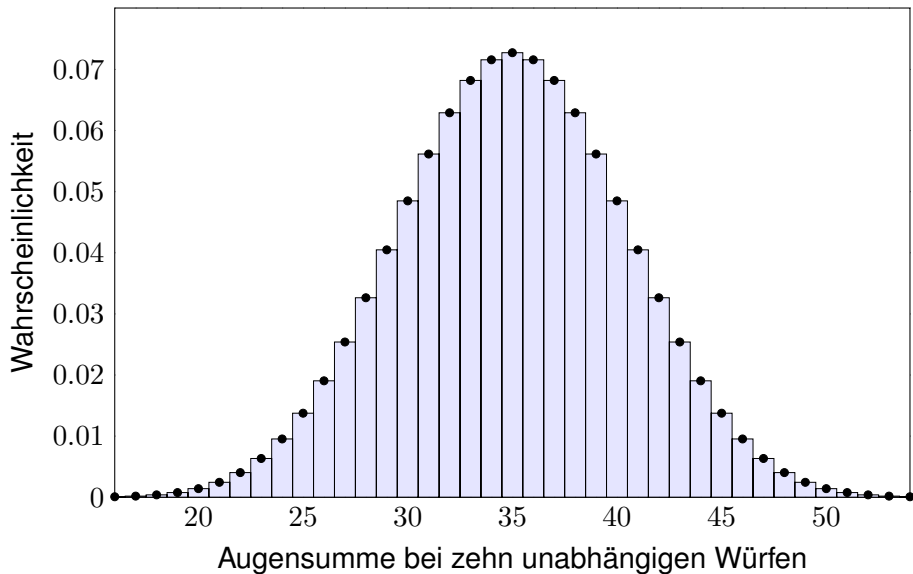
Viermaliger Wurf eines fairen Würfels

Die hier entstehende Glockenkurve wird uns noch viel nützen:
Sie illustriert den **zentralen Grenzwertsatz** W1D, kurz ZGS.



Zehnmaliger Wurf eines fairen Würfels

Kombinatorik (Kapitel U) und Asymptotik (Kapitel V & W): Hier entsteht in immer besserer Näherung die **gaußsche Glockenkurve**! Das hilft!



Kombinatorik der Augensumme beim Würfeln

Aufgabe: Auf wie viele Weisen erhalten Sie die Augensumme s bei einmaligem Würfeln? zweimaligem Würfeln? dreimaligem Würfeln?

Lösung: (1) Bei einmaligem Würfeln ist alles klar:

Augensumme s	1	2	3	4	5	6
Möglichkeiten n_s	1	1	1	1	1	1

Die Wkt $\mathbf{P}(S=s) = 1/6$ ist gleichverteilt. Wir kodieren sie als Polynom

$$G(z) = \frac{1}{6}z^1 + \frac{1}{6}z^2 + \frac{1}{6}z^3 + \frac{1}{6}z^4 + \frac{1}{6}z^5 + \frac{1}{6}z^6 = \frac{z - z^7}{6(1 - z)}.$$

(2) Bei zweimaligem Würfeln entspricht die Summe der Augenzahlen der Faltung der Verteilung (1) mit sich selbst, als Polynom also $G(z) \cdot G(z)$:

Augensumme s	2	3	4	5	6	7	8	9	10	11	12
Möglichkeiten n_s	1	2	3	4	5	6	5	4	3	2	1

Die Wkt ist jeweils $\mathbf{P}(S=s) = n_s/6^2$, also keine Gleichverteilung mehr!
Die Wkt ist am größten für die Augensumme 7, und zwar mit Wkt $1/6$.

Kombinatorik der Augensumme beim Würfeln

(3) Bei dreimaligem Würfeln müssen wir die Rechnung gut organisieren.

Beispiel: Auf wie viele Weisen kann die Augensumme 10 entstehen?

Je nach erstem Wurf $10 = 1 + 9 = 2 + 8 = 3 + 7 = 4 + 6 = 5 + 5 = 6 + 4$,

dank voriger Tabelle also auf $4 + 5 + 6 + 5 + 4 + 3 = 27$ Möglichkeiten

Die Wkt ist demnach $\mathbf{P}(S=10) = 27/6^3 = 1/8 = 0.125$.

Nach diesem Schema berechnen wir die folgende Tabelle: Die Summe der Augenzahlen entspricht der Faltung der Verteilungen (1) und (2).

(Hier kann eine Tabellenkalkulation helfen, zum Beispiel *LibreOffice*.)

Augensumme s	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Möglichkeiten n_s	1	3	6	10	15	21	25	27	27	25	21	15	10	6	3	1

Die Wkt ist jeweils $\mathbf{P}(S=s) = n_s/6^3$, wie oben graphisch dargestellt.

Wie zu erwarten ist die Wkt am größten für die mittleren Augensummen 10 und 11, und zwar jeweils mit Wkt $1/8$. Jetzt können Sie's ausrechnen!

😊 Als Polynom entspricht dies dem Produkt $G(z)^3 = G(z) \cdot G(z) \cdot G(z)$.

Die Wkten lesen Sie als Koeffizienten von $G(z)^3 = \sum_s \mathbf{P}(S=s) z^s$ ab.

Die Kodierung als **erzeugende Funktion** (U4D) ist elegant und effizient.

Kombinatorik der Summe beim fairen Münzwurf

Würfelexperimente sind schön anschaulich, die Rechnungen sind nicht ganz banal und illustrieren bereits viele wichtige Phänomene.

Münzwürfe sind wesentlich leichter und als Modell noch wichtiger: Wir wiederholen ein Experiment n -mal und zählen die Treffer.

Aufgabe: Sie werfen n -mal eine faire Münze: 0 = Kopf, 1 = Zahl. Die Summe der Treffer bezeichnen wir mit $S_n \in \{0, 1, \dots, n\}$.

(1) Berechnen Sie die Wahrscheinlichkeit $\mathbf{P}(S_n=k)$ für alle $0 \leq k \leq n$.

(2) Berechnen Sie die erzeugende Funktion $G_n(z) = \sum_k \mathbf{P}(S_n=k) z^k$.

(3) Skizzieren Sie die Verteilung der Zufallsvariablen $S_1, S_2, S_3, S_4, \dots$

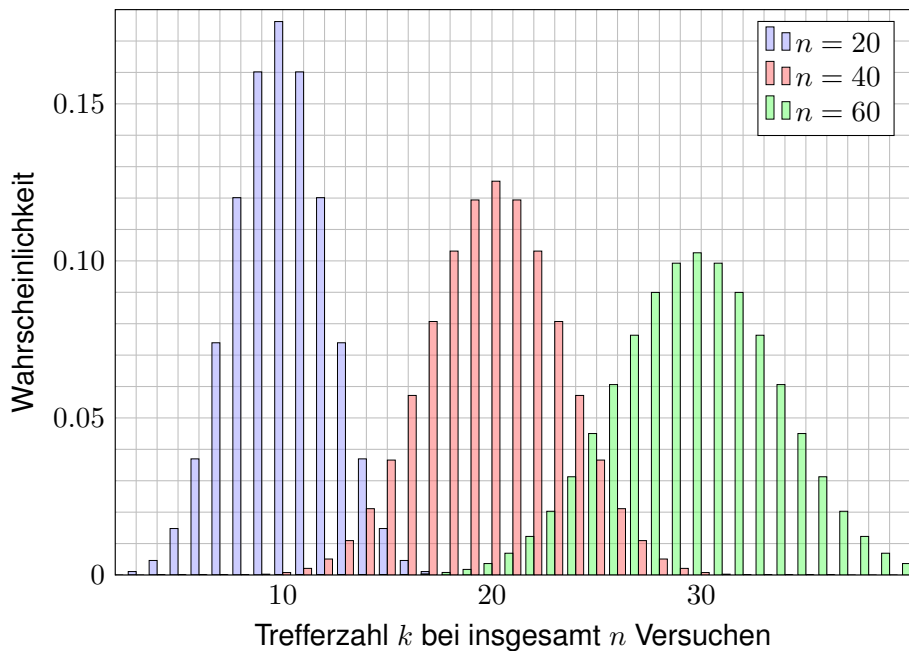
Lösung: (1) Für unabhängig wiederholte faire Münzwürfe finden wir:

$$\mathbf{P}(S_n=k) = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

😊 Die Binomialverteilung $\binom{n}{k} t^k (1-t)^{n-k}$ ist ein wichtiges Modell. Wir werden sie in den nächsten Kapiteln ausführlich untersuchen.

(2) Wir finden das Polynom $G_n(z) = \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{2}\right)^n z^k = \left(\frac{1}{2}z + \frac{1}{2}\right)^n$.

Kombinatorik der Summe beim fairen Münzwurf



Bedingte Wahrscheinlichkeit

Häufig stehen Sie vor der Aufgabe, nach Eintritt eines Ereignisses B auf die Wahrscheinlichkeit eines anderen Ereignisses A zu schließen.

Aufgabe: Angenommen, beim Würfeln wurde nicht die 6 gewürfelt. Wie groß ist dann noch die Wahrscheinlichkeit einer geraden Zahl?

Lösung: Wir setzen voraus, das Ereignis $B = \{1, 2, 3, 4, 5\}$ tritt ein. Zum Ereignis $A = \{2, 4, 6\}$ suchen wir die bedingte Wkt $\mathbf{P}_B(A)$.

Ergebnis $\omega =$	1	2	3	4	5	6
ursprüngliche Wkt $\mathbf{P}(\{\omega\}) =$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
bedingte Wkt $\mathbf{P}_B(\{\omega\}) =$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0

Die gesuchte Wkt ist demnach $\mathbf{P}_B(A) = 1/5 + 1/5 = 2/5 = 40\%$.

Interpretation: Wir würfeln sehr häufig und werten nur die Fälle, die die Bedingung B erfüllen. In $\frac{2}{5}$ dieser Fälle erhalten wir eine gerade Zahl. In den restlichen $\frac{3}{5}$ dieser Fälle erhalten wir eine ungerade Zahl.

Bedingte Wahrscheinlichkeit

Wie zuvor erklärt, beginnen wir mit einem stochastischen Modell, also einem Wahrscheinlichkeitsraum (Ω, \mathbf{P}) gemäß Definition T1c.

Definition T2A: bedingte Wahrscheinlichkeit

Sei (Ω, \mathbf{P}) ein Wahrscheinlichkeitsraum und $B \subseteq \Omega$ ein Ereignis mit Wahrscheinlichkeit $\mathbf{P}(B) > 0$. Für jedes Ereignis $A \subseteq \Omega$ definieren wir die **Wahrscheinlichkeit von A unter der Bedingung B** als Quotient

$$\mathbf{P}(A|B) := \mathbf{P}_B(A) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Beispiel: In der vorigen Aufgabe finden wir

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(\{2, 4\})}{\mathbf{P}(\{1, 2, 3, 4, 5\})} = \frac{2/6}{5/6} = \frac{2}{5} = 40\%.$$

- 😊 Bequeme Notation und präzises Rechnen mit Zufallsereignissen.
- 😊 Die Rechnung ist meist leicht, die Interpretation erfordert Übung.
- 😊 Die Definition T2A stellt eindeutig klar, was gemeint ist.

Rechenregeln für bedingte Wahrscheinlichkeiten

Aufgabe: Ist $\mathbf{P}_B : \mathfrak{P}(\Omega) \rightarrow [0, 1]$ selbst ein (diskretes) WMaß?

Lösung: Ja, \mathbf{P}_B erfüllt die beiden definierenden Eigenschaften (T1D):

(1) Normierung: Es gilt $\mathbf{P}_B(\Omega) = \mathbf{P}(\Omega \cap B) / \mathbf{P}(B) = \mathbf{P}(B) / \mathbf{P}(B) = 1$.

(2) Additivität: Für jedes Ereignis $A \subseteq \Omega$ gilt

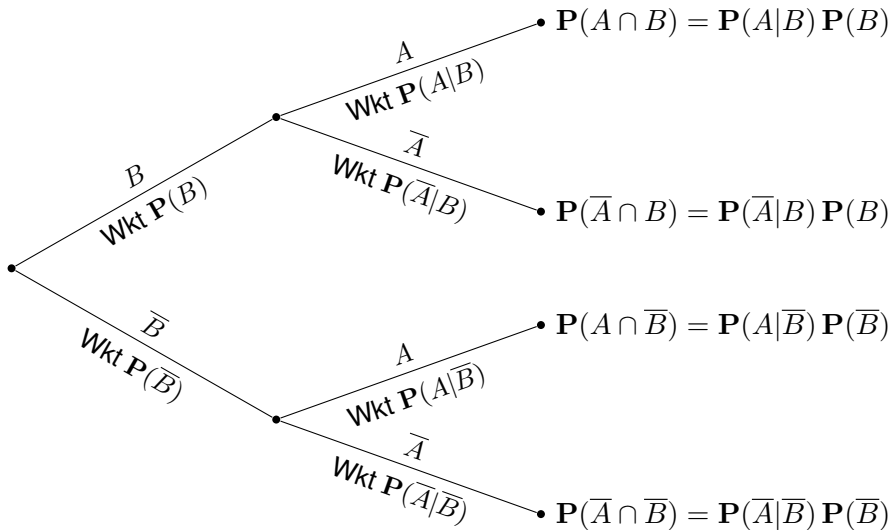
$$\begin{aligned}\mathbf{P}_B(A) &= \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\sum_{\omega \in A} \mathbf{P}(\{\omega\} \cap B)}{\mathbf{P}(B)} \\ &= \sum_{\omega \in A} \frac{\mathbf{P}(\{\omega\} \cap B)}{\mathbf{P}(B)} = \sum_{\omega \in A} \mathbf{P}_B(\{\omega\}).\end{aligned}$$

Gleiches gilt für (abzählbare) disjunkte Vereinigungen $A = \bigsqcup_{i \in I} A_i$:

$$\begin{aligned}\mathbf{P}_B(A) &= \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\sum_{i \in I} \mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} \\ &= \sum_{i \in I} \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} = \sum_{i \in I} \mathbf{P}_B(A_i)\end{aligned}$$

Aus diesen beiden Eigenschaften folgen alle weiteren Rechenregeln.

Baummodell und Pfadfinderregeln

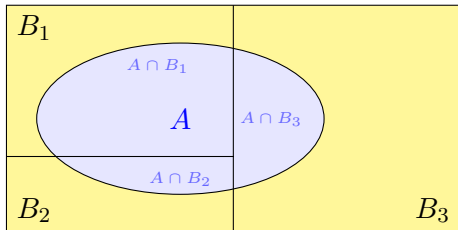


Produktregel: Die Wkt jedes Blattes ist das Produkt der bedingten Wkten entlang des Pfades.

Summenregel: Die Wkt eines Ereignisses A ist die Summe der Wkten aller Pfade, die zu A führen. Implizite Voraussetzung: Diese Zerlegung muss ganz A überdecken und disjunkt sein!

Die Formel der totalen Wahrscheinlichkeit

Sei $\Omega = B_1 \sqcup B_2 \sqcup \dots \sqcup B_\ell$ eine **Zerlegung** in disjunkte Ereignisse.
 Das bedeutet, es gilt $\Omega = B_1 \cup B_2 \cup \dots \cup B_\ell$ mit $B_i \cap B_j = \emptyset$ für $i \neq j$.
 Dies zerlegt jedes Ereignis $A \subseteq \Omega$ in disjunkte Teilmengen $A_k = A \cap B_k$:



$$A = (A \cap B_1) \sqcup (A \cap B_2) \sqcup \dots \sqcup (A \cap B_\ell)$$

Hieraus folgt die **Formel der totalen Wahrscheinlichkeit**:

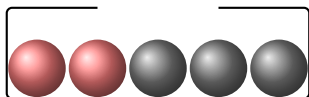
$$\mathbf{P(A) = \sum_{k=1}^{\ell} P(A \cap B_k) = \sum_{k=1}^{\ell} P(A|B_k) P(B_k)}$$

Diese Formel entspricht dem Baumdiagramm für $\Omega = B_1 \sqcup B_2 \sqcup \dots \sqcup B_\ell$.

😊 Sie gilt sogar für $P(B_k) = 0$, wobei wir $P(A|B_k)P(B_k) = 0$ setzen.

😊 Strategie: Zerlegung des Problems in einfachere Teilprobleme!

Beispiel zur totalen Wahrscheinlichkeit



Aufgabe: Aus einer Urne mit zwei roten und drei schwarzen Kugeln ziehen wir nacheinander zufällig zwei Kugeln ohne Zurücklegen. Mit welcher Wkt haben beide die gleiche Farbe?

Lösung: Sei $G = \{\text{Beide Kugeln haben die gleiche Farbe}\}$ sowie

$$R = \{\text{Erste Kugel ist rot}\} \quad \text{und} \quad S = \{\text{Erste Kugel ist schwarz}\}.$$

Dies ist eine Zerlegung, denn $R \cap S = \emptyset$ und $R \cup S = \Omega$. Wir wissen:

$$\mathbf{P}(R) = \frac{2}{5}, \quad \mathbf{P}(S) = \frac{3}{5}, \quad \mathbf{P}(G|R) = \frac{1}{4}, \quad \mathbf{P}(G|S) = \frac{2}{4}$$

Hieraus berechnen wir mühelos die gesuchte totale Wahrscheinlichkeit:

$$\mathbf{P}(G) = \mathbf{P}(G|R) \mathbf{P}(R) + \mathbf{P}(G|S) \mathbf{P}(S) = \frac{1}{4} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{3}{5} = \frac{2+6}{20} = \frac{2}{5}$$

😊 Bedingte Wkten strukturieren und vereinfachen die Rechnung!

Die Formel von Bayes: Es lebe der Dreisatz!

Erinnerung: Für alle Ereignisse $A, B \subseteq \Omega$ mit $\mathbf{P}(A), \mathbf{P}(B) > 0$ gilt

$$\mathbf{P}(A|B) \mathbf{P}(B) \stackrel{\text{Def}}{=} \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \mathbf{P}(B) = \mathbf{P}(A \cap B),$$

$$\mathbf{P}(B|A) \mathbf{P}(A) \stackrel{\text{Def}}{=} \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} \mathbf{P}(A) = \mathbf{P}(A \cap B).$$

Durch Umstellen gewinnen wir hieraus die **Formel von Bayes**:

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

😊 Dies erlaubt das „Umkehren von Bedingung und Schlussfolgerung“, zwar nicht im kausalen Sinne, aber immerhin im stochastischen Sinne.

😊 Typische Anwendung: Wir kennen die Wkten $\mathbf{P}(A)$ und $\mathbf{P}(B)$, die Bestimmung von $\mathbf{P}(A|B)$ ist leicht, aber gesucht ist $\mathbf{P}(B|A)$.

⚠ Um durch $\mathbf{P}(A)$ zu dividieren, müssen wir $\mathbf{P}(A) > 0$ voraussetzen. Die Formel gilt auch für $\mathbf{P}(B) = 0$; wir setzen dann $\mathbf{P}(A|B) \mathbf{P}(B) = 0$.

😊 Anschaulich in obiger Graphik zu $A = (A \cap B_1) \sqcup \dots \sqcup (A \cap B_\ell)$.

Die Formel von Bayes: Zusammenfassung

Soweit ist alles klar und einfach, die Formel von Bayes ist sehr leicht. In vielen Anwendungen entfaltet sie eine enorme Durchschlagskraft.

Satz T2B: Bayes 1763

Sei (Ω, \mathbf{P}) ein Wahrscheinlichkeitsraum und $\Omega = B_1 \sqcup B_2 \sqcup \dots \sqcup B_\ell$ eine Zerlegung, also $\Omega = B_1 \cup B_2 \cup \dots \cup B_\ell$ mit $B_i \cap B_j = \emptyset$ für $i \neq j$.

(1) Für alle $A \subseteq \Omega$ gilt dann die Formel der **totalen Wahrscheinlichkeit**:

$$\mathbf{P}(A) = \sum_{k=1}^{\ell} \mathbf{P}(A \cap B_k) = \sum_{k=1}^{\ell} \mathbf{P}(A|B_k) \mathbf{P}(B_k)$$

(2) Im Falle $\mathbf{P}(A) > 0$ gilt zudem die **Formel von Bayes**:

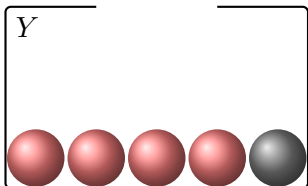
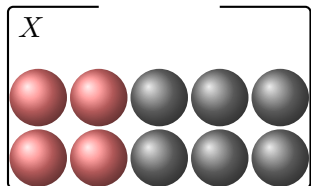
$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(A|B_i) \mathbf{P}(B_i)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B_i) \mathbf{P}(B_i)}{\sum_{k=1}^{\ell} \mathbf{P}(A|B_k) \mathbf{P}(B_k)}$$

😊 Alles gilt wörtlich genauso für abzählbare Zerlegungen, also $\ell = \infty$.

😊 Bequeme Notation und präzises Rechnen mit Zufallsereignissen.

⚠ Die Rechnung ist meist leicht, die Interpretation erfordert Übung.

Anwendungsbeispiel: ein Urnenmodell



Urne X : 4 rote und 6 schwarze; Urne Y : 4 rote und 1 schwarze Kugel.
Wir wählen zufällig eine Urne X, Y und ziehen daraus eine Kugel.

Aufgabe: (1) Welche Wkt hat das Ereignis $R = \{\text{rote Kugel}\}$?

(2) Es wurde eine rote Kugel gezogen. Mit welcher Wkt aus Urne X ?

Lösung: Dank Bayes (Satz T2B) ist die Rechnung leicht:

$$(1) \quad \mathbf{P}(R) = \mathbf{P}(R|X) \mathbf{P}(X) + \mathbf{P}(R|Y) \mathbf{P}(Y) = \frac{4}{10} \cdot \frac{1}{2} + \frac{4}{5} \cdot \frac{1}{2} = \frac{3}{5}$$

$$(2) \quad \mathbf{P}(X|R) = \frac{\mathbf{P}(R|X) \mathbf{P}(X)}{\mathbf{P}(R)} = \frac{\frac{4}{10} \cdot \frac{1}{2}}{\frac{3}{5}} = \frac{1}{5} \cdot \frac{5}{3} = \frac{1}{3}$$

😊 Die Rechnung ist Routine, doch die Interpretation erfordert Übung.

Anwendungsbeispiel: Bestehen von Prüfungen

Zu einer Prüfung treten drei Typen von Teilnehmer:innen an:

- 10% sind gut vorbereitet und haben Erfolgswkt 100% (Typ A),
- 50% sind etwas vorbereitet und haben Erfolgswkt 60% (Typ B),
- 40% sind schlecht vorbereitet und haben Erfolgswkt 25% (Typ C).

Aufgabe: (1) Wie groß ist der Anteil der erfolgreichen Teilnehmer E ?
(2) Wie groß ist darunter jeweils der Anteil der Typen A, B, C ?

Lösung: (0) Die Gesamtmenge wird disjunkt zerlegt in $\Omega = A \sqcup B \sqcup C$.
(1) Nach der Formel für die totale Wahrscheinlichkeit (T2B) erhalten wir:

$$\begin{aligned}\mathbf{P}(E) &= \mathbf{P}(E|A) \mathbf{P}(A) + \mathbf{P}(E|B) \mathbf{P}(B) + \mathbf{P}(E|C) \mathbf{P}(C) \\ &= 1.0 \cdot 0.1 \quad + 0.6 \cdot 0.5 \quad + 0.25 \cdot 0.4 \quad = 0.5\end{aligned}$$

(2) Nach Bayes (T2B) berechnen wir hieraus folgende Verteilung auf E :

$$\mathbf{P}(A|E) = \mathbf{P}(E|A) \mathbf{P}(A) / \mathbf{P}(E) = 0.1/0.5 = 0.2$$

$$\mathbf{P}(B|E) = \mathbf{P}(E|B) \mathbf{P}(B) / \mathbf{P}(E) = 0.3/0.5 = 0.6$$

$$\mathbf{P}(C|E) = \mathbf{P}(E|C) \mathbf{P}(C) / \mathbf{P}(E) = 0.1/0.5 = 0.2$$

Die Verteilung ändert sich von $(10, 50, 40)$ auf Ω zu $(20, 60, 20)$ auf E .

Anwendungsbeispiel: Test auf eine Infektion

Eine bestimmte Infektion tritt bei 2% aller Personen (zufällig) auf. Sie erfordert frühzeitige Erkennung und geeignete Maßnahmen. Ein Testverfahren für diese Infektion liefert ein positives Ergebnis bei 98% aller kranken, aber auch bei 6% aller gesunden Personen.

Aufgabe: (1) Welche Wkt haben positive Tests? (2) Wie groß ist die Wkt, dass eine positiv getestete Person krank ist? gesund ist?

Lösung: (0) Für das Ereignis $K = \{\text{krank}\}$ wissen wir $\mathbf{P}(K) = 0.02$. Für $P = \{\text{positiver Test}\}$ gilt $\mathbf{P}(P|K) = 0.98$ und $\mathbf{P}(P|\bar{K}) = 0.06$.

(1) Hieraus folgt zunächst die totale Wahrscheinlichkeit

$$\begin{aligned}\mathbf{P}(P) &= \mathbf{P}(P|K) \mathbf{P}(K) + \mathbf{P}(P|\bar{K}) \mathbf{P}(\bar{K}) \\ &= 0.98 \cdot 0.02 + 0.06 \cdot 0.98 = 0.0196 + 0.0588 = 0.0784 \approx 8\%\end{aligned}$$

(2) Die gesuchte Wahrscheinlichkeit $\mathbf{P}(K|P)$ erhalten wir mit Bayes:

$$\mathbf{P}(K|P) = \frac{\mathbf{P}(P|K) \mathbf{P}(K)}{\mathbf{P}(P)} = \frac{0.0196}{0.0784} = 0.25 = 25\%$$

Komplementär $\mathbf{P}(\bar{K}|P) \approx 75\%$. In diesem Beispiel ist eine positiv getestete Person mit größerer Wahrscheinlichkeit gesund als krank!

Anwendungsbeispiel: Test auf eine Infektion

2% krank

98% gesund

Sensitivität des Tests: 98% = Richtig-Positiv-Rate

99,96%

Wissen macht Ah! Unwissen macht Aua! Unkenntnis führt zu Fehleinschätzung, schlimmer: zu Fehlentscheidungen.

Ist unser Ergebnis plausibel? Naiv halten Sie den Test für zuverlässig und vertrauen dem Testergebnis. Das Problem ist hier, dass die Infektion eher selten ist; die allermeisten der getesteten Personen sind gesund, und hier hat der Test leider eine recht hohe Falsch-Positiv-Rate. Falsch-positive Testergebnisse kommen daher recht häufig vor.

Diese *qualitative Vorüberlegung* erklärt das Problem sehr schön anschaulich. Unsere Rechnung liefert zudem eine *quantitative Aussage*. Obwohl die Infektion mit 2% recht selten ist, werden etwa 8% aller Personen positiv getestet: Von den 98% gesunden werden 6% falsch positiv getestet!

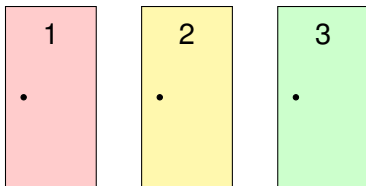
Ein positiver Test ist daher nur ein erstes Indiz, dass die Infektion vorliegen *könnte*: Die Wahrscheinlichkeit, krank zu sein, steigt von 2% vor dem Test auf 25% nach einem positiven Test. Entsprechend sinkt die Wahrscheinlichkeit, gesund zu sein, von 98% vor dem Test auf 75% nach einem positiven Test. Daher wird eine Ärzt:in möglichst versuchen, den Verdacht durch weitere, unabhängige (und genauere) Untersuchungen zu erhärten oder auszuräumen.

5.88%

Spezifität des Tests: 94% = Richtig-Negativ-Rate

Das Ziegenproblem / *Monty Hall Problem*

Bei der Spielshow *Let's make a deal* wählt die Kandidatin zwischen drei Türen: Hinter einer ist ein Auto, hinter den beiden anderen sind Ziegen.



Die Kandidatin wählt eine Tür. Der Showmaster weiß, was hinter den Türen ist, und öffnet eine der beiden anderen Türen, immer mit Ziege. Er bietet der Kandidatin nun die Möglichkeit noch einmal zu wechseln.

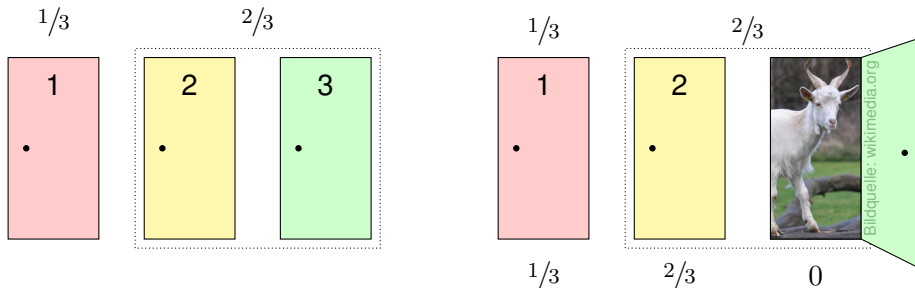
Aufgabe: Ist es vorteilhaft, die Tür zu wechseln?

Lösung: Unglaublich aber wahr: Ja! Es lohnt sich, zu wechseln!

Dies ist ein berühmtes Problem, zunächst paradox, dann lehrreich. Hierzu bieten sich drei Erklärungen: (1) Anschaulich. (2) Experimentell. (3) Ausführlich mit Bayes (Satz T2B). Versuchen Sie möglichst alle drei!

Das Ziegenproblem: anschauliche Erklärung

(1) Anschauliche Darstellung der Wahrscheinlichkeiten: Die Kandidatin wählt Tür 1, diese hat Gewinnwkt $\frac{1}{3}$, die beiden anderen zusammen $\frac{2}{3}$. Das bleibt auch so, nachdem der Moderator die/eine Ziegentür öffnet.



(2) Experimenteller Vergleich bei 3000 Spielshows:
Die Strategie „nie wechseln“ erzielt etwa 1000 Treffer (ein Drittel).
Die Strategie „immer wechseln“ erzielt die verbleibenden 2000 Treffer.

😊 Entgegen dem naiven ersten Eindruck ist es tatsächlich die bessere Strategie, die anfangs gewählte Tür im zweiten Schritt zu wechseln!
Stochastik (= Scharfsinn) hilft, mindert Risiken, optimiert Chancen.

Das Ziegenproblem: Anwendung von Bayes

Aufgabe: Lösen Sie das Ziegenproblem mit der Formel von Bayes.

Lösung: Wir betrachten die Ereignisse:

$A_1, A_2, A_3 =$ Das Auto steht hinter Tür 1, 2, 3

⚠ Annahme der Gleichverteilung: $\mathbf{P}(A_1) = \mathbf{P}(A_2) = \mathbf{P}(A_3) = 1/3$
Die Kandidatin wählt Tür 1. Wie kann der Moderator reagieren?

$M_1, M_2, M_3 =$ Der Moderator öffnet Tür 1, 2, 3

⚠ Annahme der Gleichverteilung auch bei der Türöffnung:

$$\begin{aligned} \mathbf{P}(M_2|A_1) &= 1/2, & \mathbf{P}(M_2|A_2) &= 0, & \mathbf{P}(M_2|A_3) &= 1, \\ \mathbf{P}(M_3|A_1) &= 1/2, & \mathbf{P}(M_3|A_2) &= 1, & \mathbf{P}(M_3|A_3) &= 0. \end{aligned}$$

😊 Beide Annahmen kann die Kandidatin erzwingen, indem er die drei Türen für seinen Gebrauch zufällig nummeriert, also *symmetrisiert*.
Der Moderator öffnet demnach Tür 2 oder Tür 3 jeweils mit Wkt $1/2$:

$$\begin{aligned} \mathbf{P}(M_3) &= \mathbf{P}(M_3|A_1)\mathbf{P}(A_1) + \mathbf{P}(M_3|A_2)\mathbf{P}(A_2) + \mathbf{P}(M_3|A_3)\mathbf{P}(A_3) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}. \end{aligned}$$

Das Ziegenproblem: Anwendung von Bayes

Angenommen der Moderator öffnet Tür 3. Was sollte die Kandidatin tun?
Die Formel von Bayes ergibt für die Gewinnwkt der Türen 1 und 2:

$$\mathbf{P}(A_1|M_3) = \frac{\mathbf{P}(M_3|A_1) \mathbf{P}(A_1)}{\mathbf{P}(M_3)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$\mathbf{P}(A_2|M_3) = \frac{\mathbf{P}(M_3|A_2) \mathbf{P}(A_2)}{\mathbf{P}(M_3)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

😊 Das entspricht der anschaulichen Erklärung, wie oben skizziert.
Es ist also tatsächlich vorteilhaft zu wechseln! Die Rechnung in allen anderen Fällen verläuft aus Symmetriegründen genauso.

