Stochastik

1 Wahrscheinlichkeitsraum

1.1 Definition

Ein **diskreter Wahrscheinlichkeitsraum** (Ω, Pr) besteht aus einer endlichen oder abzählbar unendlichen Menge Ω und einer Abbildung $Pr : \Omega \to [0,1] \subseteq \mathbb{R}$ mit $\sum\limits_{\omega \in \Omega} Pr(\omega) = 1$.

Die Elemente von Ω nennt man **Ergebnisse**, die Teilmengen von Ω **Ereignisse**.

Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$: $Pr(A) := \sum\limits_{\omega \in A} Pr(\omega)$.

$\to$ Erweiterung zu einem **Wahrscheinlichkeitsmaß** auf dem **Meßraum** $Pr(\Omega)$.

1.2 Begriffe und Sätze

- **Laplace-Raum:** $Pr(\omega) = \frac{1}{|\Omega|} \; \forall \, \omega \in \Omega$.
- **Bedingte Wahrscheinlichkeit:** $Pr(A|B) := \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}$ für $Pr(B) > 0$ (Satz von Bayes)

  Korollar: Sind $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse mit $Pr(A_1 \cap \ldots \cap A_n) > 0$, dann gilt:
  $Pr(A_1 \cap \ldots \cap A_n) = Pr(A_1) \cdot Pr(A_2|A_1) \cdots Pr(A_n|a_{n-1} \cap \ldots \cap A_1)$
- Zwei Ereignisse $A, B \subseteq \Omega$ heißen **voneinander unabhängig** gdw.
  $Pr(A \cap B) = Pr(A) \cdot Pr(B)$.
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

2 Zufallsvariable

2.1 Definition

Eine **Zufallsvariable** über einem WR (Ω, Pr) ist eine Abbildung $X : \Omega \to W_X \subseteq \mathbb{R}$. W_X heißt **Wertemenge** von X.

Schreibweise: $Pr(X = r) := Pr(\{\omega \in \Omega \mid X(\omega) = r\})$ $\quad Pr(X \leq r) := Pr(\{\omega \in \Omega \mid X(\omega) \leq r\})$

Zu einer Zufallsvariable X definiert man die

- **Dichtefunktion** $f_X : W_X \to [0,1]$ durch $f_X(r) = Pr(X = r)$
- **Verteilungsfunktion** $F_X : W_X \to [0,1]$ durch $F_X(r) = Pr(X \leq r)$

Aus einer numerischen Zufallsvariable X konstruiert man einen WR (W_X, Pr_X). Pr_X ist durch Dichte oder Verteilung von X bestimmt.

2.2 Mehrere Zufallsvariablen

Sind $X_1, \ldots, X_n$ Zufallsvariablen über WR $(\Omega_1, Pr_1), \ldots, (\Omega_n, Pr_n)$, dann ist $X := X_1 + \ldots + X_n$ eine Zufallsvariable über $(\Omega_1, Pr_1) \times \ldots \times (\Omega_n, Pr_n)$.

Die Zufallsvariablen $X_1, \ldots, X_n$ heißen **unabhängig**, wenn für jede Teilmenge $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$, $I \neq \emptyset$ gilt:

$$Pr(X_{i_1} = r_1, \ldots, X_{i_k} = r_k) = Pr(X_{i_1} = r_1) \cdot \ldots \cdot Pr(x_{i_k} = r_k)$$

2.3 Erwartungswert

Als **Erwartungswert** der Zufallsvariablen X definiert man

$$E(X) = \mu := \sum_{r \in W_X} r\, f_X(r),$$

falls diese Summe konvergiert.

Rechenregeln:

$E(aX + b) = aE(X) + b \quad a, b \in \mathbb{R}$ (Linearität)
$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$

2.4 Varianz und Standardabweichung

Als **Varianz** einer Zufallsgröße X bezeichnet man die Größe

$$Var(X) = \sigma^2 := E((X-\mu)^2) = \sum_{r \in W_X} (r-\mu)^2\, f_X(r)$$

$\sigma = \sqrt{Var(X)}$ heißt **Standardabweichung**.

Rechenregeln

$Var(X) = E(X^2) - [E(X)]^2 \qquad Var(aX + b) = a^2\, Var(X)$
Für unabhängige Zufallsvariable X_i gilt: $Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$

3 Wichtige Verteilungen

3.1 Bernoulli-Verteilung

Parameter: p

$W_X = \{0, 1\} \qquad f_X(r) = \begin{cases} p & r = 1 \\ 1-p & r = 0 \end{cases} \qquad E(X) = p \qquad Var(X) = p(1-p)$

3.2 Binomialverteilung

Parameter: p, n

$W_X = \{0, \ldots, n\} \qquad f_X(r) = \binom{n}{r} p^r (1-p)^{n-r} =: b(r, n, p) \qquad E(X) = np \qquad Var(X) = np(1-p)$

Entstehung: Sind $X_1, \ldots, X_n$ Bernoulli-verteilt mit dem Parameter p und unabhängig, dann ist $X := X_1 + \ldots + X_n$ binomialverteilt mit den Parametern n und p.

