Informationstheorie, FS 2017 Dr. Luis Haug
ETH Zürich, Dept. Informatik Nico Gorbach

Merkblatt Diskrete Wahrscheinlichkeitstheorie

Diskrete Wahrscheinlichkeitsräume

Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar (Ω, P) bestehend aus einer endlichen oder höchstens abzählbar unendlichen Menge Ω und einer *Verteilung* P, d.h. einer Abbildung $P : \Omega \to [0,1]$, so dass

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

gilt. Die Elemente $\omega \in \Omega$ bezeichnen wir als *Elementarereignisse*, und wir intepretieren $P(\omega)$ als die "Wahrscheinlichkeit, dass ω eintritt".

Ereignisse und ihre Wahrscheinlichkeiten

Die Elemente der Potenzmenge 2^Ω von Ω (d.h. die Teilmengen $A \subseteq \Omega$) bezeichnen wir als *Ereignisse*. Die Verteilung $P : \Omega \to [0,1]$ induziert durch die Vorschrift

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

eine Abbildung $P : 2^\Omega \to [0,1]$ (die wir der Einfachheit halber wieder mit P bezeichnen und ebenfalls Verteilung nennen). Wir interpretieren $P(A)$ als die "Wahrscheinlichkeit, dass das Ereignis A eintritt". Es gelten folgende Regeln:

a) $P(\emptyset) = 0, P(\Omega) = 1$,

b) $A \subset B \Longrightarrow P(A) \leq P(B)$,

c) $P(A \cup B) - P(A \cap B) = P(A) + P(B)$ für alle Ereignisse A, B,

d) $P(\bigcup_{i \in \mathbb{N}} A_i) \leq \sum_{i \in \mathbb{N}} P(A_i)$.

Unabhängigkeit von Ereignissen

Zwei Ereignisse $A, B \subseteq \Omega$ heissen *unabhängig*, falls $P(A \cap B) = P(A)P(B)$ gilt. Allgemeiner heisst eine Familie von Ereignissen $(A_i)_{i \in I}$ unabhängig, wenn für jede endliche Teilmenge $J \subseteq I$ gilt: $P\big(\bigcap_{j \in J} A_j\big) = \prod_{j \in J} P(A_j)$.

Bedingte Wahrscheinlichkeiten

Seien $A, B \subseteq \Omega$ Ereignisse. Die *bedingte Wahrscheinlichkeit von A gegeben B* ist definiert als

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

falls $P(B) > 0$ (andernfalls ist $P(A|B)$ undefiniert).

Zufallsvariablen

Gegeben seien ein diskreter Wahrscheinlichkeitsraum (Ω, P) sowie eine Menge $\mathcal{X}$. Eine *Zufallsvariable* mit Werten in $\mathcal{X}$ ist eine Abbildung $X : \Omega \to \mathcal{X}$.

Durch Zufallsvariablen definierte Ereignisse

Für eine Teilmenge $A \subset \mathcal{X}$ beschreiben wir das Ereignis $X^{-1}(A) \subset \Omega = \{\omega \in \Omega \mid X(\omega) \in A\}$ mit "X nimmt einen Wert in A an". Für solche durch Zufallsvariablen definierten Ereignisse verwenden wir Schreibweisen der Form

$$\{X \in A\} = X^{-1}(A), \quad \{X = x\} = X^{-1}(\{x\}), \quad \text{etc.}$$

und für die entsprechenden Wahrscheinlichkeiten schreiben wir

$$P(X \in A), \quad P(X = x), \quad \text{etc.}$$

Verteilung von Zufallsvariablen

Das Bild von X ist die Teilmenge $X(\Omega) = \{X(\omega) \in \mathcal{X} \mid \omega \in \Omega\}$ von $\mathcal{X}$. Gemeinsam mit der durch

$$P_X(x) = P(X = x)$$

gegebenen Abbildung $P_X : X(\Omega) \to [0, 1]$ ist $(X(\Omega), P_X)$ selbst ein diskreter Wahrscheinlichkeitsraum. Wir bezeichnen P_X als die *Verteilung* der Zufallsvariablen X.

Gemeinsame Verteilung

Für Zufallsvariablen $X : \Omega \to \mathcal{X}$ und $Y : \Omega \to \mathcal{Y}$ können wir das Paar (X, Y) als Zufallsvariable mit Werten in $\mathcal{X} \times \mathcal{Y}$ auffassen. Die zugehörige Verteilung bezeichen wir als *gemeinsame Verteilung* von X und Y und bezeichnen sie mit $P_{XY} : X(\Omega) \times Y(\Omega) \to [0, 1]$. Die gemeinsame Verteilung bestimmt die Verteilungen P_X und P_Y eindeutig durch

$$P_X(x) = \sum_{y \in \mathcal{Y}(\Omega)} P_{XY}(x, y), \quad P_Y(y) = \sum_{x \in \mathcal{X}(\Omega)} P_{XY}(x, y).$$

Unabhängigkeit von Zufallsvariablen

Zwei Zufallsvariablen X und Y heissen *unabhängig*, falls

$$P_{XY}(x, y) = P_X(x)P_Y(y)$$

für alle $x \in X(\Omega)$ und alle $y \in Y(\Omega)$ gilt.