Ein häufiger Einwand gegen diese Rechnung ist folgender: Was wäre, wenn der Moderator die beiden Türen ungleich öffnet? Sei allgemein $\mathbf{P}(M_2|A_1) = \alpha \in [0, 1]$ und $\mathbf{P}(M_3|A_1) = 1 - \alpha$. Wir haben oben vereinfachend $\alpha = 1/2$ angenommen. Die allgemeine Rechnung verläuft so: Zunächst folgt $\mathbf{P}(M_2) = (1 + \alpha)/3$ und $\mathbf{P}(M_3) = (2 - \alpha)/3$. Die obige Rechnung ergibt nun $P(A_1|M_3) = (1 - \alpha)/(2 - \alpha)$ und $P(A_2|M_3) = 1/(2 - \alpha)$. Es ist auch hier immer noch vorteilhaft zu wechseln! Der Sonderfall $\alpha = 0$ bedeutet: Hat der Moderator die Wahl, Tür 2 oder 3 zu öffnen, dann nimmt er immer Tür 3. Hier gilt Gleichstand $P(A_1|M_3) = P(A_2|M_3) = 1/2$. In allen verbleibenden Fällen *muss* der Moderator Tür 2 öffnen, und hier gilt $P(A_1|M_2) = 0$ gegen $P(A_3|M_2) = 1$. Über alle Fälle summiert bleibt die Schlussfolgerung bestehen!

Stochastisch un/abhängige Ereignisse

Wir betrachten weiterhin einen WRaum (Ω, \mathbf{P}) gemäß T1c. Zwei Ereignisse A und B sind **stochastisch unabhängig**, wenn die Wahrscheinlichkeit von A nicht von B abhängt:

$$\mathbf{P}(A|B) = \mathbf{P}(A)$$

Das setzt $\mathbf{P}(B) > 0$ voraus und bedeutet ausgeschrieben:

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

also $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$. Dies machen wir nun zur Definition:

Definition T2c: stochastisch unabhängige Ereignisse

Sei (Ω, \mathbf{P}) ein Wahrscheinlichkeitsraum und $A, B \subseteq \Omega$ zwei Ereignisse. Wir nennen A und B **(stochastisch) unabhängig** bezüglich \mathbf{P} , wenn

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B).$$

In Worten: Die Wkt des Schnitts $A \cap B$ ist das Produkt der Wkten. Andernfalls heißen die Ereignisse A und B **(stochastisch) abhängig**.

Stochastisch un/abhängige Ereignisse

Aufgabe: Sie würfeln einmal mit einem fairen Würfel.
Dies ist ein Laplace-Experiment auf $\Omega = \{1, 2, 3, 4, 5, 6\}$.

$$A := \{\text{gerade Zahl}\} = \{2, 4, 6\}$$

$$B := \{\text{höchstens } 5\} = \{1, 2, 3, 4, 5\}$$

Sind A und B stochastisch unabhängig?

Lösung: Wir finden zunächst $A \cap B = \{2, 4\}$.

Es gilt $\mathbf{P}(A) = \frac{1}{2}$ und $\mathbf{P}(B) = \frac{5}{6}$, aber $\mathbf{P}(A \cap B) = \frac{1}{3} \neq \frac{1}{2} \cdot \frac{5}{6}$.

😊 Die Ereignisse A und B sind demnach **stochastisch abhängig**.

😊 Anschaulich plausibel. Definition T2c sagt genau, was zu prüfen ist.

Quantitativer Vergleich: Es gilt $\mathbf{P}(A|B) = \frac{2}{5} < \frac{1}{2} = \mathbf{P}(A)$ und $\mathbf{P}(B|A) = \frac{2}{3} < \frac{5}{6} = \mathbf{P}(B)$.
Wir sprechen von einer negativen Korrelation: Das Eintreten von B verringert die Wkt von A .
Das Eintreten von A verringert die Wkt von B . Wir führen dies in Definition T3k aus.

Eine Eigenschaft wie Unabhängigkeit können Sie prüfen; das ist zunächst eine (Beweis)Last.
Umgekehrt jedoch, sobald Sie sie voraussetzen dürfen, ist sie für Sie von großem Nutzen!
In vielen praktischen Anwendungen erlaubt ein sorgfältiger Versuchsaufbau die Annahme der Unabhängigkeit, da die Ereignisse sich aus physikalischen Gründen nicht beeinflussen (können).

Stochastisch un/abhängige Ereignisse

Aufgabe: Sie würfeln zweimal mit einem fairen Würfel.
Dies ist ein Laplace-Experiment auf $\Omega = \{1, 2, 3, 4, 5, 6\}^2$.

$A := \{\text{Die erste Augenzahl ist gerade}\}$

$B := \{\text{Die Augensumme ist gerade}\}$

Sind A und B stochastisch unabhängig?

Lösung: Wir haben $A \cap B = \{\text{Beide Augenzahlen sind gerade}\}$.

Somit finden wir $\mathbf{P}(A) = \frac{1}{2}$ und $\mathbf{P}(B) = \frac{1}{2}$ sowie $\mathbf{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}$.

😊 Die Ereignisse A und B sind demnach **stochastisch unabhängig**.

😊 Auch hier sagt Ihnen die Definition T2c genau, was zu prüfen ist.

Alternative Formulierung: Hier gilt $\mathbf{P}(A) = \frac{1}{2} = \mathbf{P}(A|B)$ und $\mathbf{P}(B) = \frac{1}{2} = \mathbf{P}(B|A)$.

Intuitiv-anschaulich ist das vielleicht etwas überraschend, denn die erste Augenzahl beeinflusst natürlich die Augensumme. Dennoch sind beide Ereignisse A und B stochastisch unabhängig.

Es geht bei der stochastischen Unabhängigkeit nicht um Kausalität o.ä., sondern allein um die Multiplikativität der Wahrscheinlichkeiten. Die Ereignisse A und B mögen noch so raffiniert miteinander verwoben sein, doch es zählt allein die Gleichung $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$.

Nutzen Sie beides: Definition und Beispiele, Präzision und Intuition, Handwerk und Kunst.

Stochastisch un/abhängige Ereignisse

Aufgabe: Sie würfeln zweimal nacheinander mit einem fairen Würfel. Sei X die erste Augenzahl, Y die zweite, $S = X + Y$ ihre Summe.

(1a) Sind die Ereignisse $X = 3$ und $S = 7$ stochastisch unabhängig?

(1b) Sind die Ereignisse $X = 3$ und $S = 6$ stochastisch unabhängig?

(2) Berechnen und interpretieren Sie ebenso die bedingten Wkten.

Lösung: (1a) Wir wissen schon $\mathbf{P}(X = 3) = \frac{1}{6}$ und $\mathbf{P}(S = 7) = \frac{1}{6}$ sowie $\mathbf{P}(X = 3, S = 7) = \mathbf{P}(\{(3, 4)\}) = \frac{1}{36}$. Beide Ereignisse sind unabhängig.

(1b) Wir wissen schon $\mathbf{P}(X = 3) = \frac{1}{6}$ und $\mathbf{P}(S = 6) = \frac{5}{36}$ und finden $\mathbf{P}(X = 3, S = 6) = \mathbf{P}(\{(3, 3)\}) = \frac{1}{36}$. Beide Ereignisse sind abhängig.

$$(2a) \quad \mathbf{P}(S = 7|X = 3) = \frac{1}{6} = \mathbf{P}(S = 7),$$

$$\mathbf{P}(X = 3|S = 7) = \frac{1}{6} = \mathbf{P}(X = 3).$$

$$(2b) \quad \mathbf{P}(S = 6|X = 3) = \frac{1}{6} \neq \frac{5}{36} = \mathbf{P}(S = 6),$$

$$\mathbf{P}(X = 3|S = 6) = \frac{1}{5} \neq \frac{1}{6} = \mathbf{P}(X = 3).$$

Nochmal zur Betonung: Es geht bei der stochastischen Unabhängigkeit nicht um Kausalität! Die Ereignisse mögen noch so raffiniert miteinander verwoben sein, für ihre stochastische Unabhängigkeit zählt allein die Multiplikativität der Wahrscheinlichkeiten. So einfach ist das.

Vierfeldertafel und Unabhängigkeit

Aufgabe: Sind folgende Ereignisse A, B stochastisch unabhängig?

Wkt	B	\bar{B}	gesamt
A	0.28	0.12	0.40
\bar{A}	0.42	0.18	0.60
gesamt	0.70	0.30	1.00

Lösung: Ja! Aus $P(A)$, $P(B)$ und Multiplikativität folgt die Tabelle.

Aufgabe: Sind folgende Ereignisse X, Y stochastisch unabhängig?

Wkt	Y	\bar{Y}	gesamt
X	0.30	0.10	0.40
\bar{X}	0.40	0.20	0.60
gesamt	0.70	0.30	1.00

Lösung: Nein! Multiplikativität gilt hier nicht; X, Y sind abhängig.

Lemma T2D: Komplemente

Sind (A, B) unabhängig, dann auch (A, \bar{B}) und (\bar{A}, B) sowie (\bar{A}, \bar{B}) .
Für jedes Ereignis $A \subseteq \Omega$ sind zudem (A, \emptyset) und (A, Ω) unabhängig.

Aufgabe: Rechnen Sie dies sorgfältig nach! **Lösung:**

$$\begin{aligned}\mathbf{P}(A \cap \bar{B}) &= \mathbf{P}(A \setminus (A \cap B)) = \mathbf{P}(A) - \mathbf{P}(A \cap B) \\ &= \mathbf{P}(A) - \mathbf{P}(A) \mathbf{P}(B) = \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A) \mathbf{P}(\bar{B})\end{aligned}$$

Jedes Ereignis A ist von \emptyset stochastisch unabhängig:

$$\mathbf{P}(A \cap \emptyset) = \mathbf{P}(\emptyset) = 0 = \mathbf{P}(A) \cdot \mathbf{P}(\emptyset)$$

Jedes Ereignis A ist von Ω stochastisch unabhängig:

$$\mathbf{P}(A \cap \Omega) = \mathbf{P}(A) = \mathbf{P}(A) \cdot \mathbf{P}(\Omega)$$

Zusammengefasst: Jedes der Ereignisse $\emptyset, A, \bar{A}, \Omega$ ist stochastisch unabhängig von jedem der Ereignisse $\emptyset, B, \bar{B}, \Omega$. Das nützt gelegentlich.

Sudoku für Wahrscheinlichkeiten

Aufgabe: Wie können Sie die Tabelle füllen? bei Unabhängigkeit?

Wkt	B	\bar{B}	ges.
A			0.36
\bar{A}			
ges.	0.25		

Wkt	B	\bar{B}	ges.
A	[0.00, 0.25]	[0.11, 0.36]	0.36
\bar{A}	[0.00, 0.25]	[0.39, 0.64]	0.64
ges.	0.25	0.75	1.00

Wkt	B	\bar{B}	ges.
A	0.09	0.27	0.36
\bar{A}	0.16	0.48	0.64
ges.	0.25	0.75	1.00

Ohne die Voraussetzung der Unabhängigkeit sind die rot markierten Einträge nicht eindeutig: Sie können ein beliebiges Feld aussuchen und frei festlegen – im offensichtlichen Intervall! Bei Unabhängigkeit gibt es jeweils nur eine Lösung (hier gelb). Ein raffinierteres Beispiel: Gleiches gilt bei der nächsten Frage, es gibt aber zwei Lösungen (quadratische Gleichung).

Wkt	B	\bar{B}	ges.
A	0.12		
\bar{A}		0.32	
ges.			

Wkt	B	\bar{B}	ges.
A	0.12	[0.00, 0.56]	[0.12, 0.68]
\bar{A}	[0.00, 0.56]	0.32	[0.32, 0.88]
ges.	[0.12, 0.68]	[0.32, 0.88]	1.00

Wkt	B	\bar{B}	ges.
A	0.12	0.08 0.48	0.20 0.60
\bar{A}	0.48 0.08	0.32	0.80 0.40
ges.	0.60 0.20	0.40 0.80	1.00

Annähernde Unabhängigkeit

Aufgabe: Drei Absender schreiben je einen Brief an einen von zwei Empfängern. Die Wahlen sind zufällig, gleichverteilt und unabhängig.

(1) Berechnen Sie $\mathbf{P}(A_1)$, $\mathbf{P}(A_2)$, $\mathbf{P}(A_1 \cap A_2)$ für die Ereignisse

$$A_i = \{\text{Der Empfänger } i \text{ erhält mindestens einen Brief}\}.$$

(2) Sind A_1, A_2 stochastisch unabhängig? (3) Was gilt für n Absender?

Lösung: Es gibt $2^3 = 8$ gleichwahrscheinliche Zuordnungen:

$$(1, 1, 1) \quad (1, 1, 2) \quad (1, 2, 1) \quad (1, 2, 2)$$

$$(2, 1, 1) \quad (2, 1, 2) \quad (2, 2, 1) \quad (2, 2, 2)$$

(1) Demnach finden wir $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 7/8$ und $\mathbf{P}(A_1 \cap A_2) = 6/8$.

(2) Wegen $7/8 \cdot 7/8 \neq 6/8$ sind A_1, A_2 stochastisch abhängig.

(3) Allgemein $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 1 - 2^{-n}$ und $\mathbf{P}(A_1 \cap A_2) = 1 - 2^{1-n}$.

Unabhängigkeit gilt hierbei nie, aber die Abhängigkeit nimmt ab:

Für $n \rightarrow \infty$ gilt $\mathbf{P}(A_1) \mathbf{P}(A_2) / \mathbf{P}(A_1 \cap A_2) \rightarrow 1$. Das ist plausibel!

Unabhängigkeit mehrerer Ereignisse

Zwei Ereignisse A, B sind unabhängig, wenn gilt:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$$


Drei Ereignisse A, B, C sind unabhängig, wenn gilt:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$$

$$\mathbf{P}(A \cap C) = \mathbf{P}(A) \mathbf{P}(C)$$

$$\mathbf{P}(B \cap C) = \mathbf{P}(B) \mathbf{P}(C)$$

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

 Die Tripelbedingung folgt nicht aus den drei Paarbedingungen!
Sie kennen dieses Problem von linearer Unabhängigkeit von Vektoren.

Definition T2E: stochastische Unabhängigkeit

Sei (Ω, \mathbf{P}) ein Wahrscheinlichkeitsraum. Eine Familie von Ereignissen $A_1, A_2, \dots, A_n \subseteq \Omega$ heißt **(stochastisch) unabhängig**, wenn folgende Produktformel für jede Auswahl $1 \leq i_1 < i_2 < \dots < i_\ell \leq n$ gilt:

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}) = \mathbf{P}(A_{i_1}) \mathbf{P}(A_{i_2}) \dots \mathbf{P}(A_{i_\ell})$$

Unabhängigkeit mehrerer Ereignisse

Sie werfen zweimal eine faire Münze: 0 = Kopf, 1 = Zahl.

Aufgabe: Sind folgende Ereignisse (A, B, C) unabhängig?

$$A = \{\text{Zahl beim ersten Wurf}\} = \{(1, 0), (1, 1)\}$$

$$B = \{\text{Zahl beim zweiten Wurf}\} = \{(0, 1), (1, 1)\}$$

$$C = \{\text{gleiches Ergebnis bei beiden Würfeln}\} = \{(0, 0), (1, 1)\}$$

Welche dieser drei Paare sind unabhängig? Ist das Tripel unabhängig?

Lösung: Der Ergebnisraum ist $\Omega = \{0, 1\}^2$ mit Gleichverteilung.

Wir berechnen die erforderlichen Wahrscheinlichkeiten:

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

$$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \frac{1}{4},$$

$$\mathbf{P}(A \cap B \cap C) = \frac{1}{4} \quad \text{und} \quad \mathbf{P}(A) \mathbf{P}(B) \mathbf{P}(C) = \frac{1}{8}.$$

😊 Jedes der Paare (A, B) und (A, C) und (B, C) ist unabhängig.

⚠️ Das Tripel (A, B, C) ist nicht unabhängig! Lesen Sie die Definition!
Sie kennen dieses Problem von linearer Unabhängigkeit von Vektoren.

Unabhängigkeit mehrerer Ereignisse

Aufgabe: In einer Urne befinden sich die Lose 112, 121, 211, 222. Sei A_i das Ereignis „1 an i -ter Stelle“. Sind A_1, A_2 unabhängig? und A_1, A_3 ? und A_2, A_3 ? Sind A_1, A_2, A_3 unabhängig?

Lösung: Wir finden $\mathbf{P}(A_1) = \mathbf{P}(A_2) = \mathbf{P}(A_3) = 1/2$ und $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1 \cap A_3) = \mathbf{P}(A_2 \cap A_3) = 1/4$: Die Ereignisse A_1, A_2, A_3 sind paarweise unabhängig. Es gibt kein Los 111, daher gilt $\mathbf{P}(A_1 \cap A_2 \cap A_3) = 0$: Das Tripel (A_1, A_2, A_3) ist insgesamt nicht unabhängig.

Aufgabe: Sie würfeln zweimal mit einem fairen Würfel. Berechnen Sie die Wkten von $A = \{\text{erster gerade}\}$ und $B = \{\text{zweiter ungerade}\}$ und $C = \{\text{Summe gleich } 7\}$ und ihrer Schnitte. Sind A, B, C unabhängig?

Lösung: Wir finden $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$ und $\mathbf{P}(C) = 1/6$ sowie $\mathbf{P}(A \cap B) = 1/4$ und $\mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \mathbf{P}(A \cap B \cap C) = 1/12$: Die Familie (A, B, C) ist paarweise unabhängig, aber nicht unabhängig. Wir kennen dieses erstaunliche Phänomen bereits. Es tritt häufig auf! Es ist keine exotische Ausnahme, sondern ein typisches Problem.

Unabhängigkeit mehrerer Ereignisse

Aufgabe: Wir würfeln zweimal mit einem fairen Würfel. Berechnen Sie die Wkten von $A = \{\text{erster gerade}\}$ und $B = \{\text{zweiter 2 oder 5}\}$ und $C = \{\text{Summe gleich 7}\}$ und ihrer Schnitte. Sind A, B, C unabhängig?

Lösung: Wir finden $\mathbf{P}(A) = 1/2$ und $\mathbf{P}(B) = 1/3$ und $\mathbf{P}(C) = 1/6$ sowie $\mathbf{P}(A \cap B) = 1/6$ und $\mathbf{P}(A \cap C) = 1/12$ und $\mathbf{P}(B \cap C) = 1/18$ und $\mathbf{P}(A \cap B \cap C) = 1/36$. Die Familie (A, B, C) ist also unabhängig.

Aufgabe: Seien (A, B, C) drei unabhängige Ereignisse $A, B, C \subseteq \Omega$. Sind dann auch die folgenden Tripel von Ereignissen unabhängig?

(1) (A, C, B) , (2) (B, A, C) , (3) (A, B, \overline{C}) , (4) $(\overline{A}, \overline{B}, \overline{C})$

Lösung: Ja! Alle Bedingungen in der Definition der Unabhängigkeit sind invariant unter Vertauschung (1,2). Wir rechnen (3) exemplarisch nach:

$$\begin{aligned}\mathbf{P}(A \cap B \cap \overline{C}) &= \mathbf{P}((A \cap B) \setminus (A \cap B \cap C)) \\ &= \mathbf{P}(A \cap B) - \mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \cdot \mathbf{P}(B) - \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C) \\ &= \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot (1 - \mathbf{P}(C)) = \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(\overline{C})\end{aligned}$$

Jedes der $3! \cdot 2^3 = 48$ so variierten Tripel ist stochastisch unabhängig.

Die Siebformel

Erinnerung: Für $A, B \subseteq \Omega$ gilt $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Aufgabe: (1) Erklären Sie graphisch und rechnerisch die Siebformel:

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

(2) Was erhalten Sie im Spezialfall, wenn A, B, C unabhängig sind?

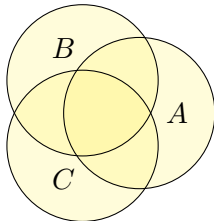
(3) Formulieren und erklären Sie die Formel für $\mathbf{P}(A \cup B \cup C \cup D)$.

(4) Formulieren Sie die allgemeine Siebformel: Aus $A = \bigcup_{i=1}^n A_i$ folgt $\prod_{i=1}^n (\mathbf{I}_A - \mathbf{I}_{A_i}) = 0$ und somit $\mathbf{I}_A = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|-1} \mathbf{I}_{A_J}$.

Lösung: (1a) Die Summe $\mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C)$ zählt manche Elemente doppelt oder gar dreifach.

Nach Korrektur $-\mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C)$ zählen die Elemente in $A \cap B \cap C$ gar nicht mehr.

Nach Korrektur $+\mathbf{P}(A \cap B \cap C)$ stimmt alles.



Die Siebformel

(1b) Wir nutzen $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ und finden:

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

(2) Sind A, B, C unabhängig, so finden wir die vertraute Produktformel:

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(\overline{A \cap B \cap C}) = 1 - [1 - \mathbf{P}(A)][1 - \mathbf{P}(B)][1 - \mathbf{P}(C)]$$

(4) Für $A_1, \dots, A_n \subseteq \Omega$ und $A = \bigcup_{i=1}^n A_i$ gilt $\prod_{i=1}^n (\mathbf{I}_A - \mathbf{I}_{A_i}) = 0$, denn für jedes $x \in A$ ist mindestens ein Faktor Null. Für jede Indexmenge $J \subseteq \{1, \dots, n\}$ sei $A_J := \bigcap_{i \in J} A_i$ die Schnittmenge, speziell $A_\emptyset = A$. Wir erhalten die **Siebformel**, auch **Inklusions-Exklusions-Prinzip**:

$$\begin{aligned} \mathbf{I}_A &= \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|-1} \mathbf{I}_{A_J}, \\ \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|-1} \mathbf{P}(A_J). \end{aligned}$$

Siebformel vs Produktformel

Aufgabe: (nach Lewis Carroll) Projekte im Research & Development leiden erfahrungsgemäß zu 70% an irrealen Zielen, 75% an schlechter Kommunikation, 80% an Missmanagement, 85% an Zeitmangel.

Wie viele Projekte leiden an allem, also an irrealen Zielen, an schlechter Kommunikation, an Missmanagement und natürlich an Zeitmangel. . .

(1) höchstens? (2) mindestens? (0) wenn diese unabhängig sind?

Nennen Sie Mengen $I, K, M, Z \subseteq \Omega$, die diese Schranken realisieren!

M	Z	Z			
76...80	81...85	86...90	91...95	96...100	
M	Z	M	Z	M	Z
51...55	56...60	61...65	66...70	71...75	
I	K	I	K	I	K
M	Z	M	Z	M	Z
26...30	31...35	36...40	41...45	46...50	
I	K	I	K	I	K
M	Z	M	Z	M	Z
1...5	6...10	11...15	16...20	21...25	
I	K	I	K	I	K

M	Z	M	Z	M	Z	M	Z	M	Z
76...80	81...85	86...90	91...95	96...100					
K	K	K	K	K					
M	Z	M	Z	M	Z	M	Z	M	Z
51...55	56...60	61...65	66...70	71...75					
I	I	I	I	I				K	
Z	Z	Z	Z	Z	M	Z			
26...30	31...35	36...40	41...45	46...50					
I	K	I	K	I	K	I	K	I	
M	Z	M	Z	M	Z	M	Z	M	Z
1...5	6...10	11...15	16...20	21...25					
I	K	I	K	I	K	I	K	I	K

Siebformel vs Produktformel

Lösung: (0) Bei Unabhängigkeit von I, K, M, Z ergibt die Produktformel ganz einfach $\mathbf{P}(I \cap K \cap M \cap Z) = \mathbf{P}(I) \cdot \mathbf{P}(K) \cdot \mathbf{P}(M) \cdot \mathbf{P}(Z) = 0.357$.

(1) Höchstens 70% leiden an allen vier; diese Schranke ist offensichtlich.

(2) Mindestens 10% leiden an allen vier; das rechnen wir geduldig aus:

$$\begin{aligned}\mathbf{P}(I \cap K) &= \mathbf{P}(I) + \mathbf{P}(K) - \mathbf{P}(I \cup K) \\ &\geq 0.70 + 0.75 - 1 = 0.45\end{aligned}$$

$$\begin{aligned}\mathbf{P}(I \cap K \cap M) &= \mathbf{P}(I \cap K) + \mathbf{P}(M) - \mathbf{P}((I \cap K) \cup M) \\ &\geq 0.45 + 0.80 - 1 = 0.25\end{aligned}$$

$$\begin{aligned}\mathbf{P}(I \cap K \cap M \cap Z) &= \mathbf{P}(I \cap K \cap M) + \mathbf{P}(Z) - \mathbf{P}((I \cap K \cap M) \cup Z) \\ &\geq 0.25 + 0.85 - 1.00 = 0.10\end{aligned}$$

Die oben skizzierten Graphiken zeigen, dass diese Werte tatsächlich angenommen werden können. Unsere Schranken sind also optimal.

Alternativ: Dank DeMorgan gilt $\overline{I \cap K \cap M \cap Z} = \bar{I} \cup \bar{K} \cup \bar{M} \cup \bar{Z}$, also $\mathbf{P}(\overline{I \cap K \cap M \cap Z}) \leq \mathbf{P}(\bar{I}) + \mathbf{P}(\bar{K}) + \mathbf{P}(\bar{M}) + \mathbf{P}(\bar{Z}) \leq 0.90$.

Diskrete Zufallsvariablen

Beispiel: Bei n -maligem Würfeln nutzen wir $\Omega = \{1, 2, \dots, 6\}^n$ mit Gleichverteilung \mathbf{P} und $S_n : \Omega \rightarrow \mathbb{R} : (\omega_1, \dots, \omega_n) \mapsto \omega_1 + \dots + \omega_n$.

Definition T3A: Zufallsvariable, zunächst diskret

Sei (Ω, \mathbf{P}) ein WRaum. Eine reelle **Zufallsvariable** ist eine Abbildung

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega).$$

Das heißt: X ordnet jedem Ergebnis $\omega \in \Omega$ eine Zahl $X(\omega) \in \mathbb{R}$ zu. Somit transportiert X das WMaß \mathbf{P} auf Ω zum **Bildmaß** \mathbf{P}_X auf \mathbb{R} :

$$\mathbf{P}_X(\{a\}) = \mathbf{P}(X = a) := \mathbf{P}(X^{-1}(\{a\})) = \mathbf{P}(\{\omega \in \Omega \mid X(\omega) = a\})$$

$$\mathbf{P}_X(A) = \mathbf{P}(X \in A) := \mathbf{P}(X^{-1}(A)) = \mathbf{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$$

für jedes Element $a \in \mathbb{R}$ und jede Teilmenge $A \subseteq \mathbb{R}$, etwa $A = [a, b]$.

Satz T3B: Auch das Bildmaß ist ein Wahrscheinlichkeitsmaß.

Sei (Ω, \mathbf{P}) ein (diskreter) WRaum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann ist \mathbf{P}_X auf \mathbb{R} ein (diskretes) Wahrscheinlichkeitsmaß (T1D).

Diskrete Zufallsvariablen

😊 Wir setzen den WRaum (Ω, \mathbf{P}) hier vorerst als diskret voraus (T1D). Später ist $(\Omega, \mathcal{A}, \mathbf{P})$ ein allgemeiner WRaum; wir fordern dann von X zudem Messbarkeit, also $X^{-1}([a, b]) \in \mathcal{A}$ für jedes Intervall $[a, b] \in \mathbb{R}$.

Aufgabe: (1) Interpretieren Sie die Abbildung $X : \Omega \rightarrow \mathbb{R}$ stochastisch. (2) Warum ist \mathbf{P}_X auf \mathbb{R} ein diskretes WMaß? Was ist hier zu prüfen?

Lösung: (1) Die Zufallsvariable X ordnet jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega) = a \in \mathbb{R}$ zu, etwa den Gewinn oder eine Auszahlung. Oft beobachten wir nicht direkt das Ergebnis ω , sondern die abgeleitete Zufallsgröße $X(\omega) = a$. Die Menge $X^{-1}(\{a\}) := \{\omega \in \Omega \mid X(\omega) = a\}$ fasst alle Ergebnisse $\omega \in \Omega$ zusammen, die zum Wert $X(\omega) = a$ führen.

(2) Wir wenden Definition T1D an. (a) $\mathbf{P}_X(\mathbb{R}) = \mathbf{P}(X^{-1}(\mathbb{R})) = \mathbf{P}(\Omega) = 1$. (b) Ebenso prüfen wir die Additivität: Für jede Teilmenge $A \subseteq \mathbb{R}$ gilt

$$\begin{aligned} \mathbf{P}_X(A) &\stackrel{\text{Def}}{=} \mathbf{P}(X^{-1}(A)) &&\stackrel{\text{Def}}{=} \mathbf{P}(\{\omega \in \Omega \mid X(\omega) \in A\}) \\ &\stackrel{\text{Dis}}{=} \mathbf{P}\left(\bigsqcup_{a \in A} \{\omega \in \Omega \mid X(\omega) = a\}\right) &&\stackrel{\text{Def}}{=} \mathbf{P}\left(\bigsqcup_{a \in A} X^{-1}(\{a\})\right) \\ &\stackrel{\text{Add}}{=} \sum_{a \in A} \mathbf{P}(X^{-1}(\{a\})) &&\stackrel{\text{Def}}{=} \sum_{a \in A} \mathbf{P}_X(\{a\}) \end{aligned}$$

Erwartungswert einer diskreten Zufallsvariablen

Definition T3c: Erwartungswert, zunächst diskret

Die **absolute Erwartung** der Zufallsvariable $X : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ ist

$$\mathbf{E}(|X|) := \sum_{\omega \in \Omega} |X(\omega)| \cdot \mathbf{P}(\{\omega\}) \in [0, \infty].$$

Ist dieser Wert endlich, so nennen wir X **absolut summierbar**.

Wir sagen hierzu anschaulich **der Erwartungswert von X existiert**; dies gilt immer, wenn Ω endlich oder X beschränkt ist. Wir schreiben:

$$L^1(\Omega, \mathbf{P}) = L^1(\Omega, \mathbf{P}; \mathbb{R}) := \{ X : \Omega \rightarrow \mathbb{R} \mid \mathbf{E}(|X|) < \infty \}$$

Für jede Zufallsvariable $X \in L^1(\Omega, \mathbf{P})$ ist der **Erwartungswert** definiert:

$$\mathbf{E}(X) := \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\{\omega\}) = \sum_{a \in \mathbb{R}} a \cdot \mathbf{P}(X = a)$$

Satz T3D: Die Erwartung ist linear.

Für jeden WRaum (Ω, \mathbf{P}) ist die Menge $L^1(\Omega, \mathbf{P})$ ein \mathbb{R} -Vektorraum und die Erwartung $\mathbf{E} : L^1(\Omega, \mathbf{P}) \rightarrow \mathbb{R} : X \mapsto \mathbf{E}(X)$ eine \mathbb{R} -lineare Abbildung.

Erwartungswert einer diskreten Zufallsvariablen

😊 Wir setzen den WRaum (Ω, \mathbf{P}) hier vorerst als diskret voraus (T1D). Im diskreten Fall genügen Reihen, im Allgemeinen nutzen wir Integrale. Für Reihen und Integrale verfügen Sie über alle nötigen Werkzeuge.

Aufgabe: (1) Der Erwartungswert $\mathbf{E}(X)$ kann wie angegeben auf zwei Arten berechnet werden: Beide Summen ergeben denselben Wert!

(2) Beweisen Sie den Satz T3D durch sorgfältiges Nachrechnen.

Lösung: (1) Dank absoluter Summierbarkeit (A2N) dürfen wir umordnen:

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{P}(\{\omega\}) &\stackrel{\text{Um}}{\underset{\text{A2N}}{=}} \sum_{a \in \mathbb{R}} \left[\sum_{\omega \in X^{-1}(\{a\})} X(\omega) \cdot \mathbf{P}(\{\omega\}) \right] \\ &\stackrel{\text{Lin}}{=} \sum_{a \in \mathbb{R}} a \cdot \left[\sum_{\omega \in X^{-1}(\{a\})} \mathbf{P}(\{\omega\}) \right] \stackrel{\text{Def}}{=} \sum_{a \in \mathbb{R}} a \cdot \mathbf{P}(X = a) \end{aligned}$$

(2) Es gilt $\mathbf{E}(|\lambda \cdot X|) = |\lambda| \cdot \mathbf{E}(|X|)$ und $\mathbf{E}(|X + Y|) \leq \mathbf{E}(|X|) + \mathbf{E}(|Y|)$, für alle Zufallsvariablen $X, Y : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ und $\lambda \in \mathbb{R}$, daher ist $L^1(\Omega, \mathbf{P})$ ein \mathbb{R} -Vektorraum. Für $X, Y \in L^1(\Omega, \mathbf{P})$ gilt zudem $\mathbf{E}(\lambda \cdot X) = \lambda \cdot \mathbf{E}(X)$ und $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ dank Linearität (konvergenter!) Reihen. Die Abbildung $\mathbf{E} : L^1(\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ ist also \mathbb{R} -linear, wie gewünscht.

