

Einführung in die Angewandte Stochastik

Prof. Dr. Ansgar Steland
Institut für Statistik und Wirtschaftsmathematik
RWTH Aachen

Zufallsvorgänge

- Ausgang/Ergebnis steht nicht (deterministisch) fest.
- Ausgang hängt vom Zufall ab.

Wir brauchen:

- Mathematisches Modell
- Geeigneten Wahrscheinlichkeitskalkül.

Beispiel 1

- 50 Ladekabel, davon 1 defekt.
- Wkeit, dass der nächste Käufer das defekte Kabel greift?

Gesunder Menschenverstand diktiert:

$$p_k = P(„k-tes Ladekabel wird gegriffen“) = \frac{1}{50}$$

für $k = 1, \dots, 50$. Gleichwahrscheinlichkeit

Modellbildung

- Ergebnismenge:

$$\Omega = \{1, \dots, 50\} \quad \text{Menge}$$

- Wkeiten p_k , $k = 1, \dots, 50$. Hier:

$$p_k = \frac{1}{50}, \quad k = 1, \dots, 50.$$

- Ereignis **Teilmenge der Ergebnismenge**

$$A = \text{„Ladekabel mit gerader Serien-Nummer“} = \{2, 4, \dots, 50\}.$$

- Wahrscheinlichkeit von A : $P(A) = \frac{25}{50} = \frac{1}{2}$

Allgemeiner Ansatz: Modelliere Zufallsvorgang durch

- Ergebnismenge/Grundmenge $\Omega = \{\omega_1, \dots, \omega_N\}$
- $\omega \in \Omega$: Versuchsausgang/Ergebnis/Ausgang
- Ereignis: $A \subset \Omega$
- Ordne ω_j die Wkkeit $p_j \in [0, 1]$ zu.
- $\sum_{i=1}^N p_i = 1$.
- Ereignis A tritt ein: $\omega \in A$.

$$P(A) = \sum_{\omega_j \in A} P(\{\omega_j\}) = P(\{\omega \in \Omega : \omega \in A\})$$

Beispiel: Max, Niklas, Laura und Sahra wohnen in einer WG. Der Putzplan für die nächsten 2 Wochen wird ausgelost. Sie legen vier Zettel mit ihren Namen in eine Dose. Emma von nebenan spielt die Glücksfee und zieht zwei Zettel. Der zuerst Gezogene putzt in dieser Woche, der andere in der zweiten Woche.

Ergebnismenge Ω ?

Formale Darstellung des Ereignisses A , dass die Männer in beiden Wochen putzen müssen?

Wahrscheinlichkeiten sind Zahlen zwischen 0 und 1 (also Prozentzahlen).

- **Sicheres Ereignis:** $P(A) = 1$, z.B. $A = \Omega$.
- **Unmögliches Ereignis:** $P(A) = 0$, z.B. $A = \emptyset$.
- Zwei Ereignisse A und B heißen **disjunkt**, wenn $A \cap B = \emptyset$.
- Ereignisse A_1, A_2, \dots heißen **paarweise disjunkt**, wenn

$$A_i \cap A_j = \emptyset \quad \text{für alle } i, j \geq 1 \text{ mit } i \neq j$$

(d.h., wenn alle Paare von verschiedenen Mengen disjunkt sind).

Wartezeit T auf ersten Ausfall eines Cloud-Servers

- $\Omega = \mathbb{N}$ (Es gibt keine obere Schranke für T .)
- Wkeit für Wartezeit $k \in \mathbb{N}$: p_k

Die $\{p_i\} \subset [0, 1]$ erfüllen:

$$\sum_{i=1}^{\infty} p_i = 1$$

(da der Server auf jeden Fall irgendwann ausfällt).

Wichtige Beobachtungen:

- Gleichwahrscheinlichkeitsmodell ausgeschlossen!
- Es gibt eine unendliche Folge von paarweise disjunkten Ereignissen:

$$A_k = \{k\}, \quad k = 1, 2, \dots$$

(für die wir gerne Additivität von $P(\cdot)$ hätten).

