

# Stochastik für Informatiker



Dr. rer. nat. Johannes Riesterer



# Diskrete Wahrscheinlichkeitsverteilung

## Beispiel: Hash Kollision

Beim Hashing werden zufällig  $k \leq n$  Daten auf  $n$  Speicherplätze verteilt. Bezeichnen wir mit  $A_{k,n}$  die Möglichkeiten der Mehrfachbelegungen von Feldern, so ist das Komplementäre Ereignis  $A_{k,n}^c = \text{Perm}_k^n(\Omega, o.W.)$ , wobei  $\Omega$  die Menge der Verfügbaren Speicherplätze Darstellt.

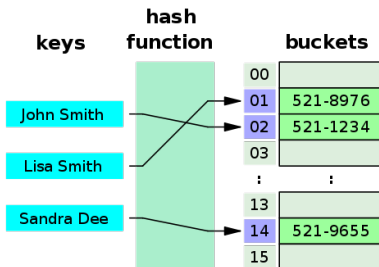


Figure: Quelle: Wikipedia

## Beispiel: Hash Kollision

$$\begin{aligned}P(A_{k,n}^c) &= \frac{\#Var_k^n(\Omega, o.W.)}{\#Var_k^n(\Omega, m.W.)} = \frac{n_k}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\&= \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right) \leq \exp\left(\sum_{i=0}^{k-1} \left(-\frac{i}{n}\right)\right) \\&\quad (\ln(1-x) \leq -x \text{ für } x < 1) \\&= \exp\left(-\frac{(k-1)k}{2n}\right)\end{aligned}$$

## Beispiel: Hash Kollision (Geburtstags-Paradoxon)

Für  $n = 365$  und  $k = 23$  ist damit  $P(A_{k,n}) > \frac{1}{2}$ . Die Wahrscheinlichkeit, dass bei einer Gruppe von mehr als 23 Leuten zwei Leute am gleichen Tag Geburtstag haben, ist also größer als  $\frac{1}{2}$ .

## $\sigma$ -Algebra

Es sei  $\Omega$  eine Menge und  $\mathcal{A} \subset \mathcal{P}(\Omega)$  ein System von Teilmengen.  $\mathcal{A}$  heißt  $\sigma$ -Algebra (Ereignis-Algebra) falls gilt:

$$(i) \Omega \in \mathcal{A}$$

$$(ii) A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

$$(iii) A_i \in \mathcal{A} \Rightarrow \bigcup_i A_i \in \mathcal{A}$$

$$(A^c = \Omega - A)$$

## Interpretation

Die Grundmenge  $\Omega$  ist ein Ereignis. Das nicht-Eintreffen eines Ereignisses ist ein Ereignis. Die Vereinigung von Ereignissen ist ein Ereignis.

## Axiome von Kolmogorov

Ein Wahrscheinlichkeitsraum ist ein Tripel  $(\Omega, \mathcal{A}, P)$  bestehend aus der Grundmenge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathcal{A} \subset \mathcal{P}(\Omega)$  und einer Abbildung  $P : \mathcal{A} \rightarrow [0, 1]$

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

$$(ii) P\left(\bigcup_i A_i\right) = \sum_i P(A_i), \text{ mit } A_i \cap A_j = \emptyset \text{ f\"ur } i \neq j$$

Die Elemente von  $\Omega$  werden elementare Ereignisse und die von  $\mathcal{A}$  Ereignisse genannt. Mengen mit  $P(M) = 0$  werden Nullmengen genannt.

## Interpretation

Die Grundmenge ist ein sicheres Ereignis. Die Wahrscheinlichkeit von überschneidungsfreien Ereignissen addiert sich.

## Diskrete Wahrscheinlichkeitsverteilung

Ein diskreter Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$ , bei dem die Grundmenge  $\Omega$  abzählbar ist.

## Beispiel: Laplace Wahrscheinlichkeit

$\Omega$  endlich  $\mathcal{A} = \mathcal{P}(\Omega)$ , und  $P(A) = \frac{\#A}{\#\Omega}$ .