Bedingte Verteilungen von Zufallsvariablen

Für eine Zufallsvariable $X : \Omega \to \mathcal{X}$ und ein Ereignis $A \subseteq \Omega$ mit $P(A) > 0$ definieren wir die *bedingte Verteilung von X gegeben A* als die durch

$$P_{X|A}(x) = P(X = x|A)$$

gegebene Abbildung $P_{X|A} : X(\Omega) \to [0, 1]$.

Für Zufallsvariablen $X : \Omega \to \mathcal{X}$ und $Y : \Omega \to \mathcal{Y}$ definieren wir die *bedingte Verteilung von X gegeben Y* als die durch

$$P_{X|Y}(x, y) = P(X = x|Y = y)$$

gegebene Abbildung $P_{X|Y} : X(\Omega) \times Y(\Omega) \to [0, 1]$. Man beachte, dass $P_{X|Y}$ nicht notwendigerweise eine Verteilung im obigen Sinne ist (weil im Allgemeinen $\sum_{x,y} P_{X|Y}(x, y) \neq 1$ gilt); allerdings ist für jedes $y \in Y(\Omega)$ mit $P(Y = y) > 0$ die Abbildung $P_{X|Y}(\cdot, y) : X(\Omega) \to [0, 1]$ eine Verteilung auf $X(\Omega)$ im üblichen Sinn.

Erwartungswerte und Varianzen

Sei $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable für die die Reihe $\sum_{x\in X(\Omega)} |x|P_X(x)$ konvergiert. Wir sagen dann, dass X einen *Erwartungswert* $E(X)$ besitzt und definieren diesen durch

$$E(X) = \sum_{x\in X(\Omega)} xP_X(x).$$

Falls sowohl X als auch X^2 Erwartungswerte besitzen, sagen wir, dass die *Varianz* $V(X)$ von X existiert und definieren diese durch

$$V(X) = \sum_{x\in X(\Omega)} (x - E(X))^2 P_X(x).$$

(Allgemeiner kann man Erwartungswerte und Varianzen für Zufallsvariablen mit Werten in $\mathbb{R}^n$ definieren).

Es gelten folgende Rechenregeln für Erwartungswerte und Varianzen:

$$E(\alpha X) = \alpha E(X), \quad E(X + Y) = E(X) + E(Y)$$

für $\alpha \in \mathbb{R}$ und für Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$ (Linearität des Erwartungswertes), und

$$V(\alpha X) = \alpha^2 V(X)$$

für $\alpha \in \mathbb{R}$, sowie

$$V(X + Y) = V(X) + V(Y)$$

falls $X, Y : \Omega \to \mathbb{R}$ unahbängig sind.

Markov-Ungleichung

Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable deren Erwartungswert existiert. Dann gilt

$$P\big(|X| \geq \varepsilon\big) \leq \frac{E(|X|)}{\varepsilon}$$

für jedes $\varepsilon > 0$.

Chebyshev-Ungleichung

Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Varianz $V(X)$. Dann gilt

$$P\big(|X - E(X)| \geq \varepsilon\big) \leq \frac{V(X)}{\varepsilon^2}.$$

Konvexe Funktionen

Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heisst *konvex*, falls für alle $a, b \in \mathbb{R}$ und jedes $t \in [0,1]$ die Ungleichung

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b)$$

gilt, und *strikt konvex*, falls Gleichheit nur für $t \in \{0,1\}$ gilt. f heisst *(strikt) konkav* falls $-f$ (strikt) konvex ist.

Jensen-Ungleichung

Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, deren Erwartungswert existiert, und sei $f : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion. Dann gilt

$$E(f(X)) \geq f(E(X)).$$

Falls f *strikt* konvex ist, gilt Gleichheit genau dann, wenn $P(X = E(X)) = 1$ ist.
Für konkave Funktionen $g : \Omega \to \mathbb{R}$ gilt entsprechend

$$E(g(X)) \leq g(E(X)).$$

Falls g *strikt* konkax ist, gilt Gleichheit genau dann, wenn $P(X = E(X)) = 1$ ist.

Konvergenz von Zufallsvariablen

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und sei X eine weitere Zufallsvariable.

- Wir sagen, dass X_n *fast sicher gegen X konvergiert*, falls

  $$P\big(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1$$

  gilt.

- Wir sagen, dass X_n *in Wahrscheinlichkeit gegen X konvergiert*, falls für jedes $\varepsilon > 0$

  $$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

  gilt.

Schwaches Gesetz der Grossen Zahlen

Seien $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die paarweise unabhängig sind und die alle den gleichen Erwartungswert E und die gleiche Varianz V besitzen. Sei $S_n = X_1 + \cdots + X_n$. Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P\big(|\tfrac{1}{n}S_n - E| > \varepsilon) = 0,$$

d.h. der Mittelwert $\frac{1}{n}S_n$ konvergiert in Wahrscheinlichkeit gegen E.