- Rechnen mit endlichen Summen, $\sum_{k=1}^N P(A_k)$, genügt nicht, es treten auch Reihen auf:

$$\sum_{k=1}^{\infty} P(A_k) = P(A_1) + P(A_2) + \dots$$

Grundbegriffe: Beispiele/Motivation

Harmonische Reihe (divergent)

$$\sum_{i=1}^{\infty} \frac{1}{i} = +\infty$$

Konvergente Reihen sind die geometrische Reihe

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}, \quad |q| < 1,$$

die Exponentialreihe

$$e^x = \exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

sowie

$$\sum_{k=0}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

Im Fall $\Omega = \mathbb{N}$ sind nicht alle Folgen $(p_i : i \in \mathbb{N})$ möglich.

Beispiel: $p_i = c/i$ für $i \in \mathbb{N}$ und eine positive Konstante. Dann gilt:

Beispiel: $p_i = \frac{6}{\pi^2} \frac{1}{i^2}$ für $i \in \mathbb{N}$. Dann gilt:

Beispiel: $p_i = 0.8 \cdot (0.2)^i$ für $i \in \mathbb{N}_0$:

Definition (Ereignisalgebra)

Ist Ω höchstens abzählbar unendliche Menge, dann heißt jede Teilmenge $A \subset \Omega$ **Ereignis** und die Potenzmenge

$$\text{Pot}(\Omega) = \{A : A \subset \Omega\}$$

heißt **Ereignisalgebra**. Die Mengen $A = \{\omega\}$, $\omega \in \Omega$, heißen **Elementarereignisse**.

Ereignisse sind also Mengen. Alle Operationen und Regeln für Mengen übertragen sich daher auf Ereignisse.

Definition

UND-, ODER- und komplementäres Ereignis $A \subset \Omega$ und $B \subset \Omega$ Ereignisse.

UND-Ereignis

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

ODER-Ereignis

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

NICHT-Ereignis (komplementäres Ereignis):

$$\bar{A} = A^c = \{\omega \mid \omega \in \Omega \text{ und } \omega \notin A\} = \Omega \setminus A$$

Hinweis: Diese Mengenoperationen entsprechen also den logischen Grundoperationen.

Die wichtigsten Regeln:

$$\textcircled{1} A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$\textcircled{2} A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$\textcircled{3} \overline{(A \cup B)} = \bar{A} \cap \bar{B}$$

$$\textcircled{4} \overline{(A \cap B)} = \bar{A} \cup \bar{B}$$

1) und 2) sind die Distributivgesetze, 3) und 4) die Regeln von DeMorgan.

Allgemeiner:

$$\textcircled{1} A \cap (B_1 \cup \dots \cup B_n) = (A \cap B_1) \cup \dots \cup (A \cap B_n)$$

$$\textcircled{2} A \cup (B_1 \cap \dots \cap B_n) = (A \cup B_1) \cap \dots \cap (A \cup B_n)$$

$$\textcircled{3} \overline{(A_1 \cup \dots \cup A_n)} = \bar{A}_1 \cap \dots \cap \bar{A}_n$$

$$\textcircled{4} \overline{(A_1 \cap \dots \cap A_n)} = \bar{A}_1 \cup \dots \cup \bar{A}_n$$

Wahrscheinlichkeitsmaß (W-Maß), Wahrscheinlichkeitsverteilung (W-Verteilung)

Ein **W-Maß (W-Verteilung)** ist eine Abbildung, die jedem Ereignis $A \subset \Omega$ eine Zahl $P(A) \in \mathbb{R}$ zuordnet, so dass die Kolmogorov-Axiome gelten:

- 1 $0 \leq P(A) \leq 1$
- 2 $P(\Omega) = 1$
- 3 Sind A_1, A_2, \dots paarweise disjunkt, dann gilt

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots = \sum_{k=1}^{\infty} P(A_k)$$

Ein **Zufallsexperiment** ist durch Angabe von Ω und P eindeutig beschrieben, also gegeben durch das Tupel (Ω, P) .