Zufallsvariable, Verteilung und Erwartung

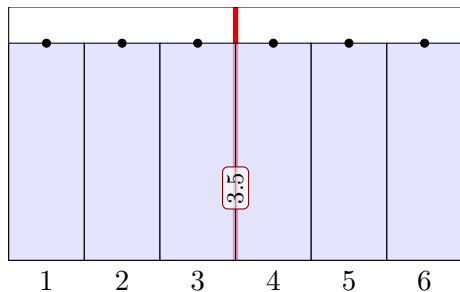
Aufgabe: Bei zweimaligem Würfeln betrachten wir die Ergebnismenge $\Omega = \{1, 2, \dots, 6\}^2$ mit der Gleichverteilung \mathbf{P} und den Zufallsvariablen

$$X, Y : \Omega \rightarrow \mathbb{R} : \quad X(\omega_1, \omega_2) = \omega_1, \quad Y(\omega_1, \omega_2) = \omega_2.$$

Berechnen Sie Verteilung und Erwartung der Zufallsvariablen X , Y , sowie $X - Y$ und $|X - Y|$, schließlich $\max(X, Y)$ und $\min(X, Y)$.

Lösung: Wir beginnen mit der Zufallsvariable $X(\omega_1, \omega_2) = \omega_1$:

$\omega_2 \uparrow$	X					
6	1	2	3	4	5	6
5	1	2	3	4	5	6
4	1	2	3	4	5	6
3	1	2	3	4	5	6
2	1	2	3	4	5	6
1	1	2	3	4	5	6
	1	2	3	4	5	6
	$\omega_1 \rightarrow$					



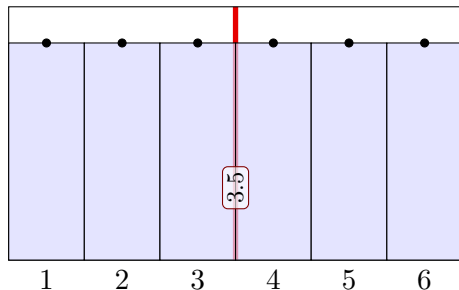
Der Erwartungswert ist hier $\mathbf{E}(X) = 3.5$, wie Sie leicht nachrechnen. Erwartungswert bedeutet nicht häufigster Wert, sondern Schwerpunkt!

Zufallsvariable, Verteilung und Erwartung

Die ZVariable Y ist verschieden von X , aber identisch verteilt:

$\omega_2 \uparrow$	Y					
6	6	6	6	6	6	6
5	5	5	5	5	5	5
4	4	4	4	4	4	4
3	3	3	3	3	3	3
2	2	2	2	2	2	2
1	1	1	1	1	1	1
	1	2	3	4	5	6

$\omega_1 \rightarrow$

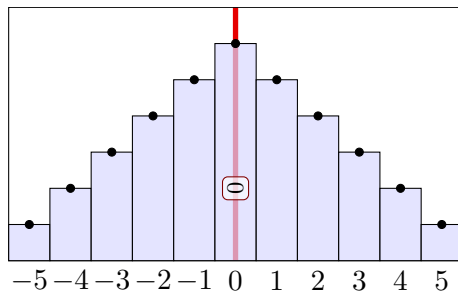


Wir sehen in den ausführlichen Tabellen links die beiden ZVariablen $X \neq Y$. Die Graphik rechts zeigt die Verteilungen, hier gilt $\mathbf{P}_X = \mathbf{P}_Y$. Dies zeigt eindrücklich: Das Bildmaß ist ein stark vergrößertes Abbild! Die letzte Formel von Definition T3C sagt: Der Erwartungswert $\mathbf{E}(X)$ hängt nur von der Verteilung \mathbf{P}_X ab: Aus $\mathbf{P}_X = \mathbf{P}_Y$ folgt $\mathbf{E}(X) = \mathbf{E}(Y)$. Diese Situation ist typisch für Anwendungen, in denen eine Messung / ein Experiment n -mal unabhängig wiederholt wird: Alle n ZVariablen X_1, \dots, X_n sind unabhängig und identisch verteilt. Dazu später mehr!

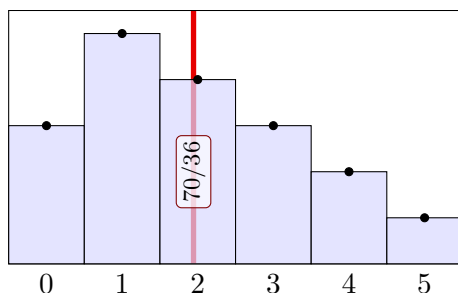
Zufallsvariable, Verteilung und Erwartung

Für $X - Y$ gilt $\mathbf{E}(X - Y) = \mathbf{E}(X) - \mathbf{E}(Y) = 0$ dank Linearität.

$\omega_2 \uparrow$	$Z = X - Y$						
6	-5	-4	-3	-2	-1	0	
5	-4	-3	-2	-1	0	1	
4	-3	-2	-1	0	1	2	
3	-2	-1	0	1	2	3	
2	-1	0	1	2	3	4	
1	0	1	2	3	4	5	
	1	2	3	4	5	6	ω_1



$\omega_2 \uparrow$	$Z = X - Y $						
6	5	4	3	2	1	0	
5	4	3	2	1	0	1	
4	3	2	1	0	1	2	
3	2	1	0	1	2	3	
2	1	0	1	2	3	4	
1	0	1	2	3	4	5	
	1	2	3	4	5	6	ω_1

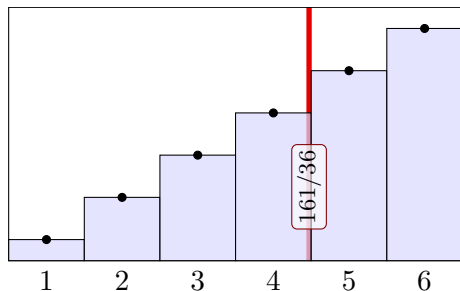


Zufallsvariable, Verteilung und Erwartung

Für die Zufallsvariablen $\max(X, Y)$ und $\min(X, Y)$ finden wir:

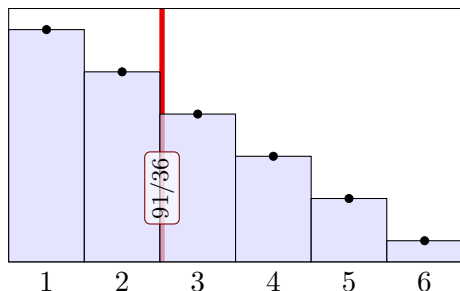
$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
6	6	6	6	6	6	6
5	5	5	5	5	5	6
4	4	4	4	4	5	6
3	3	3	3	4	5	6
2	2	2	3	4	5	6
1	1	2	3	4	5	6

$Z = \max(X, Y)$



$\omega_2 \backslash \omega_1$	1	2	3	4	5	6
6	1	2	3	4	5	6
5	1	2	3	4	5	5
4	1	2	3	4	4	4
3	1	2	3	3	3	3
2	1	2	2	2	2	2
1	1	1	1	1	1	1

$Z = \min(X, Y)$



Definition T3E: Varianz und höhere Momente

Sei (Ω, \mathbf{P}) ein WRaum und $X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega)$ eine Zufallsvariable. Zudem sei $\mathbf{E}(|X|) < \infty$, so dass der Erwartungswert $\mu = \mathbf{E}(X)$ existiert.

Die **Varianz** der ZVariable X ist die mittlere quadratische Abweichung:

$$\sigma^2(X) = \mathbf{V}(X) := \mathbf{E}[(X - \mu)^2] = \mathbf{E}(X^2) - \mathbf{E}(X)^2 \geq 0$$

Die **Streuung** oder **Standardabweichung** von X ist die Quadratwurzel

$$\sigma(X) := \sqrt{\mathbf{V}(X)} \geq 0.$$

Das (absolute, zentrierte) **n -te Moment** von X ist gegeben durch

$$\rho^n(X) := \mathbf{E}[|X - \mu|^n].$$

Anschaulich ist der Erwartungswert μ der Schwerpunkt der Verteilung. Die Varianz $\mathbf{E}[(X - \mu)^2]$ entspricht dem Trägheitsmoment. Varianz und höhere Momente messen, wie weit die Werte um μ streuen: Große Abweichungen wiegen schwerer, quadratisch oder sogar hoch n .

Erwartung und Varianz von Zufallsvariablen

Aufgabe: Berechnen Sie das 1./2./3. Moment für $X = \mathbf{I}_A \sim B(1, t)$.

Lösung: Hier ist $X : \Omega \rightarrow \{0, 1\}$ eine Zufallsvariable mit Verteilung

$$\mathbf{P}(X=1) = \mathbf{P}(A) = t \quad \text{und} \quad \mathbf{P}(X=0) = \mathbf{P}(\Omega \setminus A) = 1 - t =: \bar{t}.$$

Ihre Erwartung und 1./2./3. Momente berechnen wir nach Definition:

$$\mu = \mathbf{E}(X) = 0 \cdot \mathbf{P}(X=0) + 1 \cdot \mathbf{P}(X=1) = t$$

$$\begin{aligned} \mathbf{E}(|X - \mu|^1) &= |0 - \mu|^1 \cdot \mathbf{P}(X=0) + |1 - \mu|^1 \cdot \mathbf{P}(X=1) \\ &= t^1 \cdot (1 - t) + (1 - t)^1 \cdot t = 2t(1 - t) \end{aligned}$$

$$\begin{aligned} \mathbf{E}(|X - \mu|^2) &= |0 - \mu|^2 \cdot \mathbf{P}(X=0) + |1 - \mu|^2 \cdot \mathbf{P}(X=1) \\ &= t^2 \cdot (1 - t) + (1 - t)^2 \cdot t = t(1 - t) \end{aligned}$$

$$\begin{aligned} \mathbf{E}(|X - \mu|^3) &= |0 - \mu|^3 \cdot \mathbf{P}(X=0) + |1 - \mu|^3 \cdot \mathbf{P}(X=1) \\ &= t^3 \cdot (1 - t) + (1 - t)^3 \cdot t = t\bar{t}(t^2 + \bar{t}^2) \end{aligned}$$

😊 Zum Vergleich: Auch $\mathbf{E}(X^2) - \mathbf{E}(X)^2 = t - t^2$ liefert die Varianz.

Warnende Beispiele: divergente Reihen

Aufgabe: Seien $X, Y : \Omega \rightarrow \mathbb{Z}$ Zufallsvariablen mit Verteilung

$$\mathbf{P}[X=(-2)^k] = \frac{a}{2^k} \quad \text{und} \quad \mathbf{P}[Y=k] = \frac{b}{k^2} \quad \text{für} \quad k = 1, 2, 3, \dots$$


- (1) Sind dies tatsächlich WVerteilungen? Für welche a bzw. b ?
 (2) Existiert der Erwartungswert $\mathbf{E}(X)$ bzw. $\mathbf{E}(Y)$?

Lösung: (1) Ja: Es gilt jeweils $p(k) \geq 0$ und $\sum_{k=1}^{\infty} p(k) = 1$, wobei wir die Konstanten $a = 1$ und $b = 6/\pi^2$ setzen müssen. I324

(2) Nein, die zum Erwartungswert benötigte Reihe konvergiert nicht:

$$\sum_{k=1}^{\infty} x_k \cdot \mathbf{P}[X=x_k] = \sum_{k=1}^{\infty} (-2)^k \cdot \frac{a}{2^k} = -1 + 1 - 1 + 1 - 1 \pm \dots$$

$$\sum_{k=1}^{\infty} k \cdot \mathbf{P}[Y=k] = \sum_{k=1}^{\infty} k \cdot \frac{b}{k^2} = \frac{6}{\pi^2} \left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots \right)$$

 Einen Erwartungswert können wir hier also nicht definieren!
 Zur Rechnung für die geometrische Verteilung siehe T421.

Warnende Beispiele: die Zeta-Verteilung

Aufgabe: Sei $Z : \Omega \rightarrow \mathbb{N}$ eine Zufallsvariable mit der **Zeta-Verteilung**

$$\mathbf{P}[Z=k] = \frac{c}{k^\alpha} \quad \text{für } k = 1, 2, 3, \dots$$

(1) Für welche $\alpha, c \in \mathbb{R}$ ist dies tatsächlich eine WVerteilung?

(2) Für welche $\alpha \in \mathbb{R}$ existiert der Erwartungswert $\mu = \mathbf{E}[Z]$?
 zudem endliche Varianz $\sigma^2 = \mathbf{V}(Z) = \mathbf{E}(|Z - \mu|^2)$?
 und endliches n -tes Moment $\rho^n = \mathbf{E}(|Z - \mu|^n)$?

Lösung: (1) Es muss $c \in]0, \infty[$ gelten und $c^{-1} = \sum_{k=1}^{\infty} 1/k^\alpha =: \zeta(\alpha)$.
 Die Reihe $\sum_{k=1}^{\infty} 1/k^\alpha$ konvergiert jedoch nur für $\alpha > 1$. (Warum?)
 Für $\alpha \leq 1$ ist dies also keine Wahrscheinlichkeitsverteilung!

(2) Die Reihe $\mathbf{E}[Z] = \sum_{k=1}^{\infty} k/k^\alpha = \zeta(\alpha - 1)$ konvergiert nur für $\alpha > 2$.
 Für $1 < \alpha \leq 2$ hingegen existiert der Erwartungswert nicht!

Die Reihe $\mathbf{E}[Z^2] = \sum_{k=1}^{\infty} k^2/k^\alpha = \zeta(\alpha - 2)$ konvergiert nur für $\alpha > 3$.
 Für $2 < \alpha \leq 3$ existiert die Erwartung, aber die Varianz ist unendlich!

Allgemein: Die Reihe $\mathbf{E}[Z^n] = \sum_{k=1}^{\infty} k^n/k^\alpha = \zeta(\alpha - n)$ konvergiert nur
 für $\alpha > n + 1$. Für $1 < \alpha \leq n$ hingegen ist das n -te Moment unendlich!

Erwartung und Varianz von Zufallsvariablen

Aufgabe: (1) Sei $\mu = \mathbf{E}(X)$. Zeigen Sie die nützliche Formel

$$\sigma^2 = \mathbf{V}(X) = \mathbf{E}[(X - \mu)^2] \stackrel{!}{=} \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

(2) Wie verhalten sich Erwartung und Varianz bei Verschiebung zu $Y = X + c$ mit $c \in \mathbb{R}$ und Skalierung zu $Z = aX$ mit $a \in \mathbb{R}$?

(3) Was gilt demnach für $Y = X - \mu$ und $Z = (X - \mu)/\sigma$?

Nachrechnen: (1) Wir fordern $\mathbf{E}(|X|) < \infty$ und setzen $\mu = \mathbf{E}(X)$ gemäß Definition T3A. Dank Linearität der Erwartung T3D gilt dann

$$\begin{aligned} 0 \leq \mathbf{E}[(X - \mu)^2] &= \mathbf{E}[X^2 - 2X\mu + \mu^2] = \mathbf{E}[X^2] - 2\mathbf{E}[X]\mu + \mu^2 \\ &= \mathbf{E}(X^2) - 2\mu^2 + \mu^2 = \mathbf{E}(X^2) - \mathbf{E}(X)^2 \end{aligned}$$

😊 Die Varianz $\mathbf{V}(X) = \mathbf{E}[(X - \mu)^2] = \mathbf{E}(X^2) - \mathbf{E}(X)^2$ können Sie also mit jeder der beiden Formeln ausrechnen. Beide sind nützlich.

⚠ Dies ist eine besondere Eigenschaft des zweiten Moments. Für alle anderen Momente $\mathbf{E}[|X - \mu|^n]$ mit $n \neq 2$ gilt dies nicht!

Erwartung und Varianz von Zufallsvariablen

(2) Dank der Normierung $\mathbf{P}(\Omega) = 1$ und der Linearität der Erwartung gilt immer $\mathbf{E}(Y) = \mathbf{E}(X + c) = \mathbf{E}(X) + \mathbf{E}(c) = \mathbf{E}(X) + c$, sowie $\mathbf{E}(Z) = \mathbf{E}(aX) = a\mathbf{E}(X)$. Bei Verschiebung und Skalierung gilt:

$$\begin{aligned}\rho^n(X + c) &= \mathbf{E}[|X + c - \mathbf{E}(X + c)|^n] = \mathbf{E}[|X + c - \mathbf{E}(X) - c|^n] \\ &= \mathbf{E}[|X - \mathbf{E}(X)|^n] = \rho^n(X)\end{aligned}$$

$$\begin{aligned}\rho^n(aX) &= \mathbf{E}[|aX - \mathbf{E}(aX)|^n] = \mathbf{E}[|aX - a\mathbf{E}(X)|^n] \\ &= |a|^n \mathbf{E}[|X - \mathbf{E}(X)|^n] = |a|^n \rho^n(X)\end{aligned}$$

(3) Speziell für $Y = X - \mu$ gilt somit $\mathbf{E}(Y) = 0$ und $\mathbf{V}(Y) = \mathbf{V}(X) = \sigma^2$.
Speziell für $Z = (X - \mu)/\sigma$ gilt somit $\mathbf{E}(Z) = 0$ und $\mathbf{V}(Z) = 1$.

😊 Wir nennen $Y = X - \mu$ die **Zentrierung** und $Z = (X - \mu)/\sigma$ die **Normierung** von X . Dies lässt sich umkehren zu $X = \mu + \sigma Z$.

😊 Für die Zufallsvariable X ist die natürliche Betrachtung die (μ, σ) -**Skala** mit Ursprung in μ und Längeneinheit $\sigma > 0$.

Erwartung und Varianz von Zufallsvariablen

Aufgabe: Bei zweimaligem Würfeln betrachten wir die Ergebnismenge $\Omega = \{1, 2, \dots, 6\}^2$ mit der Gleichverteilung \mathbf{P} und den Zufallsvariablen

$$X, Y : \Omega \rightarrow \mathbb{R} : \quad X(\omega_1, \omega_2) = \omega_1, \quad Y(\omega_1, \omega_2) = \omega_2.$$

Berechnen Sie Erwartung und Varianz der Zufallsvariablen X , Y , sowie $X - Y$, $|X - Y|$, schließlich $\max(X, Y)$ und $\min(X, Y)$.

Lösung: Die Verteilungen dieser ZVaren wurden oben ausgeführt. Wir nutzen direkt die Definition T3c. Für die Zufallsvariable X finden wir:

$$\mathbf{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

$$\mathbf{E}(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$$

$$\sigma(X) = \sqrt{35/12} \approx 1.70783$$

Die ZVariable Y ist verschieden von X , aber identisch verteilt: Erwartung, Varianz und Streuung sind demnach dieselben.

Erwartung und Varianz von Zufallsvariablen

😊 Dank Linearität T3D gilt $\mathbf{E}(X - Y) = \mathbf{E}(X) - \mathbf{E}(Y) = 0$. Also:

$$\begin{aligned}\mathbf{V}(X - Y) &= \mathbf{E}[(X - Y)^2] \\ &= 0^2 \cdot \frac{6}{36} + 1^2 \cdot \frac{10}{36} + 2^2 \cdot \frac{8}{36} + 3^2 \cdot \frac{6}{36} + 4^2 \cdot \frac{4}{36} + 5^2 \cdot \frac{2}{36} = \frac{210}{36} = \frac{35}{6} \\ \sigma(X - Y) &= \sqrt{35/6} \approx 2.41523\end{aligned}$$

😊 Das Ergebnis erhalten wir bequemer dank Unabhängigkeit (T3H) der beiden ZVaren X, Y mittels Additivität der Varianzen (T3I):

$$\mathbf{V}(X - Y) = \mathbf{V}(X) + \mathbf{V}(-Y) = \mathbf{V}(X) + \mathbf{V}(Y) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$$

Für die Zufallsvariable $|X - Y|$ finden wir entsprechend:

$$\begin{aligned}\mathbf{E}(|X - Y|) &= 0 \cdot \frac{6}{36} + 1 \cdot \frac{10}{36} + 2 \cdot \frac{8}{36} + 3 \cdot \frac{6}{36} + 4 \cdot \frac{4}{36} + 5 \cdot \frac{2}{36} = \frac{70}{36} = \frac{35}{18} \\ \mathbf{V}(|X - Y|) &= \mathbf{E}[|X - Y|^2] - \mathbf{E}[|X - Y|]^2 = \frac{35}{6} - \left(\frac{35}{18}\right)^2 = \frac{665}{324} \\ \sigma(|X - Y|) &= \sqrt{665/324} \approx 1.43264\end{aligned}$$

Die Rechnungen für $\max(X, Y)$ und $\min(X, Y)$ verlaufen genauso: Wir wenden die Definition T3C direkt auf die Verteilungen an.

Chebyshev: Große Abweichungen sind unwahrscheinlich!

Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit Erwartung $\mu = \mathbf{E}(X)$.

Anschaulich gilt: Die Werte schwanken zufällig um μ , und große Abweichungen vom Erwartungswert sind eher unwahrscheinlich.

Wie unwahrscheinlich? Wir fragen also nach $\mathbf{P}(|X - \mu| \geq c)$.

Die Streuung σ misst hierzu die typische Breite der Verteilung um μ .

Die Ungleichungen von Chebyshev geben bequeme Abschätzungen!

😊 Wir müssen hierzu über die Verteilung von X nichts weiter wissen als nur ihren Erwartungswert $\mu = \mathbf{E}(X)$ und ihre Varianz $\sigma^2 = \mathbf{V}(X)$! Dies sind hier (und überall) die beiden wichtigsten Kenngrößen.

😊 Selbst wenn Sie mehr wissen oder mehr herausfinden könnten, die Chebyshev–Abschätzungen sind allgemein, unschlagbar einfach und oft sehr nützlich als erste Überschlagsrechnung.

⚠️ Anschließend werden wir den zentralen Grenzwertsatz W1D kennen und nutzen lernen. Wenn der ZGS anwendbar ist, dann ermöglicht er präzisere Schätzungen. In diesen Fällen ist Chebyshev zwar einfacher, aber auch viel gröber und kann in Genauigkeit meist nicht konkurrieren.

Chebychev: Große Abweichungen sind unwahrscheinlich!

Satz T3F: die Ungleichungen von Chebychev, 1867

Sei $X : (\Omega, \mathbf{P}) \rightarrow \mathbb{R} : \omega \mapsto X(\omega)$ eine ZVariable mit Erwartung $\mu = \mathbf{E}(X)$ und Varianz $\sigma^2 = \mathbf{E}[(X - \mu)^2]$ bzw. n -tem Moment $\rho^n = \mathbf{E}[|X - \mu|^n]$. Für alle $c, k, h > 0$ gelten universelle „worst case“ Abschätzungen:

$$\begin{array}{ll}
 (1) & \mathbf{P} \left[|X - \mu| \geq c \right] \leq \frac{\sigma^2}{c^2} \text{ bzw. } \frac{\rho^2}{c^2} \\
 (2) & \mathbf{P} \left[|X - \mu| \geq k\sigma \right] \leq \frac{1}{k^2} \\
 (3) & \mathbf{P} \left[X \geq \mu + k\sigma \right] \leq \frac{1}{1 + k^2} \\
 (4) & \mathbf{P} \left[\mu - k\sigma < X < \mu + k\sigma \right] \geq 1 - \frac{1}{k^2} \\
 (5) & \mathbf{P} \left[\mu - h\sigma < X < \mu + k\sigma \right] \geq \frac{4(hk - 1)}{(h + k)^2}
 \end{array}$$

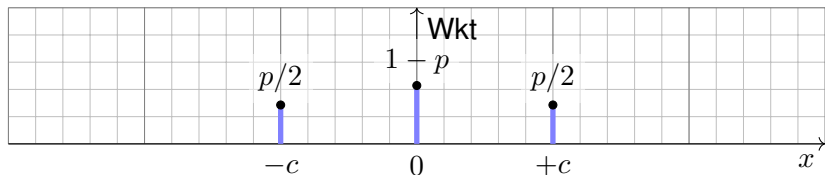
Die Wkt großer Abweichungen fällt mindestens quadratisch, mit $1/k^2$.
 Kennen Sie nur μ und σ bzw. ρ , so sind diese Ungleichungen optimal.

Beispiel zur beidseitigen Chebychev–Abschätzung

Die beidseitige Chebychev–Ungleichung (1,2) besagt:

$$\mathbf{P}[|X - \mu| \geq c] \leq \frac{\sigma^2}{c^2}$$

Aufgabe: Überprüfen Sie diese für folgende Verteilung. Ist sie scharf?



Lösung: Es gilt $\mu = 0$ und $\sigma^2 = c^2 p$. Die Ungleichung ist scharf:

$$p = \mathbf{P}[|X - \mu| \geq c] \leq \frac{\sigma^2}{c^2} = p$$

😊 Die Abschätzung ist meist grob, dafür aber universell einsetzbar. Unser Beispiel zeigt, dass sie sich allgemein nicht verbessern lässt.

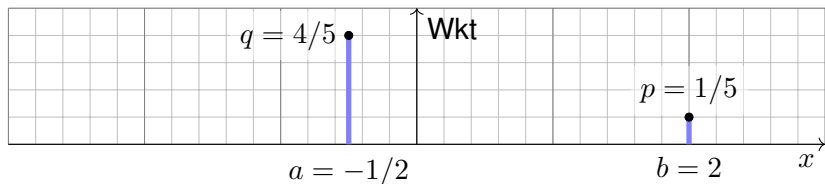
Beispiel: Speziell für $p = 1$ finden wir hier $\mathbf{P}[|X - \mu| \geq 1\sigma] = 100\%$.

Beispiel zur einseitigen Chebychev–Abschätzung

Die einseitige Chebychev–Ungleichung (3) für $a < \mu < b$ besagt:

$$\mathbf{P}[X \leq a] \leq \frac{1}{1 + \left(\frac{a-\mu}{\sigma}\right)^2} \quad \text{und} \quad \mathbf{P}[X \geq b] \leq \frac{1}{1 + \left(\frac{b-\mu}{\sigma}\right)^2}$$

Aufgabe: Überprüfen Sie beide für folgende Verteilung. Sind sie scharf?



Lösung: Hier gilt $\mu = 0$ und $\sigma^2 = 1$. Die Ungleichungen sind scharf:

$$q = \mathbf{P}[X \leq a] \leq \frac{1}{1 + a^2} = \frac{4}{5} \quad \text{bzw.} \quad p = \mathbf{P}[X \geq b] \leq \frac{1}{1 + b^2} = \frac{1}{5}.$$

😊 Die Abschätzung ist meist grob, dafür aber universell einsetzbar. Unser Beispiel zeigt, dass sie sich allgemein nicht verbessern lässt.

Übung: Erfinden und untersuchen Sie weitere einfache Beispiele.

Chebychev–Abschätzungen auf der (μ, σ) –Skala

Aufgabe: Was sagt Chebychev für folgende typische Abschätzungen?

$$\mathbf{P}[|X - \mu| \geq 1\sigma] \leq 1/1 = 100\%$$

$$\mathbf{P}[|X - \mu| \geq 2\sigma] \leq 1/4 = 25\%$$

$$\mathbf{P}[|X - \mu| \geq 3\sigma] \leq 1/9 \approx 11\%$$

$$\mathbf{P}[|X - \mu| \geq 4\sigma] \leq 1/16 \approx 6\%$$

$$\mathbf{P}[X \geq \mu + 1\sigma] \leq 1/2 = 50\%$$

$$\mathbf{P}[X \geq \mu + 2\sigma] \leq 1/5 = 20\%$$

$$\mathbf{P}[X \geq \mu + 3\sigma] \leq 1/10 = 10\%$$

$$\mathbf{P}[X \geq \mu + 4\sigma] \leq 1/17 \approx 6\%$$

☺ Große Abweichungen vom Erwartungswert sind unwahrscheinlich. Chebychevs Abschätzungen sind grob, dafür aber universell einsetzbar.

Die erste Abschätzung ist trivialerweise immer richtig, aber nie hilfreich.

Mindestens $3/4$ der Wkt liegt in der 2σ –Umgebung $[\mu - 2\sigma, \mu + 2\sigma]$.

Mindestens $8/9$ der Wkt liegt in der 3σ –Umgebung $[\mu - 3\sigma, \mu + 3\sigma]$.

☺ Für Normalverteilungen gilt die genauere 68–95–99–Regel. V120

Beweis der beidseitigen Chebychev–Abschätzung

Nachrechnen: Die Zufallsvariable X zentrieren wir zu $Z := X - \mu$. Ihr n -tes Moment ist $\rho^n = \mathbf{E}[|Z|^n]$; für $n = 2$ ist dies die Varianz σ^2 . Wir vergleichen Z mit der zweiwertigen Zufallsvariablen $Y = c \mathbf{I}_{\{|Z| \geq c\}}$,

$$Y : \Omega \rightarrow \mathbb{R} : \omega \mapsto Y(\omega) = \begin{cases} c & \text{falls } |Z(\omega)| \geq c, \\ 0 & \text{falls } |Z(\omega)| < c. \end{cases}$$

Es gilt $|Z| \geq Y \geq 0$. Dank Monotonie der Erwartung folgt

$$\rho^n = \mathbf{E}[|Z|^n] \geq \mathbf{E}[Y^n] = c^n \cdot \mathbf{P}[|X - \mu| \geq c].$$

😊 Hieraus folgt (1). Speziell für $n = 2$ und $c = k\sigma$ erhalten wir (2). Für (3) zeigen wir als Hilfsmittel zunächst die Markov–Ungleichung:

Lemma T3G: Markov–Ungleichung

Sei $Y : (\Omega, \mathbf{P}) \rightarrow [0, \infty]$ eine nicht-negative Zufallsvariable. Für alle $b > 0$ gilt dann die Abschätzung $\mathbf{P}[Y \geq b] \leq \mathbf{E}[Y]/b$.

Beweis: Es gilt $Y \geq b \mathbf{I}_{\{Y \geq b\}}$, also folgt $\mathbf{E}[Y] \geq b \cdot \mathbf{P}[Y \geq b]$.

😊 Hieraus erhalten wir, nach Rechnung, die Ungleichungen (3–5).

Beweis der einseitigen Chebychev–Abschätzung

Aufgabe: Beweisen Sie mit Lemma T3G die Ungleichungen (3) und (5).

Lösung: Wir beweisen die einseitige Chebychev–Ungleichung (3).

Wie zuvor zentrieren wir die Zufallsvariable X gemäß $Z := X - \mu$.

Für alle $t, b \in \mathbb{R}$ mit $t + b > 0$ gelten folgende Ungleichungen:

$$\begin{aligned} \mathbf{P}[Z \geq b] &= \mathbf{P}[Z + t \geq b + t] \\ &\leq \mathbf{P}[(Z + t)^2 \geq (b + t)^2] \stackrel{\text{Markov}}{\leq} \frac{\mathbf{E}[(Z + t)^2]}{(b + t)^2} = \frac{\sigma^2 + t^2}{(b + t)^2} \end{aligned}$$

Die erste Ungleichung ist klar, die zweite ist die Markov–Ungleichung.

Die letzte Gleichung folgt dank Zentrierung $\mathbf{E}[Z] = 0$, denn damit gilt

$$\mathbf{E}[(Z + t)^2] = \mathbf{E}[Z^2] + 2\mathbf{E}[Z]t + t^2 = \sigma^2 + t^2.$$

Für $b = k\sigma$ und $t = \sigma/k$ erhalten wir die gewünschte Ungleichung (3):

$$\mathbf{P}[X - \mu \geq k\sigma] = \frac{\sigma^2 + t^2}{(k\sigma + t)^2} = \frac{\sigma^2 + \sigma^2/k^2}{(k\sigma + \sigma/k)^2} = \frac{k^2 + 1}{(k^2 + 1)^2} = \frac{1}{k^2 + 1}$$

😊 Die Rechnung ist etwas trickreich, aber jeder Schritt ist elementar.

Beweis der asymmetrischen Abschätzung

Ungleichung (4) folgt direkt aus (2); sie betont hier die Analogie zur asymmetrischen Ungleichung (5). Für $a < \mu < b$ ist diese äquivalent zu:

$$\mathbf{P}[a < X < b] \geq \frac{4[(b - \mu)(\mu - a) - \sigma^2]}{(b - a)^2},$$

$$\mathbf{P}[|X - m| \geq c] \leq \frac{\sigma^2 + (m - \mu)^2}{c^2}.$$

Zur Umformulierung setzen wir $m = (a + b)/2$ und $c = (b - a)/2$.

Für $m = \mu$ ergibt sich die symmetrische Chebychev–Ungleichung.