## Lemma

Sei  $(\Omega, \mathcal{A}, P)$  ein diskreter Wahrscheinlichkeitsraum. Dann ist für  $A \subset \mathcal{A}$

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

$$P(A^c) = 1 - P(A)$$

$$P(\emptyset) = 0$$

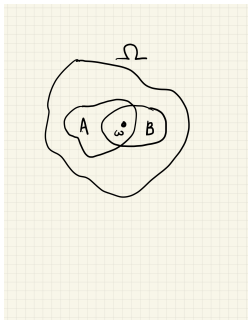
# Diskrete Wahrscheinlichkeitsverteilungen

## Herleitung der bedingten Wahrscheinlichkeit

$$\tilde{\Omega} := B$$

$$\tilde{\mathcal{A}} := \{C \cap B \mid C \in \mathcal{A}\}$$

$$\tilde{P} = \frac{P}{P(B)}$$



# Diskrete Wahrscheinlichkeitsverteilungen

## Bedingte Wahrscheinlichkeit

Für  $A, B \subset \mathcal{P}(\Omega)$  und  $P(B) > 0$  heißt

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

die bedingte Wahrscheinlichkeit (von  $A$  unter  $B$ ).

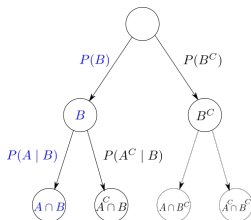


Figure: Quelle: Wikipedia

## Satz der totalen Wahrscheinlichkeit

Für eine Zerlegung  $\Omega = \bigcup_{j=1}^n B_j$ , mit  $B_i \cap B_k = \emptyset$  für  $i \neq k$

$$P(A) = \sum_{j=1}^n P(A \mid B_j) \cdot P(B_j)$$

## Satz von Bayes

Für  $A, B \subset \mathcal{P}(\Omega)$  mit  $P(B) > 0$  gilt

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

## Beweis

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

## Stochastische Unabhängigkeit

Zwei Ereignisse  $A, B$  heißen stochastisch Unabhängig, falls

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

gilt. Gleichbedeutend damit ist  $P(A|B) = P(A)$  und  $P(B|A) = P(B)$ .



## Naiver Bayes'scher Spam Filter

Gegeben ist eine E-Mail  $E$ . Wir möchten anhand des Vorkommens bestimmter Wörter  $A_1, \dots, A_n$  in der Mail entscheiden, ob es sich um eine erwünschte Mail  $H$  oder eine unerwünschte Mail  $S$  (Ham or Spam) handelt. (Typische Wörter wären zum Beispiel "reich", "casino", "Vergrößerung" ...)





## Naiver Bayes'scher Spam Filter

Aus einer Datenbank kann man das Vorkommen dieser Wörter in Spam und Ham Mails zählen und damit empirisch die Wahrscheinlichkeiten  $P(A_i|S)$  und  $P(A_i|H)$  des Vorkommens dieser Wörter in Spam und Ham Mails ermitteln. Wir gehen davon aus, dass es sich bei der Mail prinzipiell mit Wahrscheinlichkeit  $P(E = S) = P(E = H) = \frac{1}{2}$  um eine erwünschte Mail  $H$  oder eine unerwünschte Mail  $S$  handeln kann.

## Naiver Bayes'scher Spam Filter

Wir machen zudem die (naive) Annahme, dass das Vorkommen der Wörter stochastisch unabhängig ist, also

$$\begin{aligned}P(A_1 \cap \dots \cap A_n | S) &= P(A_1 | S) \cdots P(A_n | S) \\P(A_1 \cap \dots \cap A_n | H) &= P(A_1 | H) \cdots P(A_n | H)\end{aligned}$$

gilt.

## Naiver Bayes'scher Spam Filter

Mit der Formel von Bayes und der totalen Wahrscheinlichkeit können wir somit berechnen

$$\begin{aligned} P(E = S|A_1 \cap \dots \cap A_n) &= \frac{P(A_1 \cap \dots \cap A_n|S) \cdot P(S)}{P(A_1 \cap \dots \cap A_n)} \\ &= \frac{P(A_1|S) \cdots P(A_n|S) \cdot P(S)}{P(A_1 \cap \dots \cap A_n|H) + P(A_1 \cap \dots \cap A_n|S)} \\ &= \frac{P(A_1|S) \cdots P(A_n|S) \cdot P(S)}{P(A_1|H) \cdots P(A_n|H) + P(A_1|S) \cdots P(A_n|S)} \end{aligned}$$

Bemerkung:  $P(E = H|A_1 \cap \dots \cap A_n) = 1 - P(E = S|A_1 \cap \dots \cap A_n)$