Beispiel

Ist $\Omega = \{\omega_1, \dots, \omega_N\}$ eine endliche Menge, dann kann P durch eine Tabelle angegeben werden:

ω	ω_1	ω_2	\dots	ω_N
$P(\{\omega\})$	p_1	p_2	\dots	p_N

$P(A)$: Summe der p_i , die zu Elementen ω_i aus A gehören.

Welche Schreibweisen sind korrekt?

$$\square P(A) = \sum_{i=1}^N P(\{\omega_i\}), \quad \square P(A) = \sum_{i:\omega_i \in A} P(\{\omega_i\}), \quad \square P(A) = \sum_{i:\omega_i \in A} P(\{\omega_i\})$$

Zusammenhang zur deskriptiven Statistik

In der Praxis werden aus Datenbeständen oft relative Häufigkeiten f_j berechnet. Man darf mit den f_j so rechnen wie mit Wahrscheinlichkeiten.

- Merkmal mit Ausprägungen a_1, \dots, a_k wird erhoben, Stichprobe vom Umfang n
- Jede relative Häufigkeitsverteilung f_1, \dots, f_k definiert eine Wahrscheinlichkeitsverteilung.

Kategorie	a_1	a_2	\dots	a_k
rel. Hf.	f_1	f_2	\dots	f_k

Zufallsexperiment: Wähle zufällig mit Wkeit f_j die Kategorie j aus (bzw. ziehe ein Objekt aus dieser Kategorie).

- Bei **Rohdaten** x_1, \dots, x_n :

Empirische Verteilung mit Masse $\frac{1}{n}$ auf jedem x_i , $i = 1, \dots, n$.

x	x_1	x_2	\dots	x_n
rel. Hf.	$1/n$	$1/n$	\dots	$1/n$

Zufallsexperiment: Wähle zufällig ein x_i aus.

Elementare Rechenregeln

- 1 $P(\bar{A}) = 1 - P(A)$.
- 2 Für $A \subset B$ gilt: $P(B \setminus A) = P(B) - P(A)$.
- 3 Für *beliebige* Ereignisse A, B gilt:

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

- 4 Für *beliebige* Ereignisse A, B gilt:

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

Rechenregeln

Regel (3) liefert die Ungleichung

$$P(A \cup B) \leq P(A) + P(B)$$

und durch *vollständige Induktion* folgt für Ereignisse A_1, A_2, \dots, A_n die **Boolesche Ungleichung** ('Union Bound')

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

Beispiel

Studie: Unternehmensbefragung von IT-Abteilungen großer Unternehmen:

- 15% aller IT-Abteilungen haben steigende Kosten.
- 10% aller IT-Abteilungen haben fallende Umsätze.
- 80% haben weder steigende Kosten noch fallende Umsätze.

Frage: *Wie hoch ist der Anteil der IT-Abteilungen, die sowohl steigende Kosten als auch fallende Umsätze haben?*

Lösung:

Definiere die Ereignisse:

K = „IT hat steigende Kosten“

U = „IT hat fallende Umsätze“

W-Maß: Empirische Verteilung

Bek: $P(K) = 0.15$, $P(U) = 0.1$, $P(\overline{K \cup U}) = 0.8$.

Ges: $P(K \cap U)$.

Es gilt: $P(K \cup U) = P(K) + P(U) - P(K \cap U)$

Umstellen:

$$\begin{aligned} P(K \cap U) &= P(K) + P(U) - P(K \cup U) \\ &= 0.15 + 0.1 - 0.2 = 0.05. \end{aligned}$$

Gelte nun:

- 60% haben steigende Kosten.
- 50% haben fallende Umsätze.
- 30% haben steigende Kosten und fallende Umsätze.