Die letzte Ungleichung folgt erneut dank Markov–Ungleichung T3G:

$$\mathbf{P}[|X - m| \geq c] = \mathbf{P}[|X - m|^2 \geq c^2] \leq \frac{\mathbf{E}[|X - m|^2]}{c^2} = \frac{\sigma^2 + (m - \mu)^2}{c^2}$$

Auch diese Abschätzung lässt sich nicht verbessern. Obiges Beispiel:

Zu $0 < t < 1$ sei $b = \sqrt{(1 - t)/t}$ und $a = -1/b = -\sqrt{t/(1 - t)}$ sowie

$$\mathbf{P}(X=a) = 1 - t \quad \text{und} \quad \mathbf{P}(X=b) = t.$$

Es gilt $\mathbf{E}(X) = a(1 - t) + bt = 0$ und $\mathbf{V}(X) = a^2(1 - t) + b^2t = 1$.

In diesem Beispiel sind die obigen Ungleichungen scharf.

Unabhängigkeit von Zufallsvariablen

Zwei Zufallsvariablen X_1, X_2 sind unabhängig, wenn das Ergebnis von X_1 nicht die Wkten von X_2 beeinflusst. Dies präzisieren wir wie folgt:

Definition T3H: Unabhängigkeit von Zufallsvariablen

Eine Familie $X = (X_1, \dots, X_n)$ von Zufallsvariablen $X_k : \Omega \rightarrow \mathbb{R}$ heißt **(stochastisch) unabhängig**, wenn für alle Intervalle $I_1, \dots, I_n \subseteq \mathbb{R}$ gilt

$$\mathbf{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbf{P}(X_1 \in I_1) \cdots \mathbf{P}(X_n \in I_n).$$

Anders gesagt: Im WRaum (Ω, \mathbf{P}) ist die Familie (A_1, \dots, A_n) der Ereignisse $A_k = X_k^{-1}(I_k) = \{\omega \in \Omega \mid X_k(\omega) \in I_k\}$ unabhängig (T2C). Das bedeutet: Die gemeinsame Verteilung von (X_1, \dots, X_n) in \mathbb{R}^n ist die Produktverteilung $\mathbf{P}_{(X_1, \dots, X_n)} = \mathbf{P}_{X_1} \otimes \cdots \otimes \mathbf{P}_{X_n}$, siehe U1A und W1A. Diese Formulierungen gelten ebenso für kontinuierliche Verteilungen, zum Beispiel auf einem analogen Messgerät abgelesene Werte. Für **diskrete Zufallsvariablen** ist Unabhängigkeit äquivalent zu

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbf{P}(X_1 = a_1) \cdots \mathbf{P}(X_n = a_n).$$

Unabhängigkeit von Zufallsvariablen

Aufgabe: Bei zweimaligem Würfeln betrachten wir die Ergebnismenge $\Omega = \{1, 2, \dots, 6\}^2$ mit der Gleichverteilung \mathbf{P} und den Zufallsvariablen

$$X, Y : \Omega \rightarrow \mathbb{R} : \quad X(\omega_1, \omega_2) = \omega_1, \quad Y(\omega_1, \omega_2) = \omega_2.$$

- (1) Sind X und Y stochastisch unabhängig?
- (2) Sind $X + Y$ und $X - Y$ stochastisch unabhängig?
- (3) Sind $\min(X, Y)$ und $\max(X, Y)$ stochastisch unabhängig?


Lösung: (1) Ja, denn $\mathbf{P}(X=a, Y=b) = \mathbf{P}(X=a) \cdot \mathbf{P}(Y=b)$.

(2) Nein, die Multiplikativität gilt hier nicht:

Zum Beispiel gilt $\mathbf{P}(X+Y=3, X-Y=0) = 0$,
aber $\mathbf{P}(X+Y=3) = 2/36$ und $\mathbf{P}(X-Y=0) = 6/36$.

(3) Nein, die Multiplikativität gilt hier nicht:

Zum Beispiel gilt $\mathbf{P}(\min(X, Y)=1, \max(X, Y)=6) = 2/36$,
aber $\mathbf{P}(\min(X, Y)=1) = 7/36$ und $\mathbf{P}(\max(X, Y)=6) = 7/36$.

 Zum Nachweis der Unabhängigkeit muss man *alle* Produkte prüfen.
Zum Nachweis der Abhängigkeit hingegen genügt *ein* Gegenbeispiel.

Unabhängigkeit von Zufallsvariablen

Aufgabe: Seien $X_1, \dots, X_n : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ diskrete Zufallsvariablen; ihre Zusammenfassung $X = (X_1, \dots, X_n) : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}^n$ ist ebenso diskret. Zeigen Sie die Äquivalenz der beiden Unabhängigkeitsbedingungen:

$$(1) \quad \mathbf{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbf{P}(X_1 \in I_1) \cdots \mathbf{P}(X_n \in I_n)$$

für alle Intervalle $I_1, \dots, I_n \subseteq \mathbb{R}$.

$$(2) \quad \mathbf{P}(X_1 = a_1, \dots, X_n = a_n) = \mathbf{P}(X_1 = a_1) \cdots \mathbf{P}(X_n = a_n)$$

für jeden Punkt $a = (a_1, \dots, a_n) \in \mathbb{R}^n$.

😊 Bedingung (1) ist vielseitiger und bequemer anzuwenden. Zudem ist dies die richtige Definition auch für kontinuierlich verteilte Zufallsvariablen. Daher erheben wir Eigenschaft (1) zur Definition T3H.

😊 Bedingung (2) nützt nur für diskrete Zufallsvariablen, ist dann aber leichter zu prüfen. Genau hierfür nutzen wir die Äquivalenz „(1) \Leftrightarrow (2)“. Erste Beispiele kennen wir aus der vorigen Aufgabe.

Unabhängigkeit von Zufallsvariablen

Lösung: Die Implikation „(1) \Rightarrow (2)“ ist eine Einschränkung, also trivial: Hierzu genügt es, die einpunktigen Intervalle $I_k = \{a_k\}$ zu betrachten.

Wir zeigen die Umkehrung „(2) \Rightarrow (1)“: Hierzu sei $I = I_1 \times \dots \times I_n \subset \mathbb{R}^n$ ein Quader und darin $a = (a_1, \dots, a_n) \in I$ ein Punkt. Wir finden:

$$\begin{aligned}
 \mathbf{P}(X_1 \in I_1, \dots, X_n \in I_n) &= \mathbf{P}(X \in I) \\
 &\stackrel{\substack{X \\ \text{diskret}}}{=} \sum_{a \in I} \mathbf{P}(X = a) \\
 &\stackrel{I}{=} \sum_{a_1 \in I_1} \dots \sum_{a_n \in I_n} \mathbf{P}(X_1 = a_1, \dots, X_n = a_n) \\
 &\stackrel{(2)}{\text{punktweise}} \sum_{a_1 \in I_1} \dots \sum_{a_n \in I_n} \mathbf{P}(X_1 = a_1) \dots \mathbf{P}(X_n = a_n) \\
 &\stackrel{\substack{\text{Summe} \\ \text{distributiv}}}{=} \left[\sum_{a_1 \in I_1} \mathbf{P}(X_1 = a_1) \right] \dots \left[\sum_{a_n \in I_n} \mathbf{P}(X_n = a_n) \right] \\
 &\stackrel{\substack{X_i \\ \text{diskret}}}{=} \mathbf{P}(X_1 \in I_1) \dots \mathbf{P}(X_n \in I_n)
 \end{aligned}$$

Unabhängigkeit von Indikatorfunktionen

Aufgabe: (1) Wann sind zu zwei Ereignissen $A, B \subseteq \Omega$ die Indikatorfunktionen $X = \mathbf{I}_A$ und $Y = \mathbf{I}_B$ unabhängig?

(2) Wann gilt Multiplikativität $\mathbf{E}(\mathbf{I}_A \cdot \mathbf{I}_B) = \mathbf{E}(\mathbf{I}_A) \cdot \mathbf{E}(\mathbf{I}_B)$?

Lösung: (1) Sei $A \subseteq \Omega$ ein Ereignis und \mathbf{I}_A seine Indikatorfunktion,

$$\mathbf{I}_A : \Omega \rightarrow \mathbb{R} : \omega \mapsto \mathbf{I}_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A, \\ 0 & \text{falls } \omega \notin A. \end{cases}$$

Unabhängigkeit von \mathbf{I}_A und \mathbf{I}_B bedeutet vier äquivalente Bedingungen:

$$\mathbf{P}(\mathbf{I}_A=1, \mathbf{I}_B=1) = \mathbf{P}(A \cap B) \stackrel{!}{=} \mathbf{P}(\mathbf{I}_A=1) \mathbf{P}(\mathbf{I}_B=1) = \mathbf{P}(A) \mathbf{P}(B)$$

$$\mathbf{P}(\mathbf{I}_A=1, \mathbf{I}_B=0) = \mathbf{P}(A \cap \bar{B}) \stackrel{!}{=} \mathbf{P}(\mathbf{I}_A=1) \mathbf{P}(\mathbf{I}_B=0) = \mathbf{P}(A) \mathbf{P}(\bar{B})$$

$$\mathbf{P}(\mathbf{I}_A=0, \mathbf{I}_B=1) = \mathbf{P}(\bar{A} \cap B) \stackrel{!}{=} \mathbf{P}(\mathbf{I}_A=0) \mathbf{P}(\mathbf{I}_B=1) = \mathbf{P}(\bar{A}) \mathbf{P}(B)$$

$$\mathbf{P}(\mathbf{I}_A=0, \mathbf{I}_B=0) = \mathbf{P}(\bar{A} \cap \bar{B}) \stackrel{!}{=} \mathbf{P}(\mathbf{I}_A=0) \mathbf{P}(\mathbf{I}_B=0) = \mathbf{P}(\bar{A}) \mathbf{P}(\bar{B})$$

Dies ist äquivalent zur Unabhängigkeit der Ereignisse A und B ! (T2C)
(Diese vier Gleichungen sind untereinander äquivalent, siehe T2D.)

Unabhängigkeit von Indikatorfunktionen

(2) Die Indikatorfunktion \mathbf{I}_A nimmt nur die beiden Werte 0 und 1 an:

$$\mathbf{P}(\mathbf{I}_A=1) = \mathbf{P}(A) \quad \text{und} \quad \mathbf{P}(\mathbf{I}_A=0) = 1 - \mathbf{P}(A)$$

Demnach gilt $\mathbf{E}(\mathbf{I}_A) = \mathbf{P}(A)$. Für das Produkt von \mathbf{I}_A und \mathbf{I}_B gilt

$$\mathbf{I}_A \cdot \mathbf{I}_B = \mathbf{I}_{A \cap B}.$$

Wir vergleichen damit schließlich

$$\begin{aligned} \mathbf{E}(\mathbf{I}_A \cdot \mathbf{I}_B) &= \mathbf{E}(\mathbf{I}_{A \cap B}) = \mathbf{P}(A \cap B), \\ \mathbf{E}(\mathbf{I}_A) \cdot \mathbf{E}(\mathbf{I}_B) &= \mathbf{P}(A) \cdot \mathbf{P}(B). \end{aligned}$$

😊 Die Gleichheit links ist demnach äquivalent zur Gleichheit rechts: Die Zufallsvariablen $\mathbf{I}_A, \mathbf{I}_B : \Omega \rightarrow \mathbb{R}$ sind genau dann unabhängig (T3H), wenn die betrachteten Ereignisse A, B in (Ω, \mathbf{P}) unabhängig sind (T2c). Nur dann ist ihre Erwartung multiplikativ, also $\mathbf{E}(\mathbf{I}_A \cdot \mathbf{I}_B) = \mathbf{E}(\mathbf{I}_A) \cdot \mathbf{E}(\mathbf{I}_B)$.

😊 Für Indikatorfunktionen ist die Rechnung besonders einfach. Allgemein ist die Sachlage noch interessanter, siehe Satz T3i.

Erwartung und Varianz bei Summen und Produkten

Aufgabe: (0) Ist die Erwartung additiv, $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$?

(1) Ist die Erwartung multiplikativ, $\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$?

(2) Ist die Varianz additiv, $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$?

Lösung: (0) Ja! Die Erwartung $\mathbf{E} : L^1(\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ ist linear (Satz T3D).

(1) Nein! Gegenbeispiel $X : \Omega \rightarrow \{\pm 1\}$ mit $\mathbf{P}(X = \pm 1) = 1/2$: Hier gilt

$$\mathbf{E}(X \cdot X) = \mathbf{E}(1) = 1 \neq 0 = \mathbf{E}(X) \cdot \mathbf{E}(X).$$


 Die Erwartung ist immer additiv, aber i.A. nicht multiplikativ!

Alternativ: $\mathbf{E}(\mathbf{I}_A \cdot \mathbf{I}_B) = \mathbf{E}(\mathbf{I}_A) \cdot \mathbf{E}(\mathbf{I}_B)$ gilt nur für unabhängige A, B .

Allgemein: Die Varianz ist $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 \geq 0$ und i.A. > 0 .

(2) Wir setzen die Definition ein und vereinfachen geduldig:

$$\begin{aligned} \mathbf{V}(X+Y) &\stackrel{\text{Def}}{=} \mathbf{E}[(X+Y - \mathbf{E}(X+Y))^2] \stackrel{\text{Lin}}{=} \mathbf{E}[(X - \mathbf{E}(X) + Y - \mathbf{E}(Y))^2] \\ &\stackrel{\text{Bin}}{=} \mathbf{E}[(X - \mathbf{E}(X))^2] + \mathbf{E}[(Y - \mathbf{E}(Y))^2] + 2[\underbrace{\mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)}_{=: \text{Cov}(X, Y)}] \\ &\stackrel{\text{Def}}{=} \mathbf{V}(X) + \mathbf{V}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

 Die Varianzen addieren sich nur, falls die Kovarianz verschwindet.

 Dies gilt glücklicherweise bei **unabhängigen** Zufallsvariablen!

Satz T31: Fubini für unabhängige Zufallsvariablen

(1) Erwartungen unabhängiger Zufallsvariablen X, Y multiplizieren sich:

$$\mathbf{E}(X \cdot Y) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$$

Ihre Kovarianz $\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ verschwindet somit.

(2) Varianzen unabhängiger Zufallsvariablen X, Y addieren sich:

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X, Y)$$

Aufgabe: Rechnen Sie dies nach! **Lösung:** (1) Dank Unabhängigkeit:

$$\begin{aligned} \mathbf{E}(X \cdot Y) &\stackrel{\text{Def}}{=} \sum_{(a,b) \in \mathbb{R}^2} ab \mathbf{P}(X=a, Y=b) \\ &\stackrel{\text{Fub}}{\stackrel{\text{Un}}{=}} \sum_{a \in \mathbb{R}} \sum_{b \in \mathbb{R}} ab \mathbf{P}(X=a) \mathbf{P}(Y=b) \\ &\stackrel{\text{Lin}}{\stackrel{\text{Dist}}{=}} \left[\sum_{a \in \mathbb{R}} a \mathbf{P}(X=a) \right] \cdot \left[\sum_{b \in \mathbb{R}} b \mathbf{P}(Y=b) \right] \stackrel{\text{Def}}{=} \mathbf{E}(X) \cdot \mathbf{E}(Y) \end{aligned}$$

Erwartung und Varianz unabhängiger ZVaren

- Aufgabe:** (1) Einzelner Münzwurf: Berechnen Sie Erwartung und Varianz eines 0–1–Experiments $X_k : \Omega \rightarrow \{0, 1\}$ mit $\mathbf{P}(X_k=1) = t_k$.
- (2) Mehrere Münzwürfe: Bestimmen Sie Erwartung und Varianz der Trefferzahl $S = X_1 + \dots + X_n$. Welche Zusatzinformation brauchen Sie?

- Lösung:** (1) Wir erhalten $\mathbf{E}(X_k) = t_k$ und $\mathbf{V}(X_k) = t_k(1 - t_k)$.
- (2) Dank **Linearität** der Erwartung (T3D) gilt hier Additivität:

$$\mathbf{E}(S) = \mathbf{E}(X_1 + \dots + X_n) = \mathbf{E}(X_1) + \dots + \mathbf{E}(X_n) = t_1 + \dots + t_n$$

Sind die Zufallsvariablen X_1, \dots, X_n zudem **unabhängig**, so addieren sich bei der Summe $S = X_1 + \dots + X_n$ auch ihre Varianzen (T3I):

$$\mathbf{V}(S) = \mathbf{V}(X_1) + \dots + \mathbf{V}(X_n) = t_1(1 - t_1) + \dots + t_n(1 - t_n)$$

😊 Speziell für $t_1 = \dots = t_n = t$ gilt $\mathbf{E}(S) = nt$ und $\mathbf{V}(S) = nt(1 - t)$. Das ist der wichtige Spezialfall von unabhängigen, identisch verteilten 0–1–Experimenten (U3B): Die Summe ist binomialverteilt, $S \sim B(n, t)$.

Beispiel: Punkteverteilung in einer Klausur? Gilt hier Unabhängigkeit?

Erwartung und Varianz unabhängiger ZVaren

😊 Unabhängig wiederholte Experimente sind eine zentrale Methode der Natur- und Ingenieurwissenschaften. Das stochastische Modell (zur Planung) und die statistische Auswertung (nach Durchführung) gehören zum allgegenwärtigen Handwerkszeug seriöser Arbeit.

(1) Der Münzwurf ist das einfachste Zufallsexperiment: Es gibt nur zwei Ergebnisse: Kopf (0) oder Zahl (1). Wir nehmen hier nicht zwangsweise an, dass die Münze fair ist, also $\mathbf{P}(X_k=0) = \mathbf{P}(X_k=1) = 1/2$ erfüllt.

Die Wkt $\mathbf{P}(X_k=1) = t_k \in [0, 1]$ ist in unserem Modell frei wählbar.

(2) Wenn wir n Münzen werfen, so addieren sich die Erwartungen und dank Unabhängigkeit auch die Varianzen. (Für Letzteres genügt bereits die Unkorreliertheit, das ist i.A. schwächer.) Die Summe unabhängiger zweiwertiger Zufallsvariablen ist ein weiteres einfaches Modell.

(3) Die n -fache Wiederholung eines Münzwurfs ist ein wichtiges Modell: Zur Vereinfachung nehmen wir an, das Experiment wird unabhängig wiederholt und alle Wkten sind gleich. Dies führt zur Binomialverteilung U3B und zum lokalen Grenzwertsatz V3A als wichtiger Näherung.

Anwendungsbeispiel: Überbuchung

Aufgabe: Für eine Klausur sind $n = 320$ Teilnehmer angemeldet. Jeder davon tritt nur mit Wahrscheinlichkeit $t = 0.8$ tatsächlich an. Wir nehmen stochastische Unabhängigkeit an, keine Grippewelle o.ä. Mit welcher Wkt reichen 300 Plätze? (Chebychev, später LGS, exakt)

Lösung: Wir wissen $\mu = nt = 256$, $\sigma^2 = nt(1-t) = 51.2$, $\sigma \approx 7.2$.

Chebychev: Die Wkt für mehr als 300 Teilnehmer ist beschränkt durch

$$\mathbf{P}(T \geq 301) = \mathbf{P}(T - 256 \geq 45) \leq \frac{1}{1+k^2} \lesssim 2.5\% \quad \text{mit} \quad k = \frac{45}{7.2} \approx 6.22.$$

Die Schätzung 3% ist deutlich zu hoch, dafür aber bequem zu berechnen. Wir wissen über T noch viel mehr als nur μ und σ : Wir kennen die gesamte Verteilung! Die Binomialverteilung $B(n, t)$ erlaubt dank lokalem Grenzwertsatz V3A eine wesentlich genauere Abschätzung:

LGS V3A: Normalverteilung als Näherung, Tabelle: $\int_k^\infty \varphi(t) dt < 10^{-5}$,
Approximationsfehler: $|\delta| < 1/(6\sigma) < 0.023$, insgesamt Wkt $< 2.3\%$.

Exakt U3B: Die Teilnehmerzahl T ist binomialverteilt, genauer $T \sim B(n, t)$.
Durch Aufsummieren (mit Computerhilfe) erhalten wir $\mathbf{P}(T \geq 301) \lesssim 0.9 \cdot 10^{-12}$.

Es gibt $2^{320} > 2 \cdot 10^{96}$ mögliche Ausgänge dieses Experiments. Das kann offensichtlich niemand naiv aufsummieren, selbst der schnellste Computer nicht. Die exakte Rechnung organisieren wir daher wesentlich effizienter mit den Abzählmethoden des folgenden Kapitels.

Anwendungsbeispiel: Überbuchung

Aufgabe: Angenommen, im obigen Beispiel gilt $n = 320$ und $t = 0.9$. Mit welcher Wkt reichen 300 Plätze? (Chebychev, LGS, exakt)

Lösung: Wir wissen $\mu = nt = 288$, $\sigma^2 = nt(1-t) = 28.8$, $\sigma \approx 5.4$.

Chebychev: Die Wkt für mehr als 300 Teilnehmer ist beschränkt durch

$$\mathbf{P}(T \geq 301) = \mathbf{P}(T - 288 \geq 13) \leq \frac{1}{1+k^2} \lesssim 15\% \quad \text{mit} \quad k = \frac{13}{5.4} \approx 2.33.$$

Diese Größenordnung ist alarmierend: Wir sollten einen größeren Hörsaal buchen — oder genauer hinschauen! Die Binomialverteilung erlaubt eine wesentlich genauere Abschätzung:

LGS V3A: Normalverteilung als Näherung, Tabelle: $\int_k^\infty \varphi(t) dt < 0.01$,

Approximationsfehler: $|\delta| \leq 1/(6\sigma) < 0.04$, insgesamt Wkt $< 5\%$.

Exakt U3B: Die Teilnehmerzahl T ist binomialverteilt, genauer $T \sim B(n, t)$.

Durch Aufsummieren (mit Computerhilfe) erhalten wir $\mathbf{P}(T \geq 301) \lesssim 0.00684$.

Die Ungleichung von Chebychev ist vor allem in Situationen nützlich, in denen wir über die Verteilung wenig wissen (z.B. nur μ und σ), aber die Wahrscheinlichkeit großer Abweichungen abschätzen wollen. In obigen Beispielen kennen wir die Verteilung sogar genau. Der lokale Grenzwertsatz V3A und die Tabelle der Normalverteilung liefern hier präzisere Abschätzungen.

Solche Risikoabschätzungen und Optimierungen haben enorme wirtschaftliche Bedeutung. Genau dieses Szenario analysieren Fluggesellschaften bei der Überbuchung ihrer Flüge.

Was gewinnen wir durch wiederholte Messungen?

Anwendung: Jede Messung X ist zufälligen Fehlern unterworfen, Messwerte schwanken um den Erwartungswert μ mit Streuung $\sigma > 0$.

Beispiel: Sie führen $n = 400$ unabhängige Messungen X_1, X_2, \dots, X_{400} durch und erhalten Messwerte x_1, x_2, \dots, x_{400} . Sie schätzen μ durch den

Stichprobenmittelwert	$\hat{x} := \frac{1}{n} \sum_{k=1}^n x_k$	Messung = (Beispiel)	57.5
------------------------------	---	-------------------------	------

Die Streuung von X schätzen Sie (dank der Stichprobe) auf $\sigma \approx 10$. Wie genau ist \hat{x} ? Wo liegt vermutlich der wahre Erwartungswert μ ?

$$I_1 = [57, 58], \quad I_2 = [56.5, 58.5], \quad I_3 = [56, 59], \quad I_k = [57.5 \pm k \cdot 0.5]$$

Aufgabe: Welches Intervall I_k überdeckt den Erwartungswert μ mit Sicherheit $\alpha_k \geq 95\%$ bzw. $\alpha_k \geq 99\%$? (Später genauer mit ZGS, W201)

Der Erwartungswert μ ist eine Konstante. Leider ist sie uns unbekannt. Das Ziel der Messung ist, diese Konstante μ möglichst genau zu bestimmen. Als Schätzung für den Wert μ dient uns \hat{x} . Wie die Messwerte x_1, \dots, x_n ist auch \hat{x} zufälligen Schwankungen unterworfen. Wenn Sie eine neue unabhängige Stichprobe durchführen, erhalten Sie einen neuen Mittelwert \hat{x} , und mit ihm verschieben sich obige Intervalle. Wir wollen sichergehen und verlangen, dass in 95% aller Stichproben unser Intervall den wahren Wert μ überdecken soll. Welchem Intervall vertrauen wir?

Was gewinnen wir durch wiederholte Messungen?

Lösung: Die Einzelmessung entspricht einer ZVariable $X : \Omega \rightarrow \mathbb{R}$ mit Erwartung $\mu = \mathbf{E}(X)$, Varianz $\sigma^2 = \mathbf{V}(X)$ und Streuung $\sigma(X) \in]0, \infty[$.

(1) Aus unabhängigen Wiederholungen X_1, X_2, \dots, X_n bilden wir den

$$\text{empirischen Mittelwert} \quad \hat{X} := \frac{1}{n}(X_1 + X_2 + \dots + X_n).$$

Dank Linearität gilt $\mathbf{E}(\hat{X}) = \mathbf{E}(X) = \mu$ und dank Unabhängigkeit zudem

$$\mathbf{V}(\hat{X}) = \frac{1}{n^2} \mathbf{V}(X_1 + \dots + X_n) = \frac{1}{n^2} [\mathbf{V}(X_1) + \dots + \mathbf{V}(X_n)] = \frac{1}{n} \mathbf{V}(X).$$

Demnach gilt $\sigma(\hat{X}) = \sigma(X)/\sqrt{n}$, im Beispiel $\sigma(\hat{X}) \lesssim 10/\sqrt{400} = 0.5$.

(2) Für unsere Intervalle gelten die Chebychev–Abschätzungen:

$$\alpha_1 \geq 1 - 1/1^2 = 0, \quad \alpha_2 \geq 1 - 1/2^2 \geq 75\%,$$

$$\alpha_3 \geq 1 - 1/3^2 > 88\%, \quad \alpha_5 \geq 1 - 1/5^2 \geq 96\%,$$

$$\alpha_{10} \geq 1 - 1/10^2 \geq 99\%, \quad \alpha_k \geq 1 - 1/k^2.$$

😊 Diese genial-einfache Methode optimieren wir in Kapitel V und W: Unser Ziel sind schärfere Abschätzungen, also genauere Intervalle!

Was gewinnen wir durch wiederholte Messungen?

Planung: Vor der Messung sind $X_1, X_2, \dots, X_n : \Omega \rightarrow \mathbb{R}$ ZVaren. Wir kennen die Messergebnisse noch nicht, noch sind für X_j alle Werte möglich mit ihren jeweiligen Wahrscheinlichkeiten $\mathbf{P}(X_j \in I)$; daher ist es hier sinnvoll, von Erwartung $\mathbf{E}(X_j)$ und Varianz $\mathbf{V}(X_j)$ zu sprechen. Daraus berechnen wir die Zufallsvariable \hat{X} und schließlich das Intervall $I_k = [\hat{X} \pm k\sigma]$. Mit \hat{X} schwankt auch I_k . Wir wollen unser *Verfahren* (!) so einrichten, dass I_k mit hoher Wkt den gesuchten Wert μ überdeckt.

Auswertung: Nach der Messung liegen Werte $x_1, x_2, \dots, x_n \in \mathbb{R}$ vor. Dies sind keine Zufallsvariablen, sondern reelle Zahlen. Es hat daher keinen Sinn, von Erwartung $\mathbf{E}(x_k)$ oder Varianz $\mathbf{V}(x_k)$ zu sprechen. Auch das Intervall $I_k = [\hat{x} \pm k\sigma]$ ist jetzt fixiert. Entweder es überdeckt μ oder eben nicht. Es hat somit keinen Sinn, hier nach Wkt zu fragen.

Reproduzierbarkeit: Nach der Messung ist vor der Messung!

Wir wollen garantieren, dass unsere Schätzungs*methode* zukünftigen, unabhängig wiederholten Messungen standhält. Klarheit und Korrektheit benötigen solide Grundlagen und sorgfältige Ausführung!

Das Gesetz der großen Zahlen

😊 Dieses Argument und die Rechnung fassen wir als Satz zusammen:

Satz T3J: Gesetz der großen Zahlen (nach Chebychev)

Seien $X_1, \dots, X_n: (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ ZVariablen mit Erwartung $\mu = \mathbf{E}(X)$ und Varianz $\sigma^2 = \mathbf{V}(X)$ und zudem unabhängig (es genügt unkorreliert).

Der empirische Mittelwert $\hat{X} = \frac{1}{n} \sum_{k=1}^n X_k$ nähert sich der Erwartung μ :

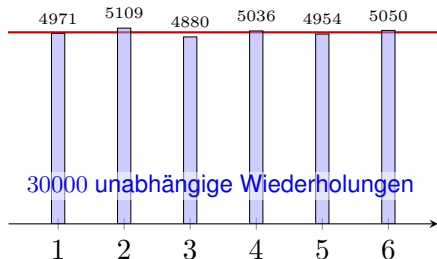
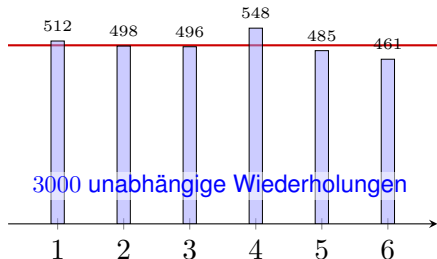
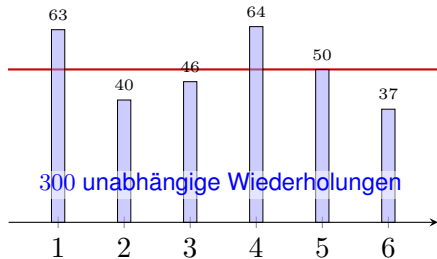
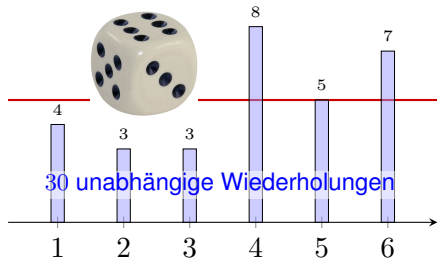
- (1) Seine Streuung $\sigma(\hat{X}) = \sigma(X)/\sqrt{n} \rightarrow 0$ fällt wie $1/\sqrt{n}$ für $n \rightarrow \infty$.
- (2) Abweichungen um $\varepsilon > 0$ werden beliebig unwahrscheinlich gemäß

$$\mathbf{P} \left[|\hat{X} - \mu| \geq \varepsilon \right] \leq \frac{\sigma^2}{n \varepsilon^2} \searrow 0.$$

Deshalb wiederholen wir Messungen *unabhängig*, um die Messgenauigkeit zu verbessern! Das Gesetz der großen Zahlen erklärt insbesondere, wie wir Wahrscheinlichkeiten messen können: Die Indikatorfunktion $X = \mathbf{I}_A$ zählt das Eintreten des Ereignisses A . Der Mittelwert \hat{X} ist die relative Häufigkeit nach n Versuchen und konvergiert gegen $\mu = \mathbf{E}(X) = \mathbf{P}(A)$. Alles wird gut!

⚠ Für praktische Anwendungen wollen Sie n nicht zu groß wählen: Wiederholte Messungen sind teuer. Chebychev T3F ist zwar einfach, aber noch recht grob. Sie wollen daher bessere Schranken für die Abweichung $\mathbf{P}(|\hat{X} - \mu| \geq \varepsilon)$. Diese liefert der ZGS! (Satz W1D)

Empirie: häufig wiederholte Experimente



Das **empirische Gesetz der großen Zahlen** ist folgende Erfahrungstatsache: Bei sehr häufiger unabhängiger Wiederholung nähern sich die relativen Häufigkeiten der Wahrscheinlichkeit an. Wir können dies nun begründen und sogar nachrechnen, wie schnell und sicher dies geschieht!

Illustration des Gesetzes der großen Zahlen

Aufgabe: Sie würfeln $n = 3600$ mal mit einem fairen Würfel. Sei S die Häufigkeit der Augenzahl 6. Wie groß sind Erwartung und Streuung? Welche zusätzlichen Annahmen fordern Sie hierzu sinnvollerweise? Wie wahrscheinlich ist $S \in [500, 700]$? $S \in [550, 650]$? $S \in [580, 620]$?

Lösung: Wir nutzen die Zufallsvariablen $X_1, X_2, \dots, X_n : \Omega \rightarrow \{0, 1\}$, wobei $X_k = 1$ bedeutet „Augenzahl 6 im k -ten Wurf“. Damit gilt:

$$\begin{array}{lll} \text{Trefferwkt} & \mathbf{P}(X_k=1) = t & = 1/6, \\ \text{Erwartung} & \mathbf{E}(X_k) = t & = 1/6, \\ \text{Varianz} & \mathbf{V}(X_k) = t(1-t) & = 5/36. \end{array}$$

Die Summe $S = X_1 + X_2 + \dots + X_n : \Omega \rightarrow \mathbb{R}$ zählt die Häufigkeit:

$$\begin{array}{lll} \text{Erwartung} & \mathbf{E}(S) = nt & = 600, \\ \text{Varianz} & \mathbf{V}(S) = nt(1-t) & = 500, \\ \text{Streuung} & \sigma(S) = \sqrt{\mathbf{V}(S)} & \approx 22.36. \end{array}$$

Satz T31: Für die Additivität der Varianzen fordern wir Unabhängigkeit!