3.3 Geometrische Verteilung

Parameter: p

$W_X = \mathbb{N}^+ \qquad f_X(r) = (1-p)p^{r-1} \qquad E(X) = \frac{1}{p} \qquad Var(X) = \frac{1-p}{p^2}$

Deutung: Wahrscheinlichkeit, dass nach r-maliger Wiederholung eines Bernoulli-verteilten Experimentes mit Erfolgswahrscheinlichkeit p ein Mißerfolg eintritt.

3.4 Poisson-Verteilung

Parameter: λ

$W_X = \mathbb{N}_0 \qquad f_x(r) = \frac{\lambda^r}{r!}e^{-\lambda} = \lim_{n\to\infty} b(r, n, \frac{\lambda}{n}) \qquad E(X) = \lambda \qquad Var(X) = \lambda$

Deutung: $f_X(r)$ ist die Wahrscheinlichkeit, dass in einem Bediensystem in einer bestimmten Zeiteinheit r Forderungen eintreffen (λ durchschnittliche Anzahl Forderungen in dieser Zeiteinheit).

4 Gesetze

4.1 Ungleichung von Markov

Es sei X eine Zufallsvariable, die nur nichtnegative Werte annimmt. Dann gilt:

$$(\forall t \in \mathbb{R}^+)\ Pr(X \geq t) \leq \frac{E(X)}{t} \qquad \text{insbesondere aber} \quad Pr(X \geq tE(X)) \leq \frac{1}{t}$$

Beweis: $E(X) = \sum_{r\in W_X} rf_X(r) \geq \sum_{\substack{r\in W_X\\ r\geq t}} rf_X(r) \geq \sum_{\substack{r\in W_X\\ r\geq t}} tf_X(r) = t \sum_{\substack{r\in W_X\\ r\geq t}} f_X(r) = tPr(X\geq t)$

4.2 Ungleichung von Tschebyscheff

Sei X eine Zufallsvariable. Dann gilt:

$$(\forall t \in \mathbb{R}^+)\ Pr(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

Beweis: $Pr(|X-E(X)|\geq t) = Pr((X-E(X))^2 \geq t^2)$ Die Zufallsvariable $(X-E(X)^2)$ hat als Erwartungswert gerade $E((X - E(X))^2) = Var(X)$. Aus der Ungleichung von Markov folgt die Behauptung.

4.3 Bernoullis Gesetz der großen Zahlen

Gegeben sei eine Zufallsvariable X sowie Zahlen $\varepsilon, \delta \in \mathbb{R}^+$. Dann gilt für $n \geq \frac{Var(X)}{\varepsilon\delta^2}$: Sind $X_1, \ldots X_n$ unabhängige Zufallsvariablen mit $f_{X_i} = f_X,\ 1 \leq i \leq n$ und setzt man
$Z := \frac{1}{n}(X_1 + \ldots + X_n)$, dann gilt:

$$Pr(|Z - E(X)| \geq \delta) \leq \varepsilon$$

5 Stochastische Prozesse mit diskreter Zeit

5.1 Definition stochastischer Prozess

Eine Familie $\{X_t | t \in T\}$ von Zufallsvariablen nennt man einen **stochastischen Prozess**.

Für $T = \mathbb{N}$ spricht man von **diskreter Zeit**, für $T = \mathbb{R}_0^+$ von **kontinuierlicher Zeit**. Bei diskreter Zeit denkt man sich einen stochastischen Prozess als Folge $X_0, X_1, \ldots$ von Zufallsvariablen.

5.2 Definition Markov-Kette

Eine (endliche) **Markov-Kette** mit diskreter Zeit über der Zustandsmenge $S = \{0, 1, \ldots, n-1\}$ besteht aus einer Folge $X_0, X_1, \ldots$ von Zufallsvariablen mit Wertemenge S und einer Startverteilung $q_0 \in (\mathbb{R}_0^+)^n$ mit $\sum_{i=1}^{n} q_{0i} = 1$. Für jede Indexmenge $I \subseteq \{0, 1, \ldots, t-1)$ und beliebige Zustände $i, j, s_k, k \in I$ muß gelten: $Pr(X_{t+1} = j \mid X_t = i) = Pr(X_{t+1} = j : \mid X_t = i, X_k = s_k\ \forall\ k \in I)$

(D. h., frühere Zustände haben keinen Einfluß auf die Übergangswahrscheinlichkeiten).

Sind die Werte $p_{ij} := Pr(X_{t+1} = j \mid X_t = i)$ unabhängig von t, so nennt man die Markov-Kette **zeithomogen**. Solche Ketten beschreibt man durch ihre **Übergangsmatrix** $P := (p_{ij})_{i \le i,j \le n}$.

Der Vektor q_t habe als i-te Komponente die Wahrscheinlichkeit $Pr(X_t = i)$. Dann gilt:

$$q_{t+1} = q_t \cdot P \qquad \rightarrow \qquad q_{t+k} = q_t \cdot P^k$$

Übergangswahrscheinlichkeit nach t Schritten: $p_{ij}^{(t)}$ ist die (i, j)-te Komponente von P^t

5.3 Periodizität

Die **Periode** eines Zustandes j ist die größte Zahl $\xi \in \mathbb{N