Frage: Wie hoch ist der Anteil der IT-Abteilungen, die genau eines der beiden Probleme haben?

Beispiel

Bek: $P(K) = 0.6$, $P(U) = 0.5$, $P(K \cap U) = 0.3$.

Ges: $P(\underbrace{(K \cup U)}_{\text{Kosten- o. Umsatzproblem}} \setminus \underbrace{(K \cap U)}_{\text{Kosten- u. Umsatzproblem}}) = ?$

Es gilt: $K \cap U \subset K \cup U$ und somit

$$p = P((K \cup U) \setminus (K \cap U)) = P(K \cup U) - P(K \cap U)$$

wobei

$$P(K \cup U) = P(K) + P(U) - P(K \cap U)$$

Einsetzen:

$$\begin{aligned} p &= P(K) + P(U) - P(K \cap U) - P(K \cap U) \\ &= 0.6 + 0.5 - 2 \cdot 0.3 = 0.5 \end{aligned}$$

Beispiel: Probabilistische Garantien

Ein ML-Algorithmus prognostiziere Variablen x_i durch \hat{x}_i , $1 \leq i \leq n$.

Die Prognose \hat{x}_i ist ε -genau, wenn $|\hat{x}_i - x_i| \leq \varepsilon$ gilt, $\varepsilon > 0$.

ε : gewünschte Fehlerschranke.

Sei $A_i = \{|\hat{x}_i - x_i| \leq \varepsilon\}$ und $B = \bigcap_{i=1}^n A_i$ (alle Prognosen sind ε -genau).

Für zufällig gewählte Inputdaten gelte die Garantie:

$$P(A_i^c) = P(|\hat{x}_i - x_i| > \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2}\right), \quad 1 \leq i \leq n.$$

Dann gilt

$$P(B) = 1 - P\left(\bigcup_{i=1}^n A_i^c\right) \geq 1 - \sum_{i=1}^n P(A_i^c) \geq 1 - 2n \exp\left(-\frac{\varepsilon^2}{2}\right)$$

Beispiel: Probabilistische Garantien

Ziel: Bestimme für gegeben max. Fehlerwahrscheinlichkeit α ein ε derart, dass mit Wahrscheinlichkeit $1 - \alpha$ alle Prognosen ε -genau sind.

Ansatz: Um $P(B) \geq 1 - \alpha$ zu gewährleisten, bestimmen wir ε so, dass die untere Schranke den Wert $1 - \alpha$ annimmt:

$$1 - 2ne^{-\varepsilon^2/2} = 1 - \alpha \Leftrightarrow 2ne^{-\varepsilon^2/2} = \alpha$$

$$\Leftrightarrow e^{-\varepsilon^2/2} = \frac{\alpha}{2n}$$

$$\Leftrightarrow \varepsilon^2/2 = -\ln\left(\frac{\alpha}{2n}\right)$$

$$\Leftrightarrow \varepsilon = \sqrt{2\ln\left(\frac{2n}{\alpha}\right)}$$

Mit Wahrscheinlichkeit $1 - \alpha$ gilt, dass alle Prognosen $\sqrt{2\ln\left(\frac{2n}{\alpha}\right)}$ -genau sind, d.h.

$$\max_{1 \leq i \leq n} |\hat{x}_i - x_i| \leq \sqrt{2\ln\left(\frac{2n}{\alpha}\right)}.$$

Laplace-Raum

(Ω, P) heißt **Laplace-Raum**, wenn

$$\Omega = \{\omega_1, \dots, \omega_K\}$$

endlich ist und das W-Maß durch

$$p(\omega) = P(\{\omega\}) = \frac{1}{K}, \quad \omega \in \Omega,$$

gegeben ist. P heißt auch **(diskrete) Gleichverteilung auf Ω** . Dann berechnen sich Wahrscheinlichkeiten durch

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl möglicher Fälle}}$$

In Laplace-Räumen reduziert sich die Berechnung von $P(A)$ auf das Abzählen von A . Dies erfordert oft kombinatorische Fähigkeiten.