Illustration des Gesetzes der großen Zahlen

Wir nutzen die (worst case) Abschätzungen von Chebychev(Satz T3F):

$$\begin{aligned}\mathbf{P}\left[|X - \mu| \geq c\right] &\leq \frac{\sigma^2}{c^2} \\ \mathbf{P}\left[|X - \mu| \geq k\sigma\right] &\leq \frac{1}{k^2}, \\ \mathbf{P}\left[\mu - k\sigma < X < \mu + k\sigma\right] &\geq 1 - \frac{1}{k^2},\end{aligned}$$

In unserem Fall gilt $\mu = 600$ und $\sigma \approx 22.36$, damit erhalten wir:

$$\mathbf{P}\left[|S - 600| \geq 101\right] \leq \frac{500}{101^2} \approx 0.05 \implies \mathbf{P}\left[S \in [500, 700]\right] \geq 0.95,$$

$$\mathbf{P}\left[|S - 600| \geq 51\right] \leq \frac{500}{51^2} \approx 0.20 \implies \mathbf{P}\left[S \in [550, 650]\right] \geq 0.80,$$

$$\mathbf{P}\left[|S - 600| \geq 21\right] \leq \frac{500}{21^2} \approx 1.13 \implies \mathbf{P}\left[S \in [580, 620]\right] \geq -0.13.$$

😊 Die ersten beiden Abschätzungen sind hilfreich, die dritte ist nutzlos. Später rechnen wir genauer mit dem lokalen Grenzwertsatz V3A (LGS).

Illustration des Gesetzes der großen Zahlen

😊 Die Ungleichungen von Chebychev sind zwar grob, aber sie erlauben uns immerhin eine erste Abschätzung, zudem schnell und bequem.

⚠️ Zunächst klingen 3600 Versuche nach viel, doch es genügt nicht: Die Ergebnisse streuen noch stark, die Messung ist recht ungenau!

Aufgabe: Bei 3600 Versuchen gilt $S \in [\mu \pm 0.167\mu]$ mit $\text{Wkt} \geq 95\%$. Wie oft müssen Sie würfeln, damit $S \in [\mu \pm 0.01\mu]$ mit $\text{Wkt} \geq 95\%$ gilt?

Lösung: Wir rechnen wie oben, mit unbekannter Versuchszahl n :

$$\mathbf{P}\left[|S - \mu| \geq 0.01\mu\right] \leq \frac{\sigma^2}{(0.01\mu)^2} = \frac{nt(1-t)}{(0.01nt)^2} = 10\,000 \frac{1-t}{nt} \stackrel{!}{\leq} 0.05$$

Wegen $(1-t)/t = 5/6 / 1/6 = 5$ muss demnach $n \geq 1\,000\,000$ gelten.

😊 Plausibel: Doppelte Genauigkeit erfordert vierfache Versuchszahl, zehnfache Genauigkeit erfordert hundertfache Versuchszahl, 17fache Genauigkeit erfordert 289fache Versuchszahl.

Genauigkeit hat ihren Preis: Wir müssen mehr messen!
Das ist zugleich Fluch und Segen der großen Zahlen.

Anwendung des Gesetzes der großen Zahlen



Präzisionswürfel aus Kunststoff

Gleiche Wkten!



Handgesägte Würfel aus Holz

Ungleiche Wkten?

Bei perfekter Symmetrie sind alle Ergebnisse gleichwahrscheinlich:

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$$

Bei handgesägten Würfeln sind die Wahrscheinlichkeiten nicht gleich:

$$p(1) = 0.15, \quad p(2) = p(3) = p(4) = p(5) = 0.16, \quad p(6) = 0.21$$

Wie findet man diese Wkten? Man muss sie messen! Aber wie genau?

Quantitativer Kompromiss: sicher vs präzise vs kostengünstig

Aufgabe: Sie wollen die Wkt $P(A) = t$ eines Ereignisses A messen:

- (1) Bis auf einen Fehler von $\varepsilon = 0.01$. Sie verlangen 98%ige Sicherheit.
- (2) Bis auf einen Fehler von $\varepsilon = 0.02$. Sie verlangen 96%ige Sicherheit.
- (3) Bis auf einen Fehler von $\varepsilon = 0.05$. Sie verlangen 93%ige Sicherheit.

Wie oft müssen Sie dazu die Messung / das Experiment wiederholen?

😊 Möglichst sicher, präzise, kostengünstig: Wählen Sie zwei von drei!
Nutzen Sie zur Vereinfachung zunächst den Satz von Chebychev T3F,
später die Binomialverteilung U3B und den lokalen Grenzwertsatz V3A.

😊 Wir setzen unser Experiment als beliebig wiederholbar voraus.
Das ist für praktische Anwendungen eine starke Annahme; manchmal
ist schon eine erste Wiederholung zu kostspielig oder gar unmöglich.
Dann muss man Wkten auf andere Weise schätzen, und sei's nur grob.

😊 Selbst wenn das Experiment im Prinzip beliebig wiederholbar ist,
so sind Wiederholungen doch meist aufwändig und kostspielig.
Begrenzte Ressourcen erlauben Ihnen nur begrenzte Genauigkeit.
Für dieses Problem soll Sie dieser Aufgabentyp sensibilisieren.

Quantitativer Kompromiss: sicher vs präzise vs kostengünstig

Lösung: Das einfache Experiment hat Erwartung $\mathbf{E}(\mathbf{I}_A) = \mathbf{P}(A) = t$ und Varianz $\mathbf{V}(\mathbf{I}_A) = t(1-t) \leq 1/4$. Wir nutzen besser n unabhängige Wiederholungen $X_1, X_2, \dots, X_n: \Omega \rightarrow \{0, 1\}$ und bilden den Mittelwert

$$\hat{X} = \frac{1}{n}(X_1 + \dots + X_n) = \text{relative Häufigkeit des Ereignisses } A.$$

Er hat die Erwartung $\mathbf{E}(\hat{X}) = t$ und dank Unabhängigkeit (!) die Varianz $\sigma^2 = \mathbf{V}(\hat{X}) = \mathbf{V}(X)/n \leq 1/4n \searrow 0$, also die Streuung $\sigma \leq 1/2\sqrt{n} \searrow 0$. Wir nutzen das Gesetz der großen Zahlen T3J nach Chebychev:

$$\alpha = \mathbf{P}[|\hat{X} - t| < \varepsilon] \geq 1 - 1/4n\varepsilon^2 \nearrow 1$$

(1) Damit schätzen wir die Wiederholungen für $\alpha = 0.98$ und $\varepsilon = 0.01$:

$$\text{Irrtumswkt: } \mathbf{P}(|\hat{X} - t| \geq k\sigma) \leq 1/k^2 \stackrel{!}{\leq} 0.02 \Leftrightarrow k \geq 8$$

$$\text{Ungenauigkeit: } 8\sigma \leq 8/2\sqrt{n} \stackrel{!}{\leq} \varepsilon = 0.01 \Leftrightarrow n \geq 160\,000$$

Für 98%ige Sicherheit genügt immer ein 8σ -Intervall um den Messwert. Zur Genauigkeit $\varepsilon = 0.01$ genügen demnach $n \geq 160\,000$ Messungen. Genauer: Dank LGS genügt $k = 2.33$, somit nur $n = 13\,573$ Messungen.

Quantitativer Kompromiss: sicher vs präzise vs kostengünstig

(2) Wir schätzen die Wiederholungen für $\alpha = 0.96$ und $\varepsilon = 0.02$:

$$\text{Irrtumswkt: } \mathbf{P}(|\hat{X} - t| \geq k\sigma) \leq 1/k^2 \stackrel{!}{\leq} 0.04 \Leftrightarrow k \geq 5$$

$$\text{Ungenauigkeit: } 5\sigma \leq 5/2\sqrt{n} \stackrel{!}{\leq} \varepsilon = 0.02 \Leftrightarrow n \geq 15\,625$$

Für 96%ige Sicherheit genügt immer ein 5σ -Intervall um den Messwert.
 Für die Genauigkeit $\varepsilon = 0.02$ genügen demnach $n \geq 15\,625$ Messungen.
 Genauer: Dank LGS genügt $k = 2.06$, somit nur $n = 2\,653$ Messungen.

(3) Wir schätzen die Wiederholungen für $\alpha = 0.93$ und $\varepsilon = 0.05$:

$$\text{Irrtumswkt: } \mathbf{P}(|\hat{X} - t| \geq k\sigma) \leq 1/k^2 \stackrel{!}{\leq} 0.07 \Leftrightarrow k \geq 4$$

$$\text{Ungenauigkeit: } 4\sigma \leq 4/2\sqrt{n} \stackrel{!}{\leq} \varepsilon = 0.05 \Leftrightarrow n \geq 1\,600$$

Für 93%ige Sicherheit genügt immer ein 4σ -Intervall um den Messwert.
 Für die Genauigkeit $\varepsilon = 0.05$ genügen demnach $n \geq 1\,600$ Messungen.
 Genauer: Dank LGS genügt $k = 1.82$, somit nur $n = 332$ Messungen.

😊 Genauigkeit und Sicherheit haben ihren Preis! Mathematik kann die Genauigkeit und die Sicherheit erhöhen und somit die Kosten senken.

Disjunkt versus unabhängig

Seien $A, B \subseteq \Omega$ Ereignisse in einem Wahrscheinlichkeitsraum (Ω, \mathbf{P}) . Für die Wahrscheinlichkeit ihrer Vereinigung gilt die einfache Formel

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

Diese Formel gilt *immer*. Meist kennen wir die Wahrscheinlichkeiten $\mathbf{P}(A)$ und $\mathbf{P}(B)$, aber wie berechnen wir den „Störterm“ $\mathbf{P}(A \cap B)$? Zwei Fälle sind besonders nützlich und wichtig:

$$\left. \begin{array}{l} A, B \text{ sind disjunkt, } A \cap B = \emptyset, \\ \text{allgemeiner } \mathbf{P}(A \cap B) = 0. \end{array} \right\} \iff \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

$$\left. \begin{array}{l} \text{Die Ereignisse } A, B \text{ sind} \\ \text{stochastisch unabhängig.} \end{array} \right\} \iff \mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$$

Das sind zwar nur die einfachsten (extremen) Fälle, aber sie treten häufig genug auf, um nützlich zu sein. Häufig nutzt (prüft oder fordert) man in Rechnungen daher Disjunktheit oder Unabhängigkeit. Das hilft.

⚠ Disjunkt *und* unabhängig gilt nur für $\mathbf{P}(A) = 0$ oder $\mathbf{P}(B) = 0$.

Kovarianz und Korrelation von Ereignissen

Wir untersuchen genauer die Un/Abhängigkeit von Ereignissen:

Im Falle $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$ sind die Ereignisse **unabhängig**.

Im Falle $\mathbf{P}(A \cap B) > \mathbf{P}(A) \mathbf{P}(B)$ sind sie **positiv korreliert**.

Im Falle $\mathbf{P}(A \cap B) < \mathbf{P}(A) \mathbf{P}(B)$ sind sie **negativ korreliert**.

Im Allgemeinen können wir dies wie folgt quantifizieren:

Definition T3K: Kovarianz und Korrelation von Ereignissen

Die **Kovarianz** von Ereignissen A, B im WRaum (Ω, \mathbf{P}) ist

$$\mathbf{Cov}(A, B) := \mathbf{P}(A \cap B) - \mathbf{P}(A) \cdot \mathbf{P}(B) = \mathbf{E}(X \cdot Y)$$

mit den zentrierten Zufallsvariablen $X = \mathbf{I}_A - \mathbf{P}(A)$ und $Y = \mathbf{I}_B - \mathbf{P}(B)$.

Im Fall $\mathbf{P}(A), \mathbf{P}(B) \in]0, 1[$ definieren wir ihre **Korrelation** durch

$$\mathbf{Cor}(A, B) := \frac{\mathbf{P}(A \cap B) - \mathbf{P}(A) \cdot \mathbf{P}(B)}{\sqrt{\mathbf{P}(A) \mathbf{P}(\bar{A}) \cdot \mathbf{P}(B) \mathbf{P}(\bar{B})}} = \frac{\mathbf{E}(X \cdot Y)}{\sqrt{\mathbf{V}(X) \cdot \mathbf{V}(Y)}} \in [-1, 1].$$

Im Fall $\mathbf{P}(A) \in \{0, 1\}$ oder $\mathbf{P}(B) \in \{0, 1\}$ setzen wir $\mathbf{Cor}(A, B) := 0$.

Kovarianz und Korrelation von Ereignissen

Die Un/Abhängigkeit von Ereignissen $A, B \subseteq \Omega$ ist eine Ja/Nein-Frage. Ihre Korrelation $\text{Cor}(A, B) \in [-1, 1]$ hingegen quantifiziert den Grad ihrer Abhängigkeit. Hierzu gelten einfache, aber nützliche Regeln:

Aufgabe: Beweisen Sie die folgenden Rechenregeln:

- (1) Genau dann gilt $\text{Cor}(A, B) = 0$, wenn A und B unabhängig sind.
- (2) Es gilt $\text{Cor}(A, A) = 1$, $\text{Cor}(A, \bar{A}) = -1$, $\text{Cor}(A, \bar{B}) = -\text{Cor}(A, B)$.
- (3) Für alle $A, B \subseteq \Omega$ gilt $\text{Cor}(A, B) \in [-1, 1]$ dank Cauchy-Schwarz.
- (4) Aus $\text{Cor}(A, B) = +1$ folgt $A = B$ *fast sicher*, also $\mathbf{P}(A \Delta B) = 0$.
- (5) Aus $\text{Cor}(A, B) = -1$ folgt $A = \bar{B}$ *fast sicher*, also $\mathbf{P}(A \Delta \bar{B}) = 0$.

Vereinbarung: Die bequeme Ausdrucksweise „fast sicher“ bedeutet „bis auf eine vernachlässigbare Menge mit Wahrscheinlichkeit Null“.

Beispiel: Wenn $\mathbf{P}(\{\omega\}) > 0$ für alle $\omega \in \Omega$ gilt, dann gilt in (4) $A = B$ bzw. in (5) $A = \bar{B}$. Es ist jedoch möglich, dass unser WRaum (Ω, \mathbf{P}) Elemente ω mit Wkt $\mathbf{P}(\{\omega\}) = 0$ enthält, dann können wir nicht $A = B$ bzw. $A = \bar{B}$ schließen, sondern nur $\mathbf{P}(A \Delta B) = 0$ bzw. $\mathbf{P}(A \Delta \bar{B}) = 0$.

Kovarianz und Korrelation von Ereignissen

Lösung: (1) Im Sonderfall $\mathbf{P}(A) \in \{0, 1\}$ oder $\mathbf{P}(B) \in \{0, 1\}$ sind A und B unabhängig, und es gilt $\mathbf{Cor}(A, B) = 0$ nach Vereinbarung. Im regulären Fall $\mathbf{P}(A), \mathbf{P}(B) \in]0, 1[$ betrachten wir den obigen Bruch: Genau dann gilt $\mathbf{Cor}(A, B) = 0$, wenn $\mathbf{P}(A \cap B) - \mathbf{P}(A) \cdot \mathbf{P}(B) = 0$.

(2) Einsetzen liefert sofort $\mathbf{Cor}(A, A) = +1$ und $\mathbf{Cor}(A, \bar{A}) = -1$. Zum Ereignis $A \subseteq \Omega$ betrachten wir $X = \mathbf{I}_A - \mathbf{P}(A)$, zum Komplement $\bar{A} = \Omega \setminus A$ entsprechend $\bar{X} = \mathbf{I}_{\bar{A}} - \mathbf{P}(\bar{A}) = (1 - \mathbf{I}_A) - (1 - \mathbf{P}(A)) = -X$. Hieraus folgt $\mathbf{Cov}(A, \bar{B}) = -\mathbf{Cov}(A, B)$ und $\mathbf{Cor}(A, \bar{B}) = -\mathbf{Cor}(A, B)$.

(3) Für $X = \mathbf{I}_A - \mathbf{P}(A)$ gilt $\mathbf{E}(X) = 0$ und $\mathbf{V}(X) = \mathbf{P}(A)\mathbf{P}(\bar{A})$. Ebenso für $Y = \mathbf{I}_B - \mathbf{P}(B)$ gilt $\mathbf{E}(Y) = 0$ und $\mathbf{V}(Y) = \mathbf{P}(B)\mathbf{P}(\bar{B})$. Wir finden $\mathbf{E}(XY) = \mathbf{E}(\mathbf{I}_A \mathbf{I}_B) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)$. Dank der Cauchy–Schwarz–Ungleichung gilt $|\mathbf{E}(XY)|^2 \leq \mathbf{E}(X^2)\mathbf{E}(Y^2)$, also $\mathbf{Cor}(X, Y)^2 \leq 1$ und somit $\mathbf{Cor}(X, Y) \in [-1, 1]$ wie gewünscht.

(4/5) Gleichheit gilt bei Cauchy–Schwarz für $Y = \lambda X$ mit $\lambda \in \mathbb{R}$. Wegen $1 = \max X - \min X = \max Y - \min Y$ bleibt nur $\lambda = \pm 1$. In $L^2(\Omega, \mathbf{P})$ gilt demnach $X = \pm Y$, also $A = B$ bzw. $A = \bar{B}$ fast sicher.

Kovarianz und Korrelation von Zufallsvariablen

Die Varianz ist nicht additiv! T331 Die nötige Korrektur liefert gerade die Kovarianz: $V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y)$. Das erinnert uns an die lineare Algebra: Die Varianz $X \mapsto V(X)$ ist eine quadratische Form, die Kovarianz $(X, Y) \mapsto \text{Cov}(X, Y)$ ist ihre **Polarisierung**.

Definition T3L: Kovarianz und Korrelation von Zufallsvariablen

Seien $X, Y : (\Omega, \mathbf{P}) \rightarrow \mathbb{R}$ Zufallsvariablen mit $\mathbf{E}[X^2], \mathbf{E}[Y^2] < \infty$.

Wir definieren ihre **Kovarianz** und ihre **Korrelation** durch

$$\text{Cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))] = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y),$$

$$\text{Cor}(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} \quad \text{bzw. } := 0, \text{ falls } V(X)V(Y) = 0.$$

Wir nennen die beiden Zufallsvariablen X, Y

- **unkorreliert**, falls $\text{Cov}(X, Y) = 0$ gilt,
- **positiv koreliert**, falls $\text{Cov}(X, Y) > 0$ gilt,
- **negativ koreliert**, falls $\text{Cov}(X, Y) < 0$ gilt.

Satz T3M: Kovarianz als Skalarprodukt

Die Kovarianz verhält sich wie ein Skalarprodukt:

$$\mathbf{Cov}(X, X) = \mathbf{V}(X) \geq 0$$

$$\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$$

$$\mathbf{Cov}(a + bX, Y) = b \mathbf{Cov}(X, Y)$$

Die Varianz $\mathbf{V}(X) = \mathbf{Cov}(X, X)$ entspricht dem Normquadrat, die Streuung $\sigma(X) = \sqrt{\mathbf{V}(X)}$ entspricht somit der Norm.

Genau dann gilt $\mathbf{V}(X) = 0$, wenn $X = \text{const}$ fast sicher.

Für die Varianz einer Summe gilt Pythagoras:

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X, Y)$$

Unkorreliert $\mathbf{Cov}(X, Y) = 0$ bedeutet orthogonal $X \perp Y$.

Dank Cauchy–Schwarz–Ungleichung gilt $\mathbf{Cor}(X, Y) \in [-1, 1]$.

Maximal gilt $\mathbf{Cor}(X, \pm X) = \pm 1$ und allgemein folgende Äquivalenz:

$\mathbf{Cor}(X, Y) = \pm 1$ genau dann, wenn $X = \pm \lambda Y$ fast sicher (mit $\lambda > 0$).

Unabhängig vs unkorreliert

Aufgabe: Seien $X, Y : \Omega \rightarrow \mathbb{R}$ zwei Zufallsvariablen. Das Paar (X, Y) nimmt die vier Werte $(\pm 1, 0), (0, \pm 1) \in \mathbb{R}^2$ jeweils mit Wkt $1/4$ an.

(1) Berechnen Sie $\mathbf{E}(X)$, $\mathbf{V}(X)$ und $\mathbf{E}(Y)$, $\mathbf{V}(Y)$ sowie $\mathbf{Cov}(X, Y)$.

(2) Sind X und Y unkorreliert? (3) Sind X und Y unabhängig?

(4) Welche Verteilung hat $S = X + Y$? Erwartung? Varianz?


Lösung: (1) Es gilt $\mathbf{P}(X=\pm 1) = \frac{1}{4}$ und $\mathbf{P}(X=0) = \frac{1}{2}$, also $\mathbf{E}(X) = 0$ und $\mathbf{V}(X) = \frac{1}{2}$. Ebenso für Y , denn Y hat dieselbe Verteilung.

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))] = \mathbf{E}(XY) = 0$$

(2) Demnach sind X und Y unkorreliert. (3) Aber nicht unabhängig:

$$\mathbf{P}(X=1) = \mathbf{P}(Y=1) = \frac{1}{4} \quad \text{aber} \quad \mathbf{P}(X=1, Y=1) = 0$$

(4) Für die Summe $S = X + Y$ gilt offensichtlich $\mathbf{P}(S=\pm 1) = \frac{1}{2}$, also $\mathbf{E}(S) = 0 = \mathbf{E}(X) + \mathbf{E}(Y)$ und $\mathbf{V}(S) = 1 = \mathbf{V}(X) + \mathbf{V}(Y)$.

 Die Variablen X und Y sind abhängig, aber dennoch unkorreliert. Diese schwächere Bedingung genügt für die Additivität der Varianzen.

Unabhängig vs unkorreliert

Aufgabe: Wir untersuchen die drei folgenden Aussagen über X, Y :

- (1) Die Zufallsvariablen X, Y sind unabhängig (gemäß Definition T3H).
- (2) Die Zufallsvariablen X, Y sind unkorreliert, also $\text{Cov}(X, Y) = 0$.
- (3) Die Varianzen addieren sich gemäß $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$.

Welche der sechs möglichen Implikationen gelten hier?

Geben Sie jeweils einen Beweis oder ein Gegenbeispiel.

Lösung: Pythagoras T3M zeigt sofort die Äquivalenz „(2) \Leftrightarrow (3)“.

Es gilt „(1) \Rightarrow (2)“: Satz T3I garantiert $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ für unabhängige Variablen, also $\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = 0$.

Die Umkehrung „(2) \Rightarrow (1)“ gilt nicht: Die vorige Aufgabe zeigt ein Gegenbeispiel. (Sie können leicht Varianten davon konstruieren.)

 Unabhängig impliziert unkorreliert, aber nicht umgekehrt!

Aufgabe: Für alle *zweiwertigen* Zufallsvariablen $X = a_0 + a_1 \mathbf{I}_A$ und $Y = b_0 + b_1 \mathbf{I}_B$ ist unabhängig äquivalent zu unkorreliert. **Lösung:** T329

 Für drei- und mehrwertige ZVariablen gilt dies allgemein nicht mehr!

Korrelation impliziert nicht Kausalität!

Aufgabe: Wir beobachten sehr häufig folgende Ereignisse:

$$R = \{\text{Es regnet}\}, \quad S = \{\text{Scheibenwischer laufen}\}$$


(1) Es gilt etwa $P(S|R) = 99\%$. Was können Sie hieraus folgern?
Verursacht Regen das Anschalten der Scheibenwischer? „Ja, klar!“

(2) Es gilt auch etwa $P(R|S) = 99\%$. Was können Sie hieraus folgern?
Verursacht Anschalten der Scheibenwischer also Regen? „Whaaa?“

 Dieser Denkfehler ist absurd, aber häufig. Wie ist er zu widerlegen?

Dieses Beispiel ist so gewählt, dass der Denkfehler möglichst auffällig ist, gar absurd.

Doch wo genau liegt der Fehler? Wir wissen alle aus Erfahrung, dass das Anschalten der Scheibenwischer nicht Regen verursacht. Woher wissen wir das? Das sagen uns nicht die obigen Wahrscheinlichkeiten, sondern folgendes aktive Experiment: Wir können willkürlich / zufällig den Scheibenwischer anschalten und den Regen beobachten. Nach zahlreicher Wiederholung dieses Experiments sind wir sicher, dass der vermutete kausale Zusammenhang nicht besteht.

 Kausalität beweist man immer nur aktiv, durch randomisierte Tests.

So lernten wir als Kinder, Zusammenhänge zu begreifen und mögliche Erklärungen zu erhärten: durch ausgiebige Tests und endlos wiederholte Spiele. Das ist eine überaus sinnvolle Strategie! Siehe hierzu en.wikipedia.org/wiki/Correlation_does_not_imply_causation.

Korrelation impliziert nicht Kausalität!

😊 Kausalität impliziert meist gewisse Wahrscheinlichkeiten.
Unser Erfahrungswissen nutzen wir zur Erstellung unserer Modelle.

☹️ Umgekehrt jedoch müssen wir dabei extrem vorsichtig sein:
Beobachtung einer Korrelation impliziert noch keine Kausalität.

⚠️ Wissenschaftliche Grundregel: Ursache und Wirkung lässt sich nur durch ein detailliertes Verständnis der Mechanismen ergründen!

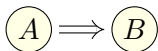
⚠️ Ein passives, statistisches Abbild der Situation ist hierfür zu grob.
Um Kausalität zu messen, benötigen wir aktiv randomisierte Tests.

Beispiel: Menschen mit hohem Vitamin-C-Gehalt im Blut erleiden seltener einen Herzinfarkt.
Kann man hieraus folgern, dass Vitamin C das Risiko mindert? Nein, daraus allein noch nicht!
Ein hoher Vitamin-C-Gehalt kann durch andere Faktoren verursacht werden (wie etwa gesunder Lebenswandel, viel Obst, viel Sport) und diese mindern das Infarktrisiko, nicht allein Vitamin C:
Es ist noch nicht schlüssig, ob zusätzlich verabreichtes Vitamin C das Risiko senkt oder erhöht.

Da sehr viele (bekannte und unbekante) Faktoren das Infarktrisiko beeinflussen, muss man mit voreiligen Schlussfolgerungen vorsichtig sein. Die Wirkung von Vitamin C auf das Infarktrisiko kann man glücklicherweise dennoch untersuchen: Sorgfältig konstruiert kann ein randomisierter kontrollierter Test (RCT) hierauf Antwort geben. Dies ist zwar extrem aufwändig, aber allgemein das beste Studiendesign in der medizinischen, psychologischen und ökonomischen Forschung.

Korrelation impliziert nicht Kausalität!

Korrelation tritt in diversen Zusammenhängen auf, manchmal mit und häufig auch ohne Kausalität:



Kausalität

„A verursacht B“

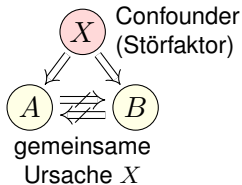


Kausalität

„B verursacht A“



Gleichzeitigkeit
ohne Kausalität



Aufgabe: Diskutieren Sie das Auftreten folgender Korrelationen:

- 1 „Regionen mit mehr Störchen haben höhere Geburtenraten.“
- 2 „Mit Schuhen zu schlafen, geht mit Kopfschmerzen einher.“
- 3 „Tabak rauchen geht häufig mit Leberschäden einher.“
- 4 „Je mehr Feinstaub, desto höher das Herzinfarkttrisiko.“
- 5 „Spieler von Ego-Shootern sind häufiger aggressiv.“
- 6 „Ein Glas Rotwein pro Tag fördert die Gesundheit.“
- 7 „Vorlesungsbesuch geht mit Klausurerfolg einher.“

Begründet dies eine Kausalität? Sehen Sie mögliche Gegenargumente? All diese Zusammenhänge (bis auf einen) wurden in Studien untersucht. Vermeiden Sie voreilige Schlüsse! Fragen Sie nach! Widersprechen Sie!

Korrelation impliziert nicht Kausalität!

- (1) Störche und Geburten sind positiv korreliert; siehe Robert Matthews: *Storks deliver Babies*. Teaching Statistics 22 (2000) 36–38. Mögliche gemeinsame Ursache: Ländliche Regionen haben (a) mehr Störche und (b) mehr Geburten. Mehr Störche anzusiedeln, erhöht nicht die Geburten.
- (2) Gemeinsame Ursache dürfte Alkoholkonsum sein. Die Schuhe auszuziehen, hilft nicht gegen Kopfschmerz. Sie wollen das ausprobieren? Machen Sie einen randomisierten kontrollierten Test!
- (3) Dieser Zusammenhang wurde intensiv untersucht. Ein wesentlicher Störfaktor: Tabakrauchen korreliert mit Alkoholtrinken, und letzteres verursacht Leberschäden. Das Rauchen aufzuhören, fördert sicherlich allgemein die Gesundheit, hat aber keinen direkten Einfluss auf Leberschäden.
- (4) Feinstaub wird heftig diskutiert, gerade in Stuttgart. Studien müssen Störfaktoren vermeiden, Beispiel: Geringes Einkommen impliziert (a) Gesundheitsrisiken und (b) schlechte Wohnlage, darunter auch hohe Luftbelastung. Vereinfacht gesagt: Wer sich's leisten kann, zieht weg.
- (5) Ein komplexes stachliges Thema! Der Oberste Gerichtshof der USA resümierte 2011 in einer Urteilsbegründung: „Nearly all of the research is based on correlation, not evidence of causation.“
- (6) Positive Korrelation besteht. Mögliche gemeinsame Ursache: Hohes Einkommen impliziert (a) gute Gesundheit und (b) Weinkonsum. Auch medizinische Zusammenhänge werden diskutiert. (Vor allem in französischen Studien? en.wikipedia.org/wiki/French_paradox ;-)
- (7) Die Korrelation ist stark. Schon drei verpasste Termine senken signifikant den Klausurerfolg. Prof. Rolf Schulmeister (Uni Hamburg) hat eine Meta-Studie von 300 empirischen Arbeiten vorgelegt: *Abwesenheit von Lehrveranstaltungen. Ein nur scheinbar triviales Problem*. (2015)

Diskrete Wahrscheinlichkeitsräume

Ein **diskreter Wahrscheinlichkeitsraum** (Ω, \mathbf{P}) besteht aus

- einer Menge Ω von möglichen Ergebnissen und
- einer normierten und additiven Abbildung $\mathbf{P} : \mathfrak{P}(\Omega) \rightarrow [0, 1]$.

Letzteres bedeutet $\mathbf{P}(\Omega) = 1$ und $\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\{\omega\})$ für alle $A \subseteq \Omega$.

Alle Ergebnisse fassen wir zur **Ergebnismenge** Ω zusammen. Jedes Element $\omega \in \Omega$ heißt **Ergebnis**. Jede Teilmenge $A \subseteq \Omega$ heißt **Ereignis**. Das Ereignis A tritt genau dann ein, wenn ein Ergebnis $\omega \in A$ eintritt. Die Wahrscheinlichkeit hierfür ist das **Wahrscheinlichkeitsmaß** $\mathbf{P}(A)$.

Spezialfall: Ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) heißt **endlich**, wenn die Ergebnismenge Ω endlich ist. Er heißt **laplacesch**, wenn zudem alle Ergebnisse $\omega \in \Omega$ gleich wahrscheinlich sind.

Für das **Laplace–Wahrscheinlichkeitsmaß** auf Ω gilt daher

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der Ergebnisse in } A}{\text{Anzahl aller Ergebnisse in } \Omega} = \frac{\text{günstige Ergebnisse}}{\text{mögliche Ergebnisse}}.$$

😊 Für Laplace–Experimente berechnet man die Wkten durch Abzählen der Ergebnisse. Hierbei helfen kombinatorische Abzählformeln.

Rechnen mit Ereignissen

Die Sprache der Mengen erlaubt uns, mit Ereignissen zu rechnen:
Klare und eindeutige Formulierung ermöglicht effiziente Rechnung.

Menge	Bedeutung als Zufallsereignis
Ω	Das sichere Ereignis: Ω tritt immer ein, $\mathbf{P}(\Omega) = 1$.
\emptyset	Das unmögliche Ereignis: \emptyset tritt nie ein, $\mathbf{P}(\emptyset) = 0$.
$A \subseteq \Omega$	Das Ereignis A tritt ein bei jedem Ergebnis $\omega \in A$.
$\Omega \setminus A$	Komplement: $\bar{A} = \Omega \setminus A$ tritt ein, wenn A nicht eintritt.
$A \subseteq B$	Teilmenge: Immer wenn A eintritt, dann auch B .
$B \setminus A$	Restmenge: Das Ereignis B tritt ein, aber nicht A .
$A \cap B$	Die Ereignisse A und B treten beide ein.
$A \cup B$	Das Ereignis A oder B tritt ein (evtl. auch beide).
$A \Delta B$	Entweder A oder B tritt ein (aber nicht beide).
$A \sqcup B$	Das Ereignis A oder B tritt ein, wobei $A \cap B = \emptyset$.

Unser Ziel sind nachvollziehbar begründete, quantitative Aussagen.

😊 Diese Notation erlaubt präzises Rechnen mit Zufallsereignissen.

Bedingte Wahrscheinlichkeit

Sei (Ω, \mathbf{P}) ein Wahrscheinlichkeitsraum und $B \subseteq \Omega$ ein Ereignis mit Wahrscheinlichkeit $\mathbf{P}(B) > 0$. Für jedes Ereignis $A \subseteq \Omega$ definieren wir die **Wahrscheinlichkeit von A unter der Bedingung B** durch

$$\mathbf{P}(A|B) := \mathbf{P}_B(A) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Sei $\Omega = B_1 \sqcup B_2 \sqcup \dots \sqcup B_\ell$ eine Zerlegung in disjunkte Ereignisse. Für $A \subseteq \Omega$ gilt dann die **Formel der totalen Wahrscheinlichkeit**:

$$\mathbf{P}(A) = \sum_{k=1}^{\ell} \mathbf{P}(A \cap B_k) = \sum_{k=1}^{\ell} \mathbf{P}(A|B_k) \mathbf{P}(B_k)$$

Im Falle $\mathbf{P}(A) > 0$ gilt zudem die **Formel von Bayes** (Satz T2B):

$$\mathbf{P}(B_i|A) = \frac{\mathbf{P}(A|B_i) \mathbf{P}(B_i)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A|B_i) \mathbf{P}(B_i)}{\sum_{k=1}^{\ell} \mathbf{P}(A|B_k) \mathbf{P}(B_k)}$$

- ☺ Bequeme Notation und präzises Rechnen mit Zufallsereignissen.
- ☺ Die Rechnung ist meist leicht, die Interpretation erfordert Übung.

Stochastische Unabhängigkeit

Eine Familie von Ereignissen $A_1, A_2, \dots, A_n \subseteq \Omega$ heißt **(stochastisch) unabhängig**, wenn alle Schnittmengen die Produktregel erfüllen:

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}) = \mathbf{P}(A_{i_1}) \mathbf{P}(A_{i_2}) \dots \mathbf{P}(A_{i_\ell})$$

für jede Auswahl $1 \leq i_1 < i_2 < \dots < i_\ell \leq n$: Die Wkt jedes Schnitts ist das Produkt der Wahrscheinlichkeiten. Für unabhängige Ereignisse lassen sich damit die Wkten aller erzeugten Ereignisse berechnen.

Das heißt, zwei Ereignisse $A, B \subseteq \Omega$ sind unabhängig, wenn gilt:

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

Drei Ereignisse $A, B, C \subseteq \Omega$ sind unabhängig, wenn gilt:

- $\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B)$,
- $\mathbf{P}(A \cap C) = \mathbf{P}(A) \mathbf{P}(C)$,
- $\mathbf{P}(B \cap C) = \mathbf{P}(B) \mathbf{P}(C)$,
- $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \mathbf{P}(B) \mathbf{P}(C)$.



Die Tripelbedingung folgt nicht aus den drei Paarbedingungen!

Verständnisfragen: Grundbegriffe

Versuchen Sie, folgende Fragen frei aber genau zu beantworten, etwa so, wie Sie dies einer Kommiliton:in / Kolleg:in erklären wollen.

Was ist ein diskreter Wahrscheinlichkeitsraum? Wann nennen wir ihn endlich? laplacesch? Was ist die endliche Gleichverteilung auf Ω ? Welche allgemeinen Rechenregeln gelten für Ereignisse und Wkten?

Welche Wkten haben die Augensummen bei zweimal Würfeln? dreimal? Wie berechnen Sie dies geschickt? Wie nutzen Sie hierzu die Faltung?

Wie definieren Sie bedingte Wahrscheinlichkeit? Ist dies ein WMaß? Wie lautet und was besagt die Formel von der totalen Wkt? und von Bayes?

Wann sind zwei Ereignisse stochastisch unabhängig? drei Ereignisse? n Ereignisse? Folgt aus paarweiser die gemeinsame Unabhängigkeit? Was genau können Sie damit berechnen? Nennen Sie Beispiele!

Was ist eine Zufallsvariable? Was ist ihre Erwartung und ihre Varianz? Was gilt für Linearkombinationen von ZV? von unabhängigen ZV? Wie lautet das Gesetz der großen Zahlen? Wie wenden Sie es an?

Verständnisfragen: typische Fehler

Aufgabe: Seien $A, B, C \subseteq \Omega$ Ereignisse in einem WRaum (Ω, \mathbf{P}) .
Gelten folgende Formeln allgemein? Geben Sie einen Beweis oder ein Gegenbeispiel! Kennen Sie zusätzliche, hinreichende Voraussetzungen?

$$(1) \quad \mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$$

$$(2) \quad \mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A)$$

$$(3) \quad \mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(B)$$

$$(4) \quad \mathbf{P}(A \sqcup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

$$(5) \quad \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

$$(6) \quad \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

$$(7) \quad \mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) \\ - \mathbf{P}(A \cap B) - \mathbf{P}(B \cap C) - \mathbf{P}(A \cap C)$$

Lösung: Zeichnen Sie Skizzen! Formeln (2),(4),(6) gelten immer.

(1) Falsch im Allgemeinen. Dies ist äquivalent zur Unabhängigkeit!

(3) Falsch im Allgemeinen. Es braucht $\mathbf{P}(B \setminus A) = 0$, z.B. $B \subseteq A$.

(5) Falsch im Allgemeinen. Es braucht $\mathbf{P}(A \cap B) = 0$, z.B. $A \cap B = \emptyset$.

(7) Falsch im Allgemeinen. Es fehlt der Korrekturterm $+\mathbf{P}(A \cap B \cap C)$.

Verständnisfragen: typische Fehler

Aufgabe: (1) „Die Wkt, die Klausur zu bestehen, beträgt 40%.

Bei drei Versuchen bestehe ich mit 120%iger Sicherheit!“ Richtig?

(2) „Die Erfolgswkt einer unbemannten Marsmission liegt bei 50%.

Es genügen also zwei Missionen, und wir haben 100%igen Erfolg!“

Lösung: (1) Offenbar ist $3 \cdot 40\% = 120\%$ nicht die richtige Antwort!

Die Wkt, alle dreimal nicht zu bestehen, ist – bei Unabhängigkeit! –

$$0.6 \cdot 0.6 \cdot 0.6 = 0.216.$$

Die Wkt, mindestens einmal zu bestehen, ist also nur $0.784 \approx 78\%$.

Sie könnten nun einwenden, dass diese Rechnung unrealistisch ist: Sobald Sie die erste Klausur bestehen, müssen und werden Sie keine weiteren Versuche unternehmen. Wenn Sie möchten, können Sie auch dieses alternative Modell ausführlich durchrechnen. Es kommt dasselbe heraus! Ernster ist folgender Einwand: Klausuren sind keine Lotterie! Unabhängigkeit gilt nicht, da sich jeder vernünftige Prüfling nach einem ersten Misserfolg sorgfältiger vorbereitet. (Nicht wahr?)

(2) Tatsächlich ist $2 \cdot 50\% = 100\%$ auch hier nicht die richtige Antwort!

Die Wkt von zwei Misserfolgen ist bei Unabhängigkeit $0.5 \cdot 0.5 = 0.25$.

Die Wkt von mindestens einem Erfolg ist demnach nur $0.75 = 75\%$.

Verständnisfragen: fair oder gezinkt?

Aufgabe: (Aus der NEW YORK TIMES, August 2011) Von drei gleich aussehenden Münzen sind zwei fair, zeigen also Kopf 0 und Zahl 1 mit gleicher Wkt, die dritte ist gezinkt und zeigt Kopf mit Wkt 60%.

Sie wählen zufällig eine der Münzen aus und machen einen Testwurf. Wenn Sie dabei Kopf erhalten, mit welcher Wkt ist Ihre Münze gezinkt? (Schöne Aufgabe. Glückliche die Zeitung, die solche Leser:innen hat!)

Lösung: Sie wählen eine faire Münze mit Wkt $P(F) = 2/3$ und die gezinkte mit Wkt $P(\bar{F}) = 1/3$. Gegeben sind zudem die Wkten $P(0|F) = P(1|F) = 1/2$ sowie $P(0|\bar{F}) = 0.6$ und $P(1|\bar{F}) = 0.4$. Dank Bayes folgt also für die gesuchte Wahrscheinlichkeit:

$$P(\bar{F}|0) = \frac{P(0|\bar{F})P(\bar{F})}{P(0|\bar{F})P(\bar{F}) + P(0|F)P(F)} = \frac{1/5}{1/5 + 1/3} = \frac{3}{8}$$

😊 Anschaulich gesagt: Vor dem Testwurf ist die Wkt $1/3 \approx 0.333$, nach dem Testwurf $3/8 = 0.375$. Das ist plausibel. Ist eine höhere Sicherheit nötig, dann sollten Sie umfangreichere Tests durchführen.

Ausfallwahrscheinlichkeit: Welcome to the machine!

Aufgabe: Drei unabhängige Bauteile A, B, C haben die Ausfallwkten

$$\mathbf{P}(A) = a = 0.3, \quad \mathbf{P}(B) = b = 0.4, \quad \mathbf{P}(C) = c = 0.5.$$

- (0) Mit welcher Wkt fällt mindestens ein Teil aus? (1) Genau 0, 1, 2, 3?
(2) Wenn nur genau ein Teil ausfällt, mit welcher Wkt ist es A, B, C ?
(3) Konstruieren Sie als Modell hierfür explizit einen WRaum (Ω, \mathbf{P}) .

Lösung: (0) Offenbar ist $0.3 + 0.4 + 0.5 = 1.2$ unsinnig! (Warum?)

Kein Teil fällt aus, wenn A funktioniert, B funktioniert, C funktioniert:

$$\{\text{Keins fällt aus}\} = \bar{A} \cap \bar{B} \cap \bar{C}$$

Dank **Unabhängigkeit** können wir diese Wkt als Produkt berechnen:

$$\mathbf{P}(\text{Keins fällt aus}) = \mathbf{P}(\bar{A}) \cdot \mathbf{P}(\bar{B}) \cdot \mathbf{P}(\bar{C}) = 0.7 \cdot 0.6 \cdot 0.5 = 0.21$$

Die Wkt, dass mindestens ein Teil ausfällt, ist demnach $1 - 0.21 = 0.79$.

😊 Komplement verkürzt die Rechnung. Frage (1) geht den langen Weg.
Beide Wege führen zum Ziel. Kennen Sie beide, wählen Sie weise!

Ausfallwahrscheinlichkeit: Welcome to the machine!

(1) Mit welcher Wkt fallen genau 0, 1, 2, 3 Teile aus? Wir zerlegen:

$$\{\text{Keins fällt aus}\} = F_0 := \bar{A} \cap \bar{B} \cap \bar{C}$$

$$\{\text{Eins fällt aus}\} = F_1 := (A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)$$

$$\{\text{Zwei fallen aus}\} = F_2 := (A \cap B \cap \bar{C}) \cup (A \cap \bar{B} \cap C) \cup (\bar{A} \cap B \cap C)$$

$$\{\text{Drei fallen aus}\} = F_3 := A \cap B \cap C$$

😊 Wir nutzen alle oben ausgeführten Rechenregeln:

Bei Vereinigung **disjunkter** Ereignisse ist die Wkt additiv.

Bei Schnitten **unabhängiger** Ereignisse ist die Wkt multiplikativ.

$$\mathbf{P}(\text{Keins fällt aus}) = \mathbf{P}(F_0) = \bar{a} \cdot \bar{b} \cdot \bar{c} = 0.21$$

$$\mathbf{P}(\text{Eins fällt aus}) = \mathbf{P}(F_1) = a \cdot \bar{b} \cdot \bar{c} + \bar{a} \cdot b \cdot \bar{c} + \bar{a} \cdot \bar{b} \cdot c = 0.44$$

$$\mathbf{P}(\text{Zwei fallen aus}) = \mathbf{P}(F_2) = a \cdot b \cdot \bar{c} + a \cdot \bar{b} \cdot c + \bar{a} \cdot b \cdot c = 0.29$$

$$\mathbf{P}(\text{Drei fallen aus}) = \mathbf{P}(F_3) = a \cdot b \cdot c = 0.06$$

😊 Plausibilitätsprüfung: Für WMaß gilt $0.21 + 0.44 + 0.29 + 0.06 = 1$.

Ausfallwahrscheinlichkeit: Welcome to the machine!

(2) Genau ein Teil fällt aus; mit welcher Wkt ist es A , B , C ?

Die Wkten berechnen wir dank Unabhängigkeit (wie zuvor):

$$\text{Nur } A \text{ fällt aus: } X := A \cap \bar{B} \cap \bar{C}, \quad \mathbf{P}(X) = 0.3 \cdot 0.6 \cdot 0.5 = 0.09$$

$$\text{Nur } B \text{ fällt aus: } Y := \bar{A} \cap B \cap \bar{C}, \quad \mathbf{P}(Y) = 0.7 \cdot 0.4 \cdot 0.5 = 0.14$$

$$\text{Nur } C \text{ fällt aus: } Z := \bar{A} \cap \bar{B} \cap C, \quad \mathbf{P}(Z) = 0.7 \cdot 0.6 \cdot 0.5 = 0.21$$

$$\text{Eins fällt aus: } F_1 := X \sqcup Y \sqcup Z, \quad \mathbf{P}(F_1) = 0.09 + 0.14 + 0.21 = 0.44$$

Für die bedingten Wahrscheinlichkeiten folgt somit:

$$\mathbf{P}(A|F_1) = \frac{\mathbf{P}(A \cap F_1)}{\mathbf{P}(F_1)} = \frac{\mathbf{P}(X)}{\mathbf{P}(F_1)} = \frac{0.09}{0.44} \approx 0.205$$

$$\mathbf{P}(B|F_1) = \frac{\mathbf{P}(B \cap F_1)}{\mathbf{P}(F_1)} = \frac{\mathbf{P}(Y)}{\mathbf{P}(F_1)} = \frac{0.14}{0.44} \approx 0.318$$

$$\mathbf{P}(C|F_1) = \frac{\mathbf{P}(C \cap F_1)}{\mathbf{P}(F_1)} = \frac{\mathbf{P}(Z)}{\mathbf{P}(F_1)} = \frac{0.21}{0.44} \approx 0.477$$

😊 Plausibilitätsprüfung: Für WMaß gilt $0.205 + 0.318 + 0.477 = 1$.

Ausfallwahrscheinlichkeit: Welcome to the machine!

(3) Wir konstruieren explizit einen geeigneten WRaum (Ω, \mathbf{P}) .

Mögliches Modell: $\Omega = \{0, 1\}^3$ mit 0 = „fehlerfrei“ und 1 = „fällt aus“.

Die zugehörigen Wkten erhalten wir dank Unabhängigkeit:

$$\mathbf{P}(\{111\}) = a \cdot b \cdot c = 0.06, \quad \mathbf{P}(\{110\}) = a \cdot b \cdot \bar{c} = 0.06,$$

$$\mathbf{P}(\{101\}) = a \cdot \bar{b} \cdot c = 0.09, \quad \mathbf{P}(\{100\}) = a \cdot \bar{b} \cdot \bar{c} = 0.09,$$

$$\mathbf{P}(\{011\}) = \bar{a} \cdot b \cdot c = 0.14, \quad \mathbf{P}(\{010\}) = \bar{a} \cdot b \cdot \bar{c} = 0.14,$$

$$\mathbf{P}(\{001\}) = \bar{a} \cdot \bar{b} \cdot c = 0.21, \quad \mathbf{P}(\{000\}) = \bar{a} \cdot \bar{b} \cdot \bar{c} = 0.21$$

😊 Hiermit ist (Ω, \mathbf{P}) tatsächlich ein WRaum: $\mathbf{P} \geq 0$ und $\mathbf{P}(\Omega) = 1$.
Unsere Ereignisse sind dann $A = \{1??\}$ und $B = \{?1?\}$ und $C = \{??1\}$
sowie ihre Komplemente $\bar{A} = \{0??\}$ und $\bar{B} = \{?0?\}$ und $\bar{C} = \{??0\}$.

Die zugehörigen Wkten erhalten wir jeweils durch Summation.

Ebenso $X = \{100\}$, $Y = \{010\}$, $Z = \{001\}$ und $F_1 = \{100, 010, 001\}$.

😊 Das explizite Modell (3) ist detailliert; das kann helfen. Oft genügt die implizite Schreibweise wie in (0–2); sie ist schnell und bequem.

⚠️ Es liegt immer ein Modell zu Grunde; im Zweifel führt man es aus.

Mikrozensus: Familienbande

Laut Mikrozensus 2009 haben von ca. 12 Mio Haushalten mit Kindern (vereinfacht) etwa 52% ein Kind, 36% zwei Kinder, 9% drei Kinder, 2% vier Kinder, 1% fünf Kinder. (Weitere vernachlässigen wir hier.) Die Wkten von Mädchen und Jungen seien gleich und unabhängig.

Aufgabe: (1) Stimmt die Schlagzeile „Jedes zweite Kind ist Einzelkind“? Wie viele sind Einzelkinder? oder eins von 2, 3, 4, 5 Geschwistern?

(2) Angenommen Sie haben später einmal drei Kinder. Mit welcher Wkt sind es zwei Mädchen und ein Junge?

(3) Angenommen Sie haben später einmal vier Kinder. Mit welcher Wkt sind es zwei Mädchen und zwei Jungs?

(4) Xanthia hat zwei Geschwister, dabei ist das ältere ein Mädchen. Mit welcher Wahrscheinlichkeit ist das andere Geschwist ein Junge?

(5) Yago hat zwei Geschwister, dabei (mindestens) eine Schwester. Mit welcher Wahrscheinlichkeit ist das andere Geschwist ein Junge?

(6) Sie lernen zufällig Zé kennen, von dem Sie zunächst nichts wissen. Wie groß ist die Wkt, dass er (mindestens) eine Schwester hat?

Mikrozensus: Familienbande

Lösung: (1) Es gab $\frac{52}{100} \cdot 12\text{Mio} = 6\,240\,000$ Einzelkinder,

$2 \cdot \frac{36}{100} \cdot 12\text{Mio} = 8\,640\,000$ Kinder mit genau einem Geschwist,

$3 \cdot \frac{9}{100} \cdot 12\text{Mio} = 3\,240\,000$ Kinder mit genau zwei Geschwistern

$4 \cdot \frac{2}{100} \cdot 12\text{Mio} = 960\,000$ Kinder mit drei Geschwistern und

$5 \cdot \frac{1}{100} \cdot 12\text{Mio} = 600\,000$ Kinder mit vier Geschwistern.

Insgesamt sind das 19 680 000 Kinder. Etwa jedes dritte ist Einzelkind.

😊 Die Zeitungsschlagzeile ist also falsch. Sie gründet, wie so oft, auf mangelnder Sachkenntnis. Man muss eine Statistik auch lesen können!

(2) In Geburtsreihenfolge: $\{\text{MMJ}, \text{MJM}, \text{JMM}\}$. Das sind drei von acht gleichwahrscheinlichen Möglichkeiten. Die Wkt beträgt $3/8 = 37.5\%$.

(3) Hier $\{\text{MMJJ}, \text{MJMJ}, \text{MJJM}, \text{JMMJ}, \text{JMJM}, \text{JJMM}\}$, also 6 von 16 gleichwahrscheinlichen Möglichkeiten. Die Wkt beträgt $6/16 = 37.5\%$.

😊 Diese beiden Fragen sind Spezialfälle der Binomialverteilung U3B.

(4) Das Geschlecht des jüngeren Geschwist ist unabhängig vom Geschlecht der älteren. Also ist die Wahrscheinlichkeit $1/2 = 50\%$.

Mikrozensus: Familienbande

(5) In Geburtsreihenfolge ist möglich: MM, MJ, JM, JJ.

Das Ereignis $A = \{\text{mindestens ein Bruder}\}$ hat Wkt $\mathbf{P}(A) = 3/4$.

Das Ereignis $B = \{\text{mindestens eine Schwester}\}$ hat Wkt $\mathbf{P}(B) = 3/4$.

Gefragt ist die Wkt $\mathbf{P}(A|B) = \mathbf{P}(A \cap B)/\mathbf{P}(B) = (1/2) / (3/4) = 2/3$.

Alternativ: Das entspricht zwei der drei Möglichkeiten MM, MJ, JM.

😊 Vergleichen Sie (4) und (5): Der Unterschied mag überraschen.

(6) Zé habe $n \in \{0, 1, 2, 3, 4\}$ Geschwister. Dann gilt:

$$\mathbf{P}(\text{Zé hat keine Schwester} \mid \text{Zé hat } n \text{ Geschwister}) = \frac{1}{2^n}$$

Als nächstes benötigen wir die Wahrscheinlichkeiten

$$\mathbf{P}(\text{Zé hat } n \text{ Geschwister}) = \frac{\text{Anzahl der Kinder mit } n \text{ Geschwistern}}{\text{Gesamtzahl aller Kinder}}.$$

Aus dem Mikrozensus (1) berechnen wir (auf drei Nachkommastellen):

$n =$	0	1	2	3	4
$p_n \approx$	0.317	0.439	0.165	0.049	0.030

Mikrozensus: Familienbande

Damit finden wir schließlich die totale Gesamtwahrscheinlichkeit (T2B):

$$\begin{aligned} & \mathbf{P}(\text{Zé hat keine Schwester}) \\ &= \sum_{n=0}^4 \mathbf{P}(\text{Zé hat } n \text{ Geschwister und keine Schwester}) \\ &= \sum_{n=0}^4 \left\{ \mathbf{P}(\text{Zé hat } n \text{ Geschwister}) \right. \\ & \quad \left. \cdot \mathbf{P}(\text{Zé hat keine Schwester} \mid \text{Zé hat } n \text{ Geschwister}) \right\} \\ &= 0.317 \cdot 1 + 0.439 \cdot \frac{1}{2} + 0.165 \cdot \frac{1}{4} + 0.049 \cdot \frac{1}{8} + 0.030 \cdot \frac{1}{16} \approx 0.586 \end{aligned}$$

Die gesuchte Wahrscheinlichkeit ist demnach das Komplement:

$$\begin{aligned} & \mathbf{P}(\text{Zé hat mindestens eine Schwester}) \\ &= 1 - \mathbf{P}(\text{Zé hat keine Schwester}) \\ &\approx 1 - 0.586 = 0.414 \end{aligned}$$

😊 Mit den Mitteln dieses Kapitels verfügen Sie über eine bequeme Notation und präzise Formeln. Der Rest ist sorgfältiges Rechnen.

Glücksritter: Ursprünge der Stochastik im Würfelspiel

Chevalier de Meré (1607–1684) war ein berufsmäßiger Spieler und suchte hierzu Rat bei dem Mathematiker Blaise Pascal (1623–1662).

Speziell plagte ihn folgendes Problem. Beim Würfelspiel war er davon ausgegangen, dass folgende Ereignisse gleich wahrscheinlich seien:

A: Bei 4 Würfeln eines Würfels tritt mindestens eine Sechs auf.

B: Bei 24 Würfeln zweier Würfel tritt mind. eine Doppelsechs auf.

Er argumentierte ganz einfach und anschaulich so: Eine Doppelsechs ist 6 mal unwahrscheinlicher, dafür führt man 6 mal mehr Würfe aus, also sind beide gleich wahrscheinlich. Damit hatte er jedoch auf Dauer Geld verloren, was ihn sehr wunderte – und vor allem sehr verdross.

Dieses praktische Problem wollte er nun mathematisch lösen.

Mit den Techniken dieses Kapitels gelingt dies spielend!

☹️ Wissen macht Ah! Unwissen macht Aua! Unkenntnis führt zu Fehleinschätzung und – schlimmer noch – zu Fehlentscheidungen.

😊 Die Mathematik bietet Ihnen hierzu extrem präzise Werkzeuge. Hier macht das den Unterschied, wie in jeder ernsthaften Anwendung.

Glücksritter: Ursprünge der Stochastik im Würfelspiel

Aufgabe: Wie wahrscheinlich sind diese Ereignisse A und B ?
Auf welches sollte unser Glücksritter demnach wetten?

Lösung: (A) Die Wahrscheinlichkeit, bei vier Würfeln keine Sechs zu werfen ist $(\frac{5}{6})^4 = \frac{625}{1296} \approx 0.4823$. Bei vier Würfeln mindestens eine Sechs zu werfen, hat somit die Gegenwahrscheinlichkeit $1 - 0.4823 \approx 0.5177$.

(B) Die Wahrscheinlichkeit, bei 24 Würfeln keine Doppelsechs zu werfen ist $(\frac{35}{36})^{24} \approx 0.5086$. Bei 24 Würfeln mindestens eine Doppelsechs zu werfen, hat somit die Gegenwahrscheinlichkeit $1 - 0.5086 \approx 0.4914$.

😊 Tatsächlich ist das Ereignis A etwas wahrscheinlicher als B .
Der Chevalier de Meré sollte demnach auf Ereignis A wetten.

⚠ Der kleine Unterschied von etwa 2.6% fällt nicht sofort auf.
Bei häufiger Wiederholung fällt er jedoch spürbar ins Gewicht.
Für professionelle Spieler wie Meré ist genaues Wissen wertvoll.

😊 Gleiches gilt für alle Prozesse, die häufig wiederholt werden,
und deren genaue Berechnung / Bewertung daher wesentlich ist.
Bei geringer Wiederholung hingegen verschwimmt die Trennschärfe.

Chancengleichheit: Wie fair ist ein fairer Würfel?

Bei einem fairen Würfel sind alle sechs Ergebnisse 1, 2, 3, 4, 5, 6 gleich wahrscheinlich. Manche verstehen „fair“ wesentlich strenger so, dass bei sechsmaligem Würfeln jedes der möglichen Ergebnisse 1, 2, 3, 4, 5, 6 auch einmal auftreten muss, sonst wär's ja irgendwie doch unfair.

- Aufgabe:** (1) Warum ist das keine vernünftige Definition von „fair“?
(2) Wie wahrscheinlich ist das? Was sagt Ihre Intuition? Rechnung?
(3) Können Sie dennoch eine realistische Vorhersage machen?

Lösung: (1) Der Würfel müsste wissen, was er noch würfeln darf. Insbesondere beim letzten Wurf wäre er festgelegt und nicht zufällig!

(2) Es gibt 6^6 Möglichkeiten (unter Beachtung der Reihenfolge). Alle sind gleich wahrscheinlich. Darunter sind $6!$ Möglichkeiten, bei denen jede der Zahlen 1, 2, 3, 4, 5, 6 genau einmal vorkommt. Die Wahrscheinlichkeit ist also $6!/6^6 = 720/46656 \approx 0.015$. Das ist unwahrscheinlich und tritt nur etwa in jeder 65. Serie auf.

(3) Das Gesetz der großen Zahlen (T3J) garantiert: Bei unabhängiger Wiederholung nähern sich die relativen Häufigkeiten den Wkten, hier also $1/6$. Sechsmalige Wiederholung ist aber noch keine große Zahl.

Efrons intransitive Würfel: Welcher ist der beste?

Aufgabe: Der Statistiker Bradley Efron erfand folgende Würfel:



Alice

$$A : 5, 5, 5, 1, 1, 1 \implies \mathbf{E}(A) = 9/3$$

$$B : 6, 6, 2, 2, 2, 2 \implies \mathbf{E}(B) = 10/3$$

$$C : 3, 3, 3, 3, 3, 3 \implies \mathbf{E}(C) = 9/3$$

$$D : 4, 4, 4, 4, 0, 0 \implies \mathbf{E}(D) = 8/3$$



Bob

Je zwei Würfel treten gegeneinander an, z.B. A gegen B . Wie groß sind die Gewinnwkten $\mathbf{P}(A > B)$ etc.? Welcher Würfel ist dabei der beste? Wie beschreiben Sie diese Situation präzise durch Zufallsvariablen?

Lösung: Unabhängigkeit! Abzählen aller Gewinnkombinationen ergibt:

$$\mathbf{P}(A > B) = 12/36 = 1/3, \quad \mathbf{P}(B > C) = 12/36 = 1/3,$$

$$\mathbf{P}(C > D) = 12/36 = 1/3, \quad \mathbf{P}(D > A) = 12/36 = 1/3,$$

$$\mathbf{P}(A > C) = 18/36 = 1/2, \quad \mathbf{P}(B > D) = 20/36 = 5/9.$$

Es gibt keinen „besten“ Würfel: Jeder wird vom nächsten geschlagen!

😊 Penney's Game: Intransitivität entsteht auch in zufälligen 0-1-Folgen beim Wettrennen von je zwei der acht Tripel: Wer schlägt hier wen?

Die geometrische Verteilung $G(q)$: Warten auf Godot

Satz T4A: geometrische Verteilung

Wir wiederholen ein Zufallsexperiment mit Misserfolgswkt $0 \leq q < 1$ unabhängig so lange, bis im n -ten Versuch der erste Erfolg eintritt.

Die Wahrscheinlichkeitsverteilung für die Wartezeit X ist gegeben durch

$$p : \mathbb{N}_{\geq 1} \rightarrow [0, 1] : n \mapsto p(n) = q^{n-1}(1 - q).$$

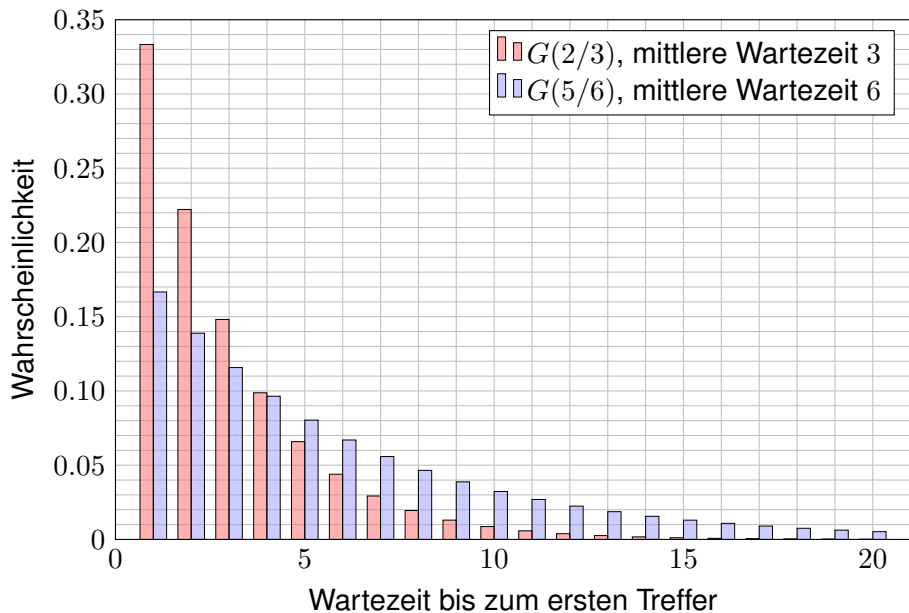
Dies nennen wir die **geometrische Verteilung** $G(q)$, geschrieben

$$X \sim G(q) = \text{Geom}(q).$$

Die mittlere Wartezeit bis zum ersten Erfolg und ihre Varianz sind

$$\mathbf{E}[X] = \sum_{n=1}^{\infty} n p(n) = \frac{1}{1 - q} \quad \text{und} \quad \mathbf{V}[X] = \frac{q}{(1 - q)^2}.$$

Beispiele: Lebensdauer eines Geräts, Wartezeit bis zum Ausfall, Würfeln bis die Sechs kommt beim Spiel *Mensch ärgere dich nicht*, Wartezeit auf den Jackpot beim Lotto oder ähnlichen Spielen. U416

Die geometrische Verteilung $G(q)$: Warten auf Godot

Erwartung der geometrischen Verteilung

Aufgabe: (0) Warum ist $G(q)$ tatsachlich eine WVerteilung?
Berechnen Sie hierzu (1) die Erwartung und (2) die Varianz.

Losung: (0) Zwei Eigenschaften sind zu prufen: Es gilt $p(n) \geq 0$ und

$$\sum_{n=1}^{\infty} p(n) = (1-q) \sum_{n=1}^{\infty} q^{n-1} = (1-q) \sum_{n=0}^{\infty} q^n = (1-q) \cdot \frac{1}{1-q} = 1.$$

(1) Wir berechnen zunachst den Erwartungswert:

$$\begin{aligned} \mathbf{E}[X] &= \sum_{n=1}^{\infty} n \mathbf{P}(X=n) = (1-q) \sum_{n=0}^{\infty} n q^{n-1} = (1-q) \sum_{n=0}^{\infty} \frac{d}{dq} q^n \\ &= (1-q) \frac{d}{dq} \left[\sum_{n=0}^{\infty} q^n \right] = (1-q) \frac{d}{dq} \left[\frac{1}{1-q} \right] = (1-q) \frac{1}{(1-q)^2} = \frac{1}{1-q} \end{aligned}$$

☺ Der geometrischen Reihe sei Dank! Hier hilft der Ableitungstrick:
Die Rechenregeln fur Reihen kennen und nutzen Sie seit der HM2.