Kombinatorisches Abzählprinzip: Ist $\Omega = \Omega_1 \times \cdots \times \Omega_k$ für ein $k \in \mathbb{N}$ und $A = A_1 \times \cdots \times A_k \subset \Omega$, dann ist

$$|A| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_k|$$

Wieviele $\omega = (\omega_1, \dots, \omega_k) \in A$ gibt es?

Zähle aus, wieviel Möglichkeiten es für ω_1 gibt. \rightarrow Genau $|A_1|$.

Zähle aus, wieviel Möglichkeiten es für ω_2 gibt. \rightarrow Genau $|A_2|$.

usw.

$\rightarrow |A|$ ist das Produkt dieser Anzahlen.

Allgemeines Abzählprinzip:

Das Abzählprinzip greift auch, wenn die Mengen der Möglichkeiten für ω_i von $\omega_{i-1}, \dots, \omega_1$ abhängen: $\omega_i \in A_{\omega_1, \dots, \omega_{i-1}}$. Dann bestimmt man für jede Kombination $(\omega_1, \dots, \omega_{i-1})$ die Anzahl der nun möglichen Festlegungen von ω_i und hat

$$\sum_{(\omega_1, \dots, \omega_{i-1})} |A_{\omega_1, \dots, \omega_{i-1}}|$$

Möglichkeiten. An einem Beispiel wird klar, was gemeint ist und wie man vorgeht:

Beispiel: Wähle aus 3 Assen zufällig zwei ohne Zurücklegen aus:

1. Zug: Ziehe aus $A_1 = \Omega = \{Ass_1, Ass_2, Ass_3\}$, also 3 Möglichkeiten.

2. Zug: Es gibt drei Fälle:

Fall 1: $\omega_1 = Ass_1$: Man zieht aus $B_{\omega_1} = B_{Ass_1} = \{Ass_2, Ass_3\}$.

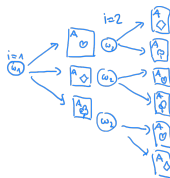
Fall 2: $\omega_1 = Ass_2$: Man zieht aus $B_{\omega_1} = B_{Ass_2} = \{Ass_1, Ass_3\}$.

Fall 3: $\omega_1 = Ass_3$: Man zieht aus $B_{\omega_1} = B_{Ass_3} = \{Ass_1, Ass_2\}$.

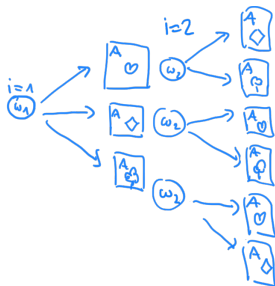
In jedem Fall gibt es 2 Möglichkeiten. Insgesamt also $3 \cdot 2 = 6$. Formel:

$$(|B_{Ass_1}| + |B_{Ass_2}| + |B_{Ass_3}|) = \sum_{\omega_1 \in A_1} |B_{\omega_1}|$$

Möglichkeiten. Übersichtliches Abzählen durch Darstellung als Baum:



Laplace-Räume und Kombinatorik



Urnenmodelle

- Ziehen in Reihenfolge mit Zurücklegen. Sei $A = \{1, \dots, N\}$.

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_1, \dots, \omega_n \in A\}$$

$$|\Omega| = N^n.$$

Als Laplace-Raum: $P(A) = \frac{|A|}{|\Omega|}$ für $A \subset \Omega$

- Ziehen in Reihenfolge ohne Zurücklegen.

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_1, \dots, \omega_n \in A, \omega_i \neq \omega_j, i \neq j\}$$

Es gilt: $|\Omega| = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1)$.

Als Laplace-Raum: $P(A) = \frac{|A|}{|\Omega|}$ für $A \subset \Omega$.