☺ Das Ergebnis ist bemerkenswert: Die Wartezeit kann beliebig gro
werden, aber ihre Erwartung ist endlich, da die Wkt schnell sinkt mit q^n .

Varianz der geometrischen Verteilung

(2) Ebenso berechnen wir $\mathbf{E}[X^2]$, dazu zunachst $\mathbf{E}[X(X - 1)]$:

$$\begin{aligned} \mathbf{E}[X(X - 1)] &= \sum_{n=0}^{\infty} n(n - 1) \mathbf{P}(X=n) = (1 - q)q \sum_{n=0}^{\infty} n(n - 1)q^{n-2} \\ &= (1 - q)q \sum_{n=0}^{\infty} \frac{d^2}{dq^2} q^n = (1 - q)q \frac{d^2}{dq^2} \left[\sum_{n=0}^{\infty} q^n \right] \\ &= (1 - q)q \frac{d^2}{dq^2} \left[\frac{1}{1 - q} \right] = (1 - q)q \frac{2}{(1 - q)^3} = \frac{2q}{(1 - q)^2} \end{aligned}$$

Hieraus erhalten wir schlielich die ersehnte Varianz:

$$\begin{aligned} \mathbf{V}[X] &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \mathbf{E}[X(X - 1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 \\ &= \frac{2q^2}{(1 - q)^2} + \frac{1}{1 - q} - \frac{1}{(1 - q)^2} = \frac{q}{(1 - q)^2} \end{aligned}$$

😊 Die Rechnung ist trickreich, aber jeder einzelne Schritt ist leicht: Fur Reihen kennen Sie alle notigen Rechenregeln aus der HM2.

😊 Auch hier hilft erneut der Ableitungstrick: Sie nutzen gewinnbringend die allgegenwartigen, extrem effizienten Rechenregeln fur Potenzreihen!

Stoppszeiten: Wurfeln bis die Eins kommt

Beim Wurfeln werden Augenzahlen addiert bis der Spieler aufhort; wurfelt er jedoch eine Eins, so endet sein Spiel mit Totalverlust/Ruin.

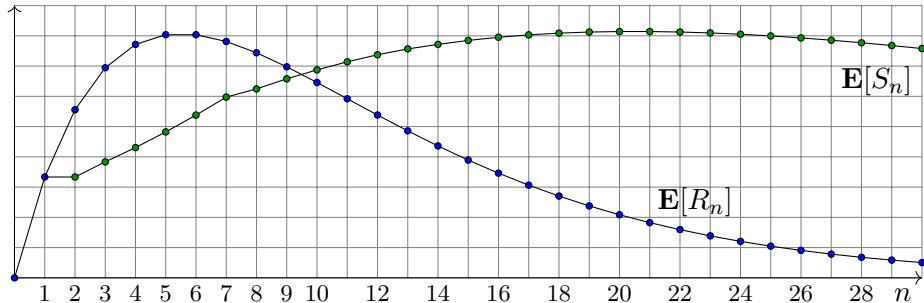
Die Spielleiterin zahlt die erreichte Augensumme in Euro, verlangt aber 10€ Einsatz. Wie spielen Sie optimal? Konnen Sie Profit generieren?

Aufgabe: Sie nutzen die Strategie R_n : „Spiele n Runden (oder Ruin).“

(1) Welchen Gewinn erwarten Sie? Welche n maximieren den Gewinn?

Losung: (1) Bei n Wurfeln erwarten wir den Gewinn $\mathbf{E}[R_n] = 4n\left(\frac{5}{6}\right)^n$.

Bei $n \in \{5, 6\}$ erreichen wir das Maximum $4 \cdot 5\left(\frac{5}{6}\right)^5 = 4 \cdot 6\left(\frac{5}{6}\right)^6 \approx 8.038$.



Stoppszeiten: Würfeln bis die Eins kommt

Die Erwartung $\mathbf{E}[R_5] \approx 8$ ist noch zu gering. Gibt es bessere Strategien? Angenommen, im laufenden Spiel beträgt Ihre aktuelle Augensumme s . Sollten Sie noch einen weiteren Zug wagen oder besser jetzt aufhören?

Aufgabe: (2) Bis zu welcher Augensumme s lohnt sich ein weiterer Zug?

Lösung: (2) Nach einem weiteren Zug ist die erwartete Augensumme

$$e(s) = \frac{0 + (s+2) + (s+3) + (s+4) + (s+5) + (s+6)}{6} = \frac{5s + 20}{6}$$

Der nächste Zug ist genau dann profitabel (im Mittel), wenn $e(s) > s$ gilt:

$$s < \frac{5s + 20}{6} \iff s < 20$$

Das legt folgende Strategie nahe: Sie würfeln bis Sie Augensumme ≥ 20 erreichen (oder aber eine Eins zuvor Ihr Spiel mit Totalverlust ruiniert.)

Ist diese flexible Strategie wirklich besser als R_n mit fest 5 Runden? Können Sie damit sogar die 10€ Einsatz übertreffen und profitieren? Dazu müssen wir die Erwartungswerte berechnen und vergleichen!

Stoppszeiten: Würfeln bis die Eins kommt

Die Strategie S_n lautet: „Spiele bis mindestens n Punkte (oder Ruin).“ Anders als bei R_n nutzen Sie Informationen des bisherigen Verlaufs!

Aufgabe: (3) Welches Ergebnis liefert die Strategie S_n mit welcher Wkt?
 (4) Welche der Strategien in $(S_n)_{n \in \mathbb{N}}$ erwartet den höchsten Gewinn?

Lösung: (3) Sei $n \in \mathbb{N}_{\geq 1}$. Die Strategie S_n liefert die Ergebnisse $0, n, n+1, n+2, n+3, n+4, n+5$ mit Wkten $a_n, b_n, c_n, d_n, e_n, f_n, g_n \in [0, 1]$. Für $n = 1$ sind diese Wkten offensichtlich $1/6, 0, 1/6, 1/6, 1/6, 1/6, 1/6$.

Rekursiv gilt dann $a_{n+1} = a_n + b_n/6$, $b_{n+1} = c_n$, $c_{n+1} = d_n + b_n/6$,
 $d_{n+1} = e_n + b_n/6$, $e_{n+1} = f_n + b_n/6$, $f_{n+1} = g_n + b_n/6$, $g_{n+1} = b_n/6$.

Mit einer Tabellenkalkulation erhalten wir mühelos die folgenden Werte; Sie finden dieses Beispiel unter eiser.m.de/lehre/HM3/Wuerfel.n.ods.

(4) Der erwartete Gewinn für die Strategie S_n ist gegeben durch
 $\mathbf{E}[S_n] = nb_n + (n+1)c_n + (n+2)d_n + (n+3)e_n + (n+4)f_n + (n+5)g_n$.

Das Maximum $\mathbf{E}[S_{20}] = \mathbf{E}[S_{21}] \approx 8.142$ liegt nur ein klein Wenig höher als zuvor $\mathbf{E}[R_5] = \mathbf{E}[R_6] \approx 8.038$ bei fester Rundenzahl. Immerhin! Für 10€ Einsatz genügt das nicht. Können Sie mehr rausholen?

Stoppszeiten: Würfeln bis die Eins kommt

S_n	0	n	$n+1$	$n+2$	$n+3$	$n+4$	$n+5$	$E[S_n]$	S_n	0	n	$n+1$	$n+2$	$n+3$	$n+4$	$n+5$	$E[S_n]$
1	.167	.000	.167	.167	.167	.167	.167	3.333	26	.714	.076	.072	.057	.041	.027	.013	7.931
2	.167	.167	.167	.167	.167	.167	.000	3.333	27	.726	.072	.069	.054	.040	.026	.013	7.855
3	.194	.167	.194	.194	.194	.028	.028	3.833	28	.738	.069	.066	.052	.038	.025	.012	7.771
4	.222	.194	.222	.222	.056	.056	.028	4.306	29	.750	.066	.063	.049	.036	.024	.012	7.679
5	.255	.222	.255	.088	.088	.060	.032	4.824	30	.761	.063	.060	.047	.035	.023	.011	7.579
6	.292	.255	.125	.125	.097	.069	.037	5.380	31	.771	.060	.058	.045	.033	.022	.011	7.474
7	.334	.125	.167	.140	.112	.079	.042	5.974	32	.781	.058	.055	.043	.032	.021	.010	7.363
8	.355	.167	.160	.133	.100	.063	.021	6.245	33	.791	.055	.053	.041	.030	.020	.010	7.248
9	.383	.160	.161	.128	.091	.049	.028	6.579	34	.800	.053	.050	.039	.029	.019	.009	7.128
10	.410	.161	.155	.118	.075	.055	.027	6.874	35	.809	.050	.048	.038	.028	.018	.009	7.005
11	.436	.155	.145	.102	.081	.054	.027	7.141	36	.818	.048	.046	.036	.026	.017	.008	6.878
12	.462	.145	.128	.107	.079	.053	.026	7.374	37	.826	.046	.044	.034	.025	.016	.008	6.750
13	.486	.128	.131	.103	.077	.050	.024	7.567	38	.833	.044	.042	.033	.024	.016	.008	6.619
14	.508	.131	.125	.098	.071	.045	.021	7.716	39	.841	.042	.040	.031	.023	.015	.007	6.487
15	.530	.125	.120	.093	.067	.043	.022	7.848	40	.848	.040	.039	.030	.022	.014	.007	6.353
16	.550	.120	.114	.088	.064	.043	.021	7.952	41	.854	.039	.037	.029	.021	.014	.007	6.219
17	.570	.114	.108	.084	.063	.041	.020	8.032	42	.861	.037	.035	.028	.020	.013	.006	6.084
18	.589	.108	.103	.082	.060	.039	.019	8.089	43	.867	.035	.034	.026	.019	.013	.006	5.949
19	.607	.103	.100	.078	.057	.037	.018	8.125	44	.873	.034	.032	.025	.018	.012	.006	5.814
20	.625	.100	.095	.074	.054	.035	.017	8.142	45	.878	.032	.031	.024	.018	.011	.006	5.680
21	.641	.095	.091	.071	.052	.034	.017	8.142	46	.884	.031	.029	.023	.017	.011	.005	5.546
22	.657	.091	.087	.068	.050	.032	.016	8.126	47	.889	.029	.028	.022	.016	.010	.005	5.412
23	.672	.087	.083	.065	.048	.031	.015	8.096	48	.894	.028	.027	.021	.015	.010	.005	5.280
24	.687	.083	.079	.062	.045	.030	.014	8.052	49	.898	.027	.026	.020	.015	.010	.005	5.149
25	.700	.079	.076	.059	.043	.028	.014	7.997	50	.903	.026	.025	.019	.014	.009	.004	5.019

Nun zu etwas völlig anderem. . .

Es ist überaus bemerkenswert, dass wir mit Wkten rechnen können! Anfangs erschien der Zufall allzu vage, unfassbar, unberechenbar. Doch auch der Zufall folgt gewissen Gesetzen und Rechenregeln.

Mit den Mitteln dieses Kapitels verfügen Sie über eine bequeme und präzise Notation zur Berechnung von Wahrscheinlichkeiten. Insbesondere die Formel von Bayes ist sehr einfach, aber nützlich: Sie klärt viele Rechnungen, die ansonsten undurchdringlich blieben.

Sie können dies in den unterschiedlichsten Situationen einsetzen: Konfidenzintervalle [T337] [W201], Testverfahren [T408], Partnerwahl [T437], Bewerbung [T439], Glücksspiel [T441] [T445] oder Googles PageRank [T453]. Versuchen wir es mal mit den großen Fragen der Menschheit. . .

There is a theory which states that if ever anybody discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

(Douglas Adams, 1952–2001)

Seul au monde?



*Man sieht nur mit der Mathematik gut.
Das Wesentliche ist für die Augen unsichtbar.*
(frei nach Antoine de Saint-Exupéry, 1900–1944)

Sind Sie das einzige denkende Wesen?

Als Kopfnuss hier eine erstaunliche Rechnung, die nicht ganz ernst gemeint ist. (Oder doch?) Es gibt Menschen, die glauben, sie seien das einzige denkende Wesen im Universum. Für den Fall, dass auch Sie sich diese Frage stellen, schlage ich Ihnen ein stochastisches Argument vor.

(a) Angenommen, es gibt genau X **denkende Wesen** im Universum. Diese Anzahl X ist Ihnen nicht bekannt, Sie nehmen deshalb die Gleichverteilung auf der Menge $\{1, 2, 3, \dots, n\}$ an, etwa $n < e^{100}$.

Wir wollen nur annehmen, dass n sehr groß ist, aber doch kleiner als die geschätzten 10^{80} Elementarteilchen im Universum. (Paralleluniversen werden nicht mitgerechnet, wie auch?)

(b) Angenommen, **nur eines** der Wesen hat zufällig die Erleuchtung: *Ich denke, also bin ich. . . Bin ich das einzige denkende Wesen?*

Das ist von allen Annahmen die willkürlichste und rechnerisch auch die einschneidendste. Andererseits: Wer sich wirklich fragt, ob er das einzige denkende Wesen im Universum ist, der wird auch vor dieser Annahme der Einzigkeit nicht zurückschrecken. Also sei's drum.

Aufgabe: Angenommen, dieses denkende Wesen sind zufällig Sie. Mit welcher Wahrscheinlichkeit können Sie nun zurückschließen, dass Sie das einzige denkende Wesen im Universum sind?

Sind Sie das einzige denkende Wesen?

Lösung: Nach Annahme der Gleichverteilung (a) wissen Sie

$$\mathbf{P}(X=1) = \mathbf{P}(X=2) = \dots = \mathbf{P}(X=n) = \frac{1}{n}.$$

Ihre Erleuchtung E hat gemäß (b) die Wahrscheinlichkeit

$$\mathbf{P}(E|X=k) = \frac{1}{k}.$$

Ihre Einzigkeit $\{X=1\}$ hat nach Bayes die Wahrscheinlichkeit

$$\begin{aligned} \mathbf{P}(X=1|E) &= \frac{\mathbf{P}(E|X=1) \mathbf{P}(X=1)}{\sum_{k=1}^n \mathbf{P}(E|X=k) \mathbf{P}(X=k)} \\ &= \frac{\frac{1}{n}}{\sum_{k=1}^n \frac{1}{k} \cdot \frac{1}{n}} = \frac{1}{\sum_{k=1}^n \frac{1}{k}} \geq \frac{1}{1 + \ln n} > \frac{1}{100} = 1\% \end{aligned}$$

Zur harmonischen Reihe siehe B302. Unter den genannten Annahmen (a) und (b) gibt es also eine nicht zu vernachlässigende Wahrscheinlichkeit von mindestens 1%, dass Sie das einzige denkende Wesen im Universum sind. Wie gut, dass wir das klären konnten. Ich gratuliere!

Objektive und subjektive Wahrscheinlichkeit

Ein Wahrscheinlichkeitsraum (Ω, \mathbf{P}) ist ein **theoretisches Modell**:
Es dient zur Grundlegung und Formulierung unserer Berechnungen.
Was sagt es über die **Wirklichkeit**? Was bedeuten Aussagen wie:
„Das Ereignis A tritt mit Wahrscheinlichkeit $\mathbf{P}(A)$ ein“?

Objektive Interpretation: Die Wahrscheinlichkeit eines Ereignisses ist die relative Häufigkeit bei **beliebig wiederholbaren Experimenten**.
Sie kann empirisch überprüft werden, wir können sie messen! T345

- Bei einem fairen Würfel erscheinen gerade Zahlen mit Wkt $1/2$.
- Bei unserem gezinkten Würfel haben gerade Zahlen Wkt 53% .
- Die Wkt für 6 Richtige beim Lotto ist Eins zu $\binom{49}{6} = 13\,983\,816$.

Subjektive Interpretation: Die Wahrscheinlichkeit drückt den Grad unseres Unwissens aus, also die „**gefühlte Wahrscheinlichkeit**“.
Dies hat nur einen vagen Sinn und ist empirisch kaum überprüfbar.

- Morgen beträgt die Regenwahrscheinlichkeit 70% .
- Die Klausur werde ich mit 90% Wahrscheinlichkeit bestehen.
- Mit Wkt 1% bin ich das einzige denkende Wesen im Universum.

Objektive und subjektive Wahrscheinlichkeit

Die objektive Sichtweise heißt auch **frequentistische Interpretation**: Wir betrachten nur Experimente, die wir „beliebig oft“ unter den „gleichen Bedingungen“ wiederholen können. Die Wkt eines Ereignisses ist der Grenzwert der relativen Häufigkeit bei „sehr vielen“ Wiederholungen.

Damit erscheint die Wkt als eine **physikalische Größe**, die zwar Messfehlern unterliegt, die wir aber beliebig genau messen können. Diese Interpretation ist inspiriert durch physikalische Laborexperimente, darunter auch Glücksspiele. Sie versagt allerdings bei Experimenten, die wir nicht wiederholen können, etwa gar nicht oder nicht oft genug.

Die subjektive Sichtweise heißt auch **Bayes'sche Interpretation**: Die Wkt $P(A) \in [0, 1]$ eines Ereignisses A drückt aus, wie überzeugt der subjektive Beobachter vom Eintreten ist. In gewissen Grenzen können wir auch dies grob messen: „Wie viel bist du bereit zu setzen?“

Damit erscheint die Wkt als eine **psychologische Größe**; das ist für viele Anwendungen pragmatisch, aber leider vollkommen subjektiv. In dieser Sichtweise ist es ganz egal, ob wirklich Zufall im Spiel ist. Oft werden die objektive und die subjektive Interpretation vermischt!

Objektive und subjektive Wahrscheinlichkeit

Beispiel: Sie kaufen ein Los mit Gewinnwkt $1:14\text{Mio}$ und reisen damit in die Berge zum Wandern. Die Ziehung findet statt, doch fernab der Zivilisation wissen Sie davon nichts. Wie groß ist jetzt Ihre Gewinnwkt? Die Aussage „Sie haben gewonnen.“ ist nun entweder wahr oder falsch, und somit keine Frage der Wahrscheinlichkeit mehr! Oder etwa doch?

Beispiel: Diese Situation können wir hier vor Ort sofort nachstellen: Die Spielleiterin wirft eine faire Münze, hält diese aber noch verdeckt. Wie stehen jetzt Ihre Gewinnchancen? objektiv? subjektiv?

Nach dem Ereignis hat es keinen Sinn, nach Wahrscheinlichkeit zu fragen.

(Albert Einstein, 1879–1955)

Beispiel: Drei gleich gute Studenten A, B, C schreiben eine Klausur. Vorläufiges Ergebnis: Zwei der drei haben die Bestnote Eins, doch die Einzelnoten sind noch nicht bekannt. Student A bezirzt die Assistentin, ihm einen (anderen) Studenten zu nennen, der eine Eins hat, und nach einigem Zögern sagt sie: „ C hat eine Eins.“ Daraufhin ist A sehr traurig, denn er fühlt seine Chance von $2/3$ auf $1/2$ gesunken. Zu Recht? T213

Objektive und subjektive Wahrscheinlichkeit

Diese grundlegenden Fragen sind nicht bloß spitzfindige Haarspalterei: Bei der Erklärung von **Konfidenzintervallen** geht es genau darum! Wissenschaft sucht objektive Aussagen und begründete Folgerungen. Äußerungen, die keine Prüfung erlauben, haben keinen objektiven Sinn.

Das ist nicht nur nicht richtig, es ist nicht einmal falsch! [not even wrong]
(Wolfgang Pauli, 1900–1958)

Präzise Aussagen und rationales Entscheiden folgen strengen Regeln: Dies ist das übliche und bewährte Vorgehen der Naturwissenschaften!

Theorie: Die präzise Beschreibung des zugrundegelegten Modells dient der wissenschaftlichen Redlichkeit: Es enthält alle Annahmen und die festgelegten Wahlen, Parameter, Kalibrierungen etc. Alle Rechnungen bauen auf genau diesen Daten auf — und auf nichts weiter.

Empirie: Ob ein Modell zur Beschreibung der Realität geeignet ist, oder ob von zwei Modellen womöglich eines besser ist als das andere, lässt sich empirisch feststellen: Wir vergleichen hierzu die Vorhersagen und prüfen sie durch Experimente – stochastisch unabhängig wiederholt!

Was nützen dem Pfau seine Federn?

Pfauen-Männchen haben ein prachtvolles Gefieder. Für das Überleben ist es leider hinderlich: Tarnung, Flucht, Energie, ... Warum lohnen sich dennoch lange Federn? Wozu dienen sie? Angenommen, es gebe „fitte“ und „unfitte“ Pfauen-Männchen. Letztere werden öfter von Raubtieren gefressen etc. Sowohl fitte als auch unfitte können kurze oder lange Federn haben. Die mit langen Federn werden aufgrund der genannten Nachteile jedenfalls öfter gefressen.

Eine Population von Pfauen-Männchen könnte etwa so aussehen:

Pfauen	fit	unfit
kurze Federn	40%	20%
lange Federn	30%	10%

Das Pfauen-Weibchen sieht nicht die Fitness, sondern nur die Federn!

Aufgabe: Welche Strategie der Partnerwahl ist für sie vorteilhaft?

- Federlänge ignorieren: Trefferquote 70 : 30 (möglich)
- Kurze Federn bevorzugen: Trefferquote 40 : 20 (schlechter)
- Lange Federn bevorzugen: Trefferquote 30 : 10 (besser!)

😊 Trotz aller Nachteile zahlen sich lange Federn doch aus!

Visuelles Ornament und sexuelle Selektion

Das auffällige Gefieder der Pfauen-Männchen wird als *visuelles Ornament* bezeichnet. Die lange Schleppe ist zwar hinderlich und vermindert das Flugvermögen, paradoxerweise kann es dennoch ein Indikator für genetische Fitness sein und Pfauen-Weibchen als Indiz für gesunden Nachwuchs dienen. (Siehe unser Zahlenbeispiel.) Wäre es nicht effizienter, Pfauen würden diesen Aufwand sparen und die Energie in bessere Überlebensstrategien stecken? Auf den ersten Blick besehen ja. Aber für ein Weibchen gibt es keine bessere Wahl: Es sieht nicht die tatsächliche Fitness, sondern nur das *Signal* der Federn. Auf Grundlage dieser unvollständigen Information muss es sich für die langen Federn entscheiden, selbst wenn das Überlebensnachteile mit sich bringt, auch für seinen eigenen Nachwuchs! Damit bleibt auch den Männchen keine Wahl, lange Federn werden zum Must-Have. Diese *sexuelle Selektion* ist der Schlussstein in Darwins Evolutionstheorie.

Auch Menschen ist dieses Phänomen vertraut: Manche Männer beeindrucken Frauen, indem sie Geld verschwenden oder andere verrückte Dinge tun. Wäre dieses Verhalten insgesamt nachteilig, so würde man vermuten, dass es auf lange Sicht ausstirbt. Das ist jedoch nicht der Fall, da Frauen dieses Verhalten als Indikator für (gesellschaftlichen) Erfolg interpretieren können und eventuell bei der Partnerwahl belohnen. (All das funktioniert selbstverständlich auch umgekehrt. . .)

Dieses Phänomen wird *Handicap-Prinzip* genannt. Beispiele gibt es viele: Produktwerbung verschwendet Geld, wird aber vom Käufer belohnt. Manch akademischer Titel ist scheinbar Zeitverschwendung, wird aber vom Arbeitgeber honoriert. Auch Ihr Studium ist nur teilweise für Ihren späteren Beruf relevant, und dennoch wird diese Anstrengung meist belohnt. Zum Beispiel gelten Leistungen in Mathematik als zuverlässiger Indikator für intellektuelle Leistungsfähigkeit. Dazu diskutieren wir den Arbeitsmarkt, extrem vereinfacht, in folgendem Gedankenexperiment.

Kann ein unnützer Dokortitel doch nützlich sein?

Eine Personalchefin sucht für eine Stelle einen Ingenieur (m/w/d). Aus Erfahrung schätzt sie die allgemeine Bewerberlage wie folgt:

Bewerber	geeignet	ungeeignet
Diplom/Master	50%	25%
Promotion	20%	5%

Als individuelle Information hat sie zunächst nur den *Abschluss* laut Bewerbungsunterlagen. Die eigentlich interessante Zielgröße der *tatsächlichen Eignung* kennt sie hingegen nicht. Eine Promotion kostet Zeit und Mühe, bringt aber für *diese* Stelle keinen direkten Nutzen.

Aufgabe: Welche Strategie ist bei ihrer Auswahl vorteilhaft?

- Abschluss ignorieren: Trefferquote 70 : 30 (möglich)
- Master einstellen: Trefferquote 50 : 25 (schlechter)
- Doktor einstellen: Trefferquote 20 : 5 (besser!)

😊 Trotz aller Nachteile kann sich eine Promotion also auszahlen ... selbst wenn sie für die eigentliche Tätigkeit nicht relevant ist!

⚠ Ineffizienz ist der Preis für unvollständige Information.

Bedingte Wkt: Vorurteil oder Gerechtigkeit?

Wir untersuchen hier zwei simple, aber frappierende Beispiele: Federn und Dokortitel. Beide sind durchaus realistisch und handfeste Illustrationen für das Konzept der *bedingten Wahrscheinlichkeit*: Diese ist nicht nur eine schöne Theorie, sondern überall tägliche Praxis.

Die Argumente in unserer fiktiven Bewerbungssituation mögen manche für ungerecht halten. In der Tat basieren Sie auf *Vorurteilen* der Arbeitgeberin – ein eher negativ besetzter Ausdruck, aber inhaltlich bedeutet es dasselbe wie bedingte Wahrscheinlichkeit: Sie nutzt ihre Erfahrung. Unter den gegebenen spärlichen Informationen ist das Vorurteil nützlicher als gar kein Urteil.

Das Grundproblem: Die primäre Zielgröße „Eignung“ ist nicht direkt zugänglich. Der sekundäre Faktor „Ornament“ ist eigentlich unwichtig, dafür aber leicht sichtbar. In Ermangelung primärer Information muss man sich mit sekundärer Information begnügen. Diese erhält dadurch eine größere Bedeutung als sie eigentlich haben sollte, und das wird als ineffizient oder ungerecht empfunden. Das ist der Preis für unvollständige Information!

Zur Beruhigung der Gemüter: Nichts hält die Arbeitgeberseite davon ab, über die erste grobe Vorinformation hinaus genauere Information zu gewinnen, zum Beispiel durch Gespräche, Tests, Assessment oder eine Probezeit. Genau das wird in der Praxis auch erfolgreich genutzt. Das ist der Vorteil, wenn man Information nicht nur *passiv* beobachtet, sondern *aktiv* herstellen kann.

Schließlich zur Ehrenrettung der Promotion, auch aus persönlicher Erfahrung: Viele Studierende empfinden große Begeisterung für ihr Fach. Dies kann sogar dazu führen, dass sie aus ehrlichem intrinsischem Interesse einer Frage auf den Grund gehen wollen und darüber sogar promovieren. Das wird durch die obigen, allzu kühl berechnenden Argumente nicht in Zweifel gezogen!

Das Runde muss ins Eckige: Bayes beim Tischkicker

Angenommen beim Tischkicker bis 10 zwischen den gleich starken Alice und Bob fallen die Tore zufällig (50 : 50) und unabhängig voneinander.

Aufgabe: Wie stehen die Gewinnchancen bei 9 : 8? bei 4 : 7?

Lösung: Wahrscheinlichkeitstabelle für den Sieg von Alice beim Stand $a : b$.

Am oberen und linken Rand ist das Spiel beendet und die Wkt ist entweder 0 oder 1.

Im Inneren ist jeder Eintrag der Mittelwert aus dem linken und dem oberen Nachbarn.

$a : b$	10	9	8	7	6	5	4	3	2	1	0
10		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0.000	0.500	0.750	0.875	0.938	0.969	0.984	0.992	0.996	0.998	0.999
8	0.000	0.250	0.500	0.688	0.813	0.891	0.938	0.965	0.980	0.989	0.994
7	0.000	0.125	0.313	0.500	0.656	0.773	0.855	0.910	0.945	0.967	0.981
6	0.000	0.063	0.188	0.344	0.500	0.637	0.746	0.828	0.887	0.927	0.954
5	0.000	0.031	0.109	0.227	0.363	0.500	0.623	0.726	0.806	0.867	0.910
4	0.000	0.016	0.063	0.145	0.254	0.377	0.500	0.613	0.709	0.788	0.849
3	0.000	0.008	0.035	0.090	0.172	0.274	0.387	0.500	0.605	0.696	0.773
2	0.000	0.004	0.020	0.055	0.113	0.194	0.291	0.395	0.500	0.598	0.685
1	0.000	0.002	0.011	0.033	0.073	0.133	0.212	0.304	0.402	0.500	0.593
0	0.000	0.001	0.006	0.019	0.046	0.090	0.151	0.227	0.315	0.407	0.500

Das Runde muss ins Eckige: Bayes beim Tischkicker

Ausführung: Wie kommt diese Rechnung zustande? Ganz einfach dank der Formel von Bayes! Beim Stand von $a : b$ gibt es zwei mögliche Fortgänge: Entweder es trifft Alice oder es trifft Bob. Wir entwickeln unsere Rechnung übersichtlich in einer Tabelle. Sei $A[a:b]$ das Ereignis „Alice gewinnt nach Stand $a : b$ “. Dank Satz T2B gilt für die totale Wahrscheinlichkeit:

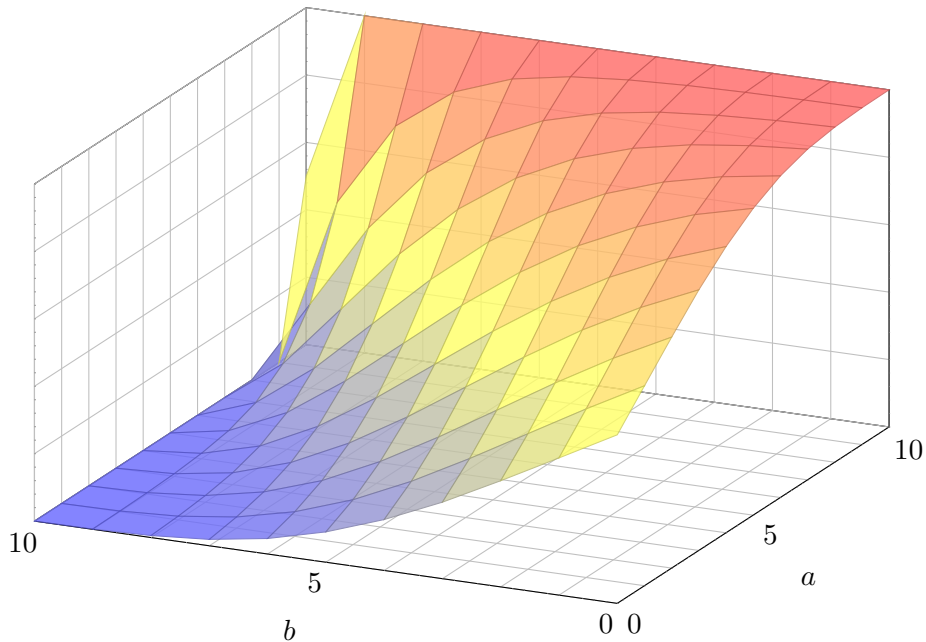
$$\begin{aligned} \mathbf{P}(A[a:b]) &= \mathbf{P}(A[a:b] \mid \text{Alice trifft}) \cdot \mathbf{P}(\text{Alice trifft}) + \mathbf{P}(A[a:b] \mid \text{Bob trifft}) \cdot \mathbf{P}(\text{Bob trifft}) \\ &= \mathbf{P}(A[a+1:b]) \cdot \mathbf{P}(\text{Alice trifft}) + \mathbf{P}(A[a:b+1]) \cdot \mathbf{P}(\text{Bob trifft}) \end{aligned}$$

☺ Mit dieser einfachen Rekursionsformel können Sie nun die gesamte Tabelle ausfüllen. Besonders bequem und automatisiert geht's mit einer Tabellenkalkulation, z.B. *LibreOffice*.

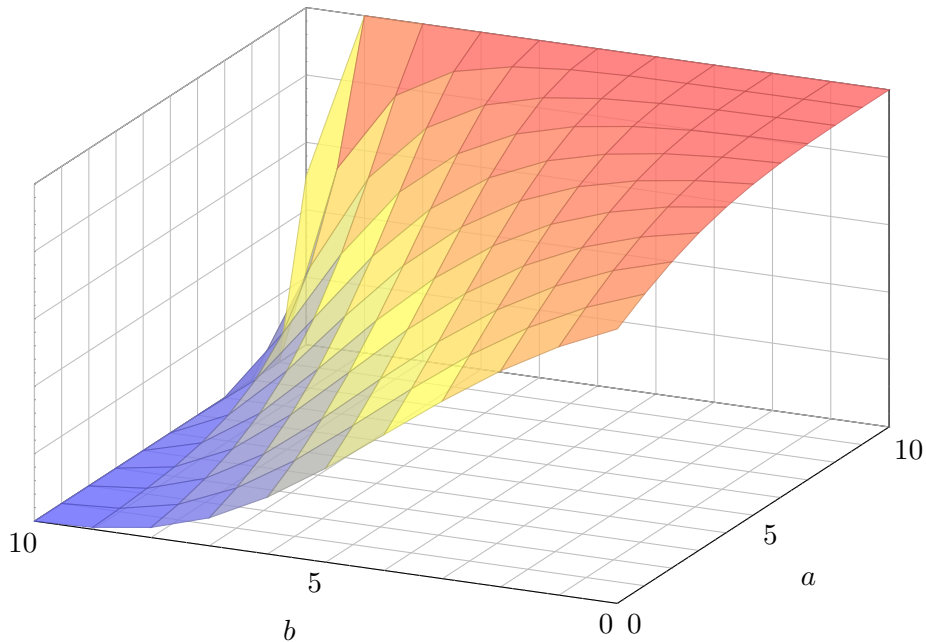
☺ Pascal lässt grüßen: Die n te Diagonale der Tabelle entspricht der n ten Zeile des Pascalschen Dreiecks (hier kumuliert und normiert). In unserer Tabelle versteckt sich die Binomialverteilung!

Geschichte: Mitte 1654 schrieb Blaise Pascal (1623–1662) an Pierre de Fermat (1607–1665) einen Brief, der berühmt wurde und als Geburtsurkunde der Wahrscheinlichkeitsrechnung gilt. Pascal löste darin zwei konkrete Probleme zu Glücksspielen, zu denen ein Freund ihn um Rat gebeten hatte, der berufsmäßige Spieler Antoine Gombaud, Chevalier de Méré (1607–1684). Erstens eine Berechnung von Würfelwahrscheinlichkeiten. T417 Zweitens das Problem der gerechten Teilung bei vorzeitigem Spielabbruch. Was heißt hier „gerecht“? Pascal und Fermat wurden sich nach ausführlicher Diskussion einig, dass der Einsatz gemäß den Gewinnwkten aufgeteilt werden sollte. Das ist die obige Tabelle. Pascal konnte diese kombinatorisch berechnen dank dem von ihm zuvor entwickelten Pascalschen Dreieck für die Binomialkoeffizienten $\binom{n}{k}$.

Das Runde muss ins Eckige: Bayes beim Tischkicker

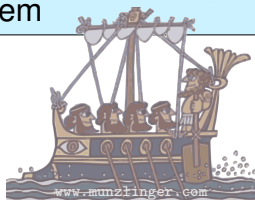
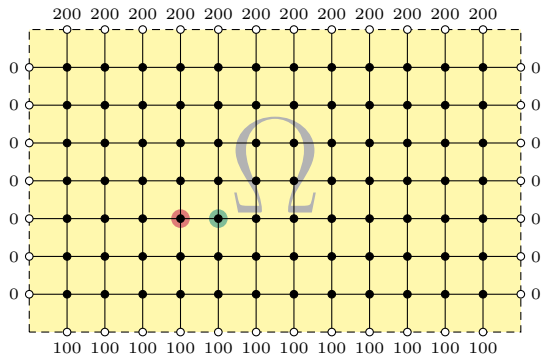


Das Runde muss ins Eckige: Bayes beim Tischkicker



Irrfahrten und Potentiale: das Dirichlet–Problem

Zufällige Irrfahrt auf einem Spielfeld $\Omega \subseteq \mathbb{Z}^2$:



Sie ziehen mit Wkt $1/4$ nach links / rechts / oben / unten. Das Spiel endet am Rand mit dem gezeigten Gewinn. Für welche Startpunkte würden Sie 100 zahlen?

Aufgabe: (1) Wie groß ist die Gewinnerwartung $u(x, y)$ auf jedem Feld? Wo ist sie maximal? Ist die gesuchte Lösung $u : \Omega \rightarrow \mathbb{R}$ eindeutig? Wie berechnet man sie? möglichst effizient? näherungsweise?

😊 Kontext und Anwendung ändern sich, die Rechnung bleibt dieselbe!
 (2) Hooke: Netz aus Massen und Federn. (3) Kirchhoff: Spannung einer elektrischen Schaltung. (4) Fourier: diskrete Wärmeleitung / Diffusion.

Irrfahrten und Potentiale: das Dirichlet–Problem

Lösung: (1) Sei $u(x, y)$ die Gewinnerwartung auf dem Feld $(x, y) \in \Omega$. In jedem Randpunkt $(x, y) \in \partial\Omega$ ist der Gewinn $u(x, y)$ fest vorgegeben. In jedem inneren Punkt $(x, y) \in \Omega^\circ$ gilt die **Mittelwerteigenschaft**:

$$u(x, y) = \frac{1}{4}u(x-1, y) + \frac{1}{4}u(x+1, y) + \frac{1}{4}u(x, y-1) + \frac{1}{4}u(x, y+1)$$

Eine solche diskrete Funktion $u : \mathbb{Z}^2 \supset \Omega \rightarrow \mathbb{R}$ nennen wir **harmonisch**.

😊 Die folgende Tabelle und Graphik zeigen die / eine Lösung u .

Wir betrachten hier das ebene Gitter \mathbb{Z}^2 und darin eine Teilmenge $\Omega \subseteq \mathbb{Z}^2$. Innere Punkte $z \in \Omega$ sind solche, deren vier direkte Nachbarn ebenfalls in Ω liegen. Bei einem Randpunkt $z \in \Omega$ liegt mindestens ein Nachbar außerhalb von Ω . **Dirichlet–Problem:** In jedem Randpunkt $z \in \partial\Omega$ ist der Wert $u(z)$ festgelegt durch die vorgegebene Randfunktion $v = u|_{\partial\Omega} : \partial\Omega \rightarrow \mathbb{R}$. Gesucht ist eine harmonische Funktion $u : \Omega \rightarrow \mathbb{R}$ mit $u|_{\partial\Omega} = v$. Existiert eine Lösung? Ist sie eindeutig? Wie können wir sie berechnen bzw. annähern? Kurzum: Ist das Problem **gut gestellt**?

Diese Anwendung ist recht faszinierend: Sie ist physikalisch-anschaulich motiviert und fördert sowohl Intuition als auch Methodik. Hier gilt das Minimum-Maximum-Prinzip (Satz T4B) und daraus können wir nicht nur die Eindeutigkeit, sondern sogar die Existenz einer Lösung ableiten. Diese einfache Differenzengleichung diskretisiert partielle Differentialgleichungen (H125) und enthüllt so unerwartete Zusammenhänge, als Modell / Näherung / Anschauung für harmonische Funktionen $\Delta u = 0$ (Satz R2C) und die Wärmeleitungsgleichung $\partial_t u = \kappa \Delta u$ (Satz S3E).

Irrfahrten und Potentiale: das Dirichlet–Problem

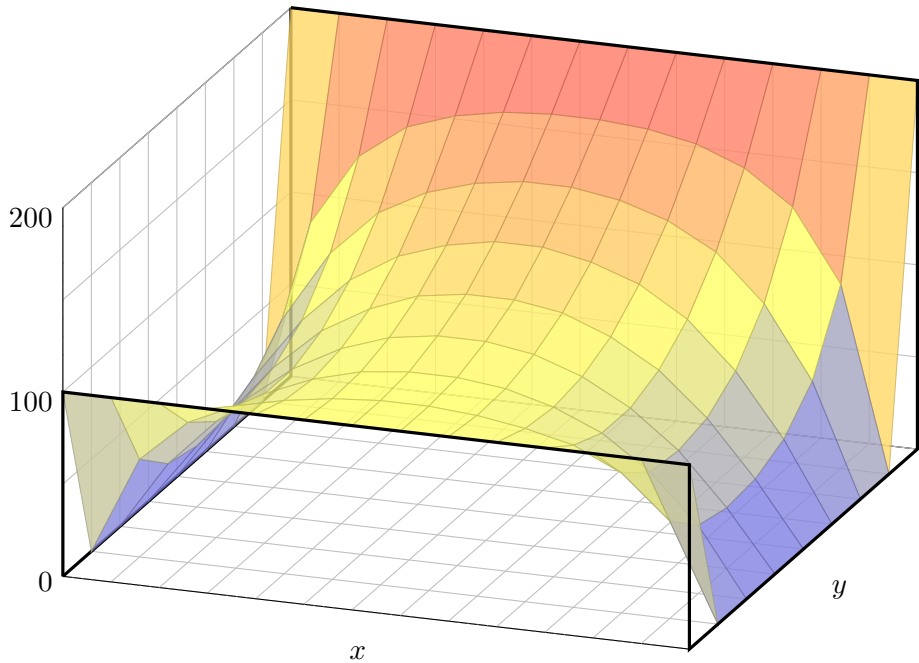
Dieses einfache Beispiel illustriert das allgemeine und überall wichtige **Dirichlet–Problem**. Die Rechnung ist trickreicher als in der vorigen Aufgabe: Dort konnten wir mit den Startwerten direkt Zeile für Zeile ausrechnen. Hier hingegen scheint die Rechnung zunächst selbstbezüglich.

	200	200	200	200	200	200	200	200	200	200	200	200	
000	100	139	158	167	172	174	174	172	167	157	139	100	000
000	061	100	125	139	147	151	151	147	139	125	100	061	000
000	043	077	102	118	127	132	132	127	118	102	077	043	000
000	035	065	088	103	113	117	117	113	103	088	065	035	000
000	033	061	081	095	104	108	108	104	095	081	061	033	000
000	037	063	081	092	099	102	102	099	092	081	063	037	000
000	053	076	088	094	098	100	100	098	094	088	076	053	000
	100	100	100	100	100	100	100	100	100	100	100	100	

Die Aufgabe führt uns zu einem **linearen Gleichungssystem** mit $7 \times 12 = 84$ Unbekannten. Für diese haben wir genau 84 Gleichungen. Das sieht vernünftig aus, bedeutet aber noch nicht, dass es genau eine Lösung gibt. Hierzu müssen wir genauer hinschauen und präzise begründen!

Meine Tabellenkalkulation *LibreOffice* weigert sich zunächst mit „Fehler 522: zirkulärer Bezug“. Mit Extras > Optionen > Calc > Iterationen rechnet sie dann iterativ obige Näherungslösung aus. Eine tabellarische Darstellung der ersten 80 Näherungsschritte finden Sie ab Seite H129.

Irrfahrten und Potentiale: das Dirichlet–Problem

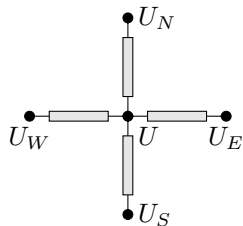


Harmonische Funktionen: Gleichgewicht!

😊 Kontext und Anwendung ändern sich, die Rechnung bleibt dieselbe!

(2) Wir betrachten Massenpunkte in $(x, y, u(x, y)) \in \mathbb{R}^3$ in Ruhelage. Jeder ist durch gleich starke Federn mit seinen Nachbarn verbunden. Somit gilt: Ruhelage = Kräftegleichgewicht \approx Mittelwerteigenschaft!

Sie können es nachrechnen! [H134] Genauer gesagt ist dies die Näherung bei geringer Krümmung.



(3) Wir betrachten die gezeigte Schaltung mit vier gleichen Widerständen. An den Nachbarpunkten liegen die Potentiale U_E, U_N, U_W, U_S an.

Ohmsches Gesetz und Kirchhoffsche Regel:

$$U = \frac{U_E + U_N + U_W + U_S}{4}$$

Ausführlich: Es gilt das Ohmsche Gesetz $I_E = (U_E - U)/R$. Die Kirchhoffsche Regel besagt hier $I_E + I_N + I_W + I_S = 0$. Einsetzen und Auflösen nach U ergibt die Mittelwerteigenschaft!

😊 Wir können Ω als Schaltung realisieren und am Rand die genannten Spannungen anlegen. Mit einem Voltmeter messen wir das Potential $u(x, y)$ im Inneren und finden obige Lösung. Physikalische Intuition suggeriert Existenz und Eindeutigkeit der Lösung, siehe Satz T4B.

Wärmeleitungsgleichung: Alles fließt!

(4) Wir betrachten Ω als ein Bauteil aus wärmeleitendem Material. Am Rand liegen die Temperaturen 0°C , 100°C , 200°C an. H125 R407
Wärme fließt von warm nach kalt proportional zur Temperaturdifferenz. Stationäre Verteilung = Fließgleichgewicht = Mittelwerteigenschaft!

😊 Unser LGS ist die diskrete Version der Potentialgleichung $\Delta u = 0$. Für $u: \mathbb{R}^2 \rightarrow \mathbb{R}$ zweimal stetig differenzierbar gilt die Taylor-Formel:

$$0 = \underbrace{u(x-1, y) - 2u(x, y) + u(x+1, y)}_{= \partial_x^2 u(x, y) + \text{h.o.t.}} + \underbrace{u(x, y-1) - 2u(x, y) + u(x, y+1)}_{= \partial_y^2 u(x, y) + \text{h.o.t.}}$$

Wir diskretisieren die Wärmeleitungsgleichung $\partial_t u = \kappa \Delta u$ für $\kappa = \frac{1}{4}$:

$$u(t+1, x, y) = \frac{u(t, x-1, y) + u(t, x+1, y) + u(t, x, y-1) + u(t, x, y+1)}{4}$$

😊 Diese Iteration nähert sich der stationären Lösung u mit $\Delta u = 0$.

Diese Beobachtung führt uns schließlich zur Methode der iterativen Näherung, die ab Seite H129 numerisch illustriert wird. Alles fließt! Dieses Näherungsverfahren beginnt mit einer (beliebigen!) Startverteilung und konvergiert recht schnell gegen die (eindeutige!) stationäre Lösung. Das ist für den Computer einfach und schnell zu rechnen. Genau das nutzt Ihre Tabellenkalkulation.

Harmonische Funktionen: Minimum-Maximum-Prinzip

Wie zuvor sei $\Omega \subseteq \mathbb{Z}^n$ ein endlicher Graph. Wir zerlegen $\Omega = \Omega^\circ \sqcup \partial\Omega$: Jeder innere Punkt $z \in \Omega^\circ$ hat alle seine $2n$ Nachbarn $z \pm e_k$ in Ω ; jeder Randpunkt $z \in \partial\Omega$ hat mindestens einen Nachbarn außerhalb von Ω . Wir nennen $u : \Omega \rightarrow \mathbb{R}$ harmonisch, wenn die Mittelwerteigenschaft gilt.

Satz T4B: Minimum-Maximum-Prinzip, Eindeutigkeit & Existenz

(1) Jede harmonische Funktion $u : \Omega \rightarrow \mathbb{R}$ nimmt Minimum und Maximum am Rand $\partial\Omega$ an: Es gilt also $\min_{\Omega} u = \min_{\partial\Omega} u$ und $\max_{\Omega} u = \max_{\partial\Omega} u$.

Für je zwei harmonische Funktionen $u, \tilde{u} : \Omega \rightarrow \mathbb{R}$ gilt demnach:

(2) Monotonie: Aus $u \leq \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u \leq \tilde{u}$ auf ganz Ω .

(3) Eindeutigkeit: Aus $u = \tilde{u}$ auf dem Rand $\partial\Omega$ folgt $u = \tilde{u}$ auf ganz Ω .

Hieraus gewinnen wir die Existenz und sogar eine Konstruktion:

(4) Existenz: Zu jeder vorgegebenen Randfunktion $v : \partial\Omega \rightarrow \mathbb{R}$ existiert genau eine harmonische Funktion $u : \Omega \rightarrow \mathbb{R}$ mit $u|_{\partial\Omega} = v$.

(5) Konstruktion: Die diskrete Wärmeleitung liefert zu v Näherungen $u_0, u_1, u_2, \dots \rightarrow u$, die gegen die eindeutige Lösung u konvergieren.

Harmonische Funktionen: Eindeutigkeit und Existenz

☺ Wir kennen das Prinzip (1-3) von harmonischen Funktionen $\Delta u = 0$ (Satz R2C) und der Wärmeleitung $\partial_t u = \kappa \Delta u$ (Satz S3E). Neu hinzu kommen Existenz (4) und Konstruktion (5).

Aufgabe: Beweisen Sie (1-4) mit linearer Algebra. **Lösung:** Die Mittelwertegenschaft (MWE) ist ein lineares Gleichungssystem: $u(z) = \frac{1}{2n} \sum_{k=1}^n u(z + e_k) + u(z - e_k)$ für alle $z \in \Omega^\circ$.

- (1) Sei $u : \Omega \rightarrow \mathbb{R}$ harmonisch. Da die Menge Ω endlich ist, existiert eine Minimalstelle $z \in \Omega$, das heißt $u(z) \leq u$. Liegt z im Inneren, so sind alle Nachbarn $z \pm e_k$ ebenfalls Minimalstellen. Es gibt einen Weg von z zu einem Randpunkt $z' \in \partial\Omega$, also ist auch dieser eine Minimalstelle.
- (2) Die Differenz $v = \tilde{u} - u$ ist harmonisch und erfüllt $v \geq 0$ auf $\partial\Omega$, nach (1) also $v \geq 0$ auf Ω . Daraus folgt sofort (3). Alternativ: $v = \tilde{u} - u$ erfüllt $v = 0$ auf $\partial\Omega$, nach (1) also $v = 0$ auf Ω .
- (4) Vorgegeben ist die Randfunktion $v : \partial\Omega \rightarrow \mathbb{R}$ als Bedingung $u|_{\partial\Omega} = v$. Wir suchen die/eine harmonische Fortsetzung $w = u|_{\Omega^\circ} : \Omega^\circ \rightarrow \mathbb{R}$ auf das Innere. Sei $N = \#\Omega^\circ$ die Anzahl innerer Punkte. Wir haben N Variablen $w(z)$ für $z \in \Omega^\circ$ und ebenso viele Gleichungen

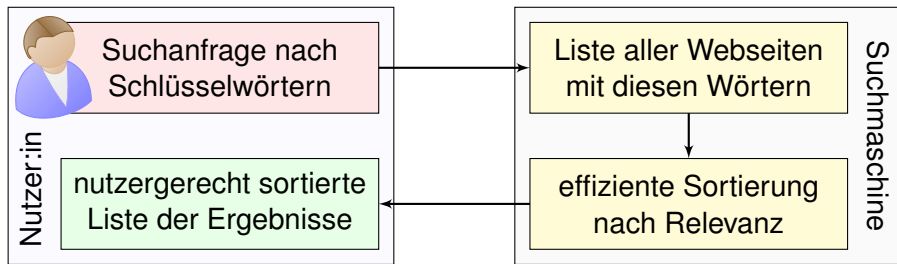
$$\Lambda_v(w)(z) := w(z) - \frac{1}{2n} \sum_{k=1}^n u(z + e_k) + u(z - e_k) \stackrel{!}{=} 0 \quad \text{wobei } u = v \sqcup w.$$

Dies definiert die affin-lineare Abbildung $\Lambda_v : \mathbb{R}^N \rightarrow \mathbb{R}^N : w \mapsto \Lambda_v(w) = \Lambda_0(w) + \Lambda_v(0)$. Die Abbildung $\Lambda_0 : \mathbb{R}^N \rightarrow \mathbb{R}^N$ ist linear. Dank (3) ist $\Lambda_0(w) = 0$ eindeutig lösbar, also Λ_0 injektiv. Dank Dimensionsformel $\dim \mathbb{R}^N = \dim \text{Kern } \Lambda_0 + \dim \text{Bild } \Lambda_0$ ist Λ_0 surjektiv, also bijektiv. Somit ist auch Λ_v bijektiv. Insbesondere existiert die ersehnte Lösung $w \in \mathbb{R}^N$ mit $\Lambda_v(w) = 0$.

- (5) Hierzu liefert die Analysis den Fixpunktsatz von Banach. T459 All unsere Argumente gelten allgemein für jeden endlichen Graphen mit Übergangswkten auf den Kanten (diskrete endliche Markov-Kette). Genau dies nutzt Googles PageRank, wie im Folgenden erläutert.

Die Mathematik hinter Google

„Wo simmer denn dran? Aha, heute krieje mer de Suchmaschin.
 Wat is en Suchmaschin? Da stelle mer uns ganz dumm. . . .“



Mathematik: Wie misst man Relevanz von Informationen?

Artificial Intelligence (AI), Machine Learning (ML), . . .

Informatik: Wie verarbeitet man enorm große Datenmengen?

Big Data, Data Mining, Data Science, . . . „Data is the new oil.“

Finanzstrategie: Wie verdient man Geld mit einem Gratisprodukt?

„If you're not paying for it, you're not the customer, you are the product.“

Die Mathematik hinter Google

Als das World Wide Web Mitte der 1990er noch klein war, da genügte es, zu einer Suchanfrage einfach alle Treffer aufzulisten. Die Liste war noch kurz, jede:r Nutzer:in konnte sie leicht selbst überblicken. Das Internet blieb jedoch nicht lange so überschaubar. . . . Das Volumen explodierte! Als Versuch einer Lösung ging 1998 die Suchmaschine Google in Betrieb und dominiert seither den Markt. Sie wird ständig weiterentwickelt. Die meisten Optimierungen hütet Google streng als Firmengeheimnis, doch das ursprüngliche Grundprinzip ist veröffentlicht und genial einfach:

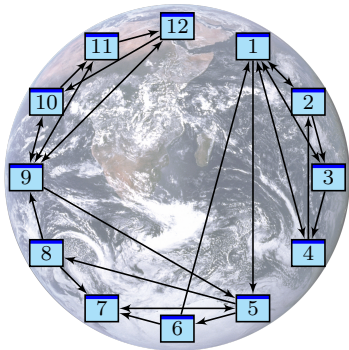
📖 Sergey Brin, Larry Page: *The anatomy of a large-scale hypertextual web search engine*. Stanford University 1998, infolab.stanford.edu/pub/papers/google.pdf

Bei vorherigen Suchmaschinen musste man endlose Trefferlisten durchforsten, bis man auf die ersten interessanten Ergebnisse stieß. Bei Google stehen sie auf wundersame Weise ganz oben. Wie ist das möglich? Die Antwort liegt (zu einem großen Teil) in folgender genial-einfachen Idee. Google misst die Popularität p_i (PageRank) jeder Seite i durch folgendes Gleichungssystem:

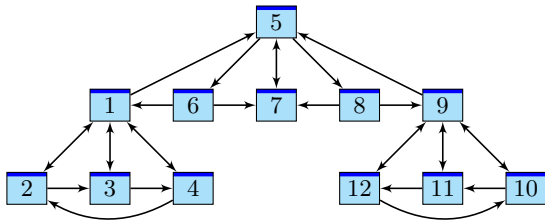
$$\text{PageRank } p_i = \frac{q}{N} + \sum_{j \rightarrow i} \frac{1-q}{\ell_j} p_j$$

Keine Angst, die Formel sieht nur auf den ersten Blick kompliziert aus. Ich werde sie anhand von Beispielen Schritt für Schritt erläutern. Wer sowas schon gesehen hat, weiß, dass es sich um eine besonders einfache Formel handelt, nämlich ein *lineares Gleichungssystem*, das keine Quadrate oder komplizierteres enthält. Schon die Formel von Pythagoras $a^2 + b^2 = c^2$ ist komplizierter.

Chaos und Struktur im Web



Miniaturbeispiel des Web als ein Graph aus Seiten $i = 1, \dots, N$ und Links $i \rightarrow j$.
Versuch einer hierarchischen Anordnung:



Eine Seite ist populär, wenn viele Seiten auf sie verweisen? Zu naiv!

Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen.

Ein zufälliger Surfer folgt von der aktuellen Seite zufällig einem der Links.

Aufgabe: Berechnen Sie die Aufenthaltswktn. Konvergieren sie gegen ein Gleichgewicht? Wie schnell? Immer dasselbe, d.h. ist es eindeutig?

😊 Im Rückblick ist die abstrakt-mathematische Idee genial einfach.

Wer diese Aufgabe bis 1998 professionell löste, ist heute Milliardär.

Chaos und Struktur im Web

Klassische Texte sind von einer Person geschrieben und linear: Ein Buch hat einen Anfang und ein Ende, typischerweise liest man es von vorne nach hinten in der Reihenfolge der Seiten. Meist gibt es zudem ein Inhaltsverzeichnis oder einen Index zum leichteren Nachschlagen. (Ja, liebe Kinder, unsere Vorfahren konnten Texte mit hunderttausend Buchstaben am Stück lesen, ohne Clicks und ohne Werbung. Man nannte das „Buch“ und speicherte es auf „Papier“. Damals!)

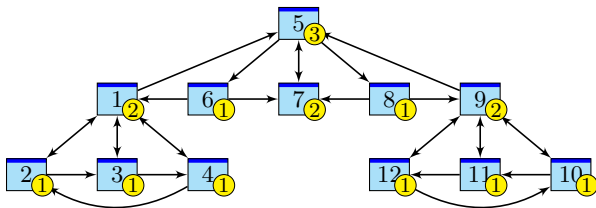
Webseiten bilden hingegen eine gänzlich andere Struktur. Niemand käme auf die Idee, das Internet von Anfang bis Ende durchzulesen: Es hat keine lineare Struktur, keine erste und keine letzte Seite, es ist zudem viel zu groß, und das meiste ist ohnehin uninteressant.

Die Webseiten verweisen gegenseitig aufeinander und bilden einen *Hypertext*. Zur Illustration betrachten wir ein Miniaturbeispiel bestehend aus 12 Webseiten. Unter den Seiten 1, 2, 3, 4 wird 1 am häufigsten zitiert. Die Seite 1 scheint daher besonders relevant oder populär. Gleiches gilt für 9, 10, 11, 12 mit 9 an der Spitze. Die Struktur von 5, 6, 7, 8 ist ähnlich mit 7 an der Spitze. Aber die Seiten 1, 7, 9, die wir schon als relevant erkannt haben, verweisen alle auf die Seite 5. Diese scheint daher populär / wichtig / zentral und für eine spätere Suche besonders relevant.

Diese Anordnung war Handarbeit. Lässt sie sich automatisieren? Nach welchen Regeln? Erster Versuch einer Bewertung: Eine Seite ist populär, wenn viele Seiten auf sie verweisen. Nachteil: Die simple Linkzählung ist zu naiv und anfällig für Manipulationen! (Linkfarmen)

Zweiter Versuch: Eine Seite ist populär, wenn viele populäre Seiten auf sie verweisen. Das klingt zunächst zirkulär, lässt sich aber in eine einfache Gleichung fassen und lösen. Ich erläutere dazu die besonders anschauliche Betrachtungsweise des zufälligen Surfers.

Irrfahrt eines zufälligen Surfers

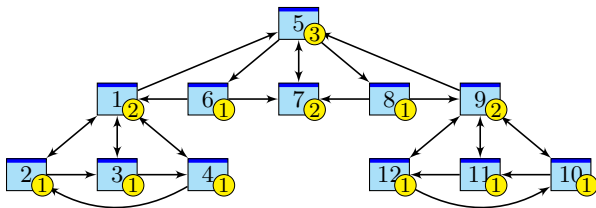


Googles Heuristik: Aufenthaltswkt \sim Popularität \sim Relevanz

Aufgabe: Berechnen Sie die Aufenthaltswktn bei Start auf Seite 7.

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$.000	.000	.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000
$t = 1$.000	.000	.000	.000	1.00	.000	.000	.000	.000	.000	.000	.000
$t = 2$.000	.000	.000	.000	.000	.333	.333	.333	.000	.000	.000	.000
$t = 3$.167	.000	.000	.000	.333	.000	.333	.000	.167	.000	.000	.000
$t = 4$.000	.042	.042	.042	.417	.111	.111	.111	.000	.042	.042	.042
$t = 5$.118	.021	.021	.021	.111	.139	.250	.139	.118	.021	.021	.021
...												
$t = 29$.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 30$.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Irrfahrt eines zufälligen Surfers



Googles Heuristik: Aufenthaltswkt \sim Popularität \sim Relevanz

Aufgabe: Berechnen Sie die Aufenthaltswktn bei Start auf Seite 1.

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$.000	.250	.250	.250	.250	.000	.000	.000	.000	.000	.000	.000
$t = 2$.375	.125	.125	.125	.000	.083	.083	.083	.000	.000	.000	.000
$t = 3$.229	.156	.156	.156	.177	.000	.083	.000	.042	.000	.000	.000
$t = 4$.234	.135	.135	.135	.151	.059	.059	.059	.000	.010	.010	.010
$t = 5$.233	.126	.126	.126	.118	.050	.109	.050	.045	.005	.005	.005
...												
$t = 69$.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059
$t = 70$.117	.059	.059	.059	.177	.059	.117	.059	.117	.059	.059	.059

Zufall und Notwendigkeit: Googles PageRank

Wir beobachten eine Diffusion: Sie konvergiert gegen eine stationäre Gleichgewichtsverteilung!
Ebenso beim Start in 1; sie konvergiert langsamer, aber schließlich zum selben Gleichgewicht!
Dank dieser Betrachtungsweise löst sich unser LGS sozusagen von allein! Verfeinertes Modell:

Sprung: Mit Wkt q startet unser Surfer neu auf irgendeiner zufälligen Seite $i \in \{1, \dots, N\}$.

Fluss: Mit Wkt $(1 - q)$ folgt er von der aktuellen Seite j zufällig irgendeinem der ℓ_j Links.

Dies führt zu folgenden Gleichungen, analog zur Wärmeleitung bzw. Potentialgleichung: T445

$$\text{Diffusion} \quad p_i(t+1) = \frac{q}{N} + \sum_{j \rightarrow i} \frac{1-q}{\ell_j} p_j(t)$$

$$\text{Gleichgewicht} \quad p_i = \frac{q}{N} + \sum_{j \rightarrow i} \frac{1-q}{\ell_j} p_j$$

Dieses verfeinerte Modell mit Teleportation lässt sich ebenso leicht berechnen. Für $q = 0.15$ entspricht es dem typischen Verhalten, sechs bis sieben Links zu folgen, bevor man neu anfängt.

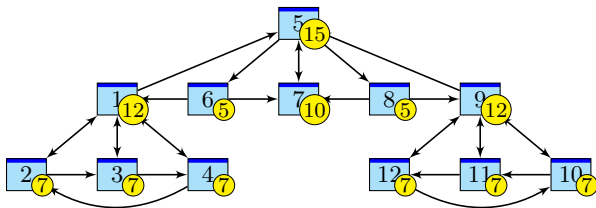
😊 Die Ergebnisse entsprechen der Nutzererwartung und sind recht robust gegen Manipulationen.

😊 Unsere obigen Beobachtungen zur Konvergenz sind nicht bloß zufällig, sondern beruhen auf mathematischen Gesetzmäßigkeiten. Diese kann man beweisen und darf sich darauf verlassen:

Der **Fixpunktsatz von Banach** garantiert bei positiver Sprunghaftigkeit $0 < q \leq 1$ Folgendes:

- (1) Es gibt genau ein Gleichgewicht p . Dieses erfüllt $p_1, \dots, p_N > 0$ und $p_1 + \dots + p_N = 1$.
- (2) Für jede Anfangsverteilung konvergiert die Diffusion gegen die Gleichgewichtsverteilung p .
- (3) Die Konvergenz ist mindestens so schnell wie die der geometrischen Folge $(1 - q)^n \searrow 0$.

Zufall und Notwendigkeit: Googles PageRank



Googles Heuristik: Aufenthaltswkt \sim Popularität \sim Relevanz

Aufgabe: Aufenthaltswkten bei Sprunghaftigkeit $q = 0.15$:

	1	2	3	4	5	6	7	8	9	10	11	12
$t = 0$	1.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
$t = 1$.013	.225	.225	.225	.225	.013	.013	.013	.013	.013	.013	.013
$t = 2$.305	.111	.111	.111	.028	.076	.087	.076	.034	.020	.020	.020
$t = 3$.186	.124	.124	.124	.158	.021	.085	.021	.071	.028	.028	.028
$t = 4$.180	.105	.105	.105	.140	.057	.075	.057	.057	.040	.040	.040
$t = 5$.171	.095	.095	.095	.126	.052	.101	.052	.087	.042	.042	.042
...												
$t = 29$.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066
$t = 30$.120	.066	.066	.066	.150	.055	.102	.055	.120	.066	.066	.066

Temporary page!

\LaTeX was unable to guess the total number of pages correctly. There was some unprocessed data that should have been added to the document. On this page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will disappear, because \LaTeX now knows how many pages to expect in the document.