Hinweis: Man muss klären, ob das Gleichwahrscheinlichkeitsmodell zutrifft! Das folgt nicht allein schon daraus, dass eine Ergebnismenge mit der eines Urnenmodells übereinstimmt.

→ Die Urnenmodelle ohne Reihenfolge besprechen wir später.

Beispiele:

a) Max hat 5 Hemden, 2 Krawatten und 1 Anzugshose. Wieviele verschiedene Möglichkeiten hat er, sich für sein Bewerbungsgespräch anzuziehen? $\rightarrow 5 \cdot 2 \cdot 1 = 10$.

(Bei der formalen Modellierung wählt man das Urnenmodell *in Reihenfolge*, da man Hals und Beine nicht vertauschen sollte.)

b) Skat: 32 Karten, 3 Spieler, je 10 Karten, 2 im Skat. Wieviele verschiedene Skatblätter ('unsortierte') gibt es für einen Spieler? 32 Möglichkeiten für 1. Karte, 31 Möglichkeiten für 2. Karte, usw., 23 Möglichkeiten für 10. Karte.

$$\rightarrow 32 \cdot 31 \cdot \dots \cdot (32 - 10 + 1) = 2.34102 \cdot 10^{14}$$

Da ein Blatt nicht von den $10!$ möglichen Anordnungen der Karten auf der Hand abhängt, erhält man die Anzahl verschiedener Blätter durch Division durch $10!$.

Beispiele:

c) Das Passwort eines IT-Systems besteht aus einem Buchstaben (A-Z) gefolgt von einer zweistelligen Zahl und einem X oder Z, z.B. G74X.

Wie wahrscheinlich ist es, den Geheimcode zu erraten?

Wäre eine dreistellige Zahl sicherer? Wie sicher ist ein 4-stelliger PIN?

Modell:

$$\Omega = \{(\omega_1, \omega_2, \omega_3) \mid \omega_1 \in \{A, \dots, Z\}, \omega_2 \in \{0, \dots, 99\}, \omega_3 \in \{X, Z\}\}$$

Abzählen von $\Omega = \{A, \dots, Z\} \times \{0, \dots, 99\} \times \{X, Z\}$:

$$|\Omega| = 26 \cdot 100 \cdot 2 = 5200$$

Die Wahrscheinlichkeit, den Code zu erraten ist $\frac{1}{5200} = 0.0001923077$.

Dreistellige Zahl: $1/1000 = 0.001$. Vierstellige Zahl: $1/10000 = 0.0001$.

Beispiele:

d) Eine faire Münze (Kopf/Zahl gleichwahrscheinlich) wird zweimal geworfen. Ist es wahrscheinlicher, dass beide Male dasselbe erscheint oder einmal Kopf und einmal Zahl?

$\omega_1 \in \{K, Z\}$ Ergebnis des 1. Wurfs, $\omega_2 \in \{K, Z\}$ Ergebnis des 2. Wurfs.
 $\omega = (\omega_1, \omega_2)$ Ergebnis des Experiments.

Ergebnismenge: $\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{K, Z\}\}$.

Alle Elementarereignisse $\{(\omega_1, \omega_2)\}$ sind gleichwahrscheinlich.

$\rightarrow |\Omega| = 2 \cdot 2 = 2^2$.

$A = \{(K, K), (Z, Z)\}$, $B = \{(K, Z), (Z, K)\}$. Es gilt $P(A) = 2/4 = 1/2$ und $P(B) = 2/4 = 1/2$. Also ist keines der Ereignisse wahrscheinlicher als das andere, sondern beide sind gleichwahrscheinlich!

Beispiele:

e) Die Münze aus c) wird nun 6-mal geworfen. Ist das Ergebnis (Z, K, Z, K, Z, K) oder (Z, Z, Z, Z, Z, K) wahrscheinlicher?

Jedes Ergebnis ist beschrieben durch $\omega = (\omega_1, \dots, \omega_6)$ mit $\omega_1, \dots, \omega_6 \in \{K, Z\}$.

Hier ist $\Omega = \{(\omega_1, \dots, \omega_6) \mid \omega_1, \dots, \omega_6 \in \{K, Z\}\}$

→ $|\Omega| = 2^6 = 64$. Es liegt ein Laplace-Experiment vor, da alle Ausgänge gleichwahrscheinlich sind. Daher gilt:

$$P(\{(Z, K, Z, K, Z, K)\}) = \frac{1}{64}, \quad P(\{(Z, Z, Z, Z, Z, K)\}) = \frac{1}{64}$$

Achtung: Beim Ausgang (Z, Z, Z, Z, Z, K) kommt zunächst 5-mal Zahl. Viele Menschen glauben, dass es nun sehr wahrscheinlich ist, dass beim nächsten Mal Kopf kommt. Dies ist allerdings ein Trugschluss: Das Ergebnis (Z, Z, Z, Z, Z, Z) hat aber ebenfalls die Wahrscheinlichkeit $1/64$. Die geworfene Münze landet (aufgrund der Gesetze der Physik) jedesmal zufällig auf der einen oder der anderen Seite, und zwar mit Wkeit $1/2$. Dies gilt ebenso für andere Glücksspiele,

Zufällige Speicherbelegung:

k Objekte sollen in einem Speicher mit n Speicherplätzen abgelegt werden. Als Speicherplatz wird jeweils eine zufällige Zahl zwischen 1 und n gewählt. A_{nk} bezeichne das Ereignis, dass eine Kollision auftritt.

Bestimme $P(A_{nk})$ und gebe eine untere Schranke an!

Das Geburtstagsproblem:

Sie sind auf einer Party mit n Gästen. Wie wahrscheinlich ist es, dass mindestens zwei Gäste am selben Tag Geburtstag haben? Ist man im Vorteil, wenn man darauf wettet?

Die Lösung ist:

$$p_n = 1 - \frac{365 \cdot \dots \cdot (365 - n + 1)}{365^n}$$

Zahlenbeispiele (gerundet):

$$p_5 = 0.027, p_{10} = 0.117, p_{20} = 0.411, p_{23} = 0.507.$$

In Beispielen wie diesen tritt der Ausdruck $\frac{n(n-1)\cdots(n-k+1)}{n^k}$ auf (setze: $n = N, k = n$).

Es gilt die Abschätzung:

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} \leq \exp\left(-\frac{(k-1)k}{2n}\right)$$

Alltag: Die Chancen stehen $75 : 25 = 0.75 : 0.25$

Chancen

Die **Chance** (odds) eines Ereignisses A mit Eintrittswahrscheinlichkeit $p = P(A)$ ist definiert durch

$$o = o(A) = \frac{p}{1-p}.$$

Es ist $o(A)$ -mal wahrscheinlicher, dass A eintritt, als dass A nicht eintritt. Durch Logarithmieren erhält man die **logarithmierten Chancen** (engl.: *log-odds*):

$$\log(o) = \log(p/(1-p)) = \log(p) - \log(1-p).$$

Symmetrieeigenschaft:

$$\log o(\bar{A}) = \log((1-p)/p) = -\log o(A)$$

Nullpunkt: $\log o(A) = 0$, falls $P(A) = 1/2$.

Odds-Ratio

$$r = \frac{o(A)}{o(B)} = \frac{P(A)/(1 - P(A))}{P(B)/(1 - P(B))}$$

Beispiel:

Gruppe 1 mit Risikofaktor, $A =$ "Person aus 1 erkrankt": $o(A) = \frac{0.3}{0.7}$.

Gruppe 2 ohne Risikofaktor: $B =$ "Person aus 2 erkrankt": $o(B) = \frac{0.1}{0.9}$.

Odds-Ratio: $r = \frac{0.3 \cdot 0.9}{0.7 \cdot 0.1} = \frac{0.27}{0.07} \approx 3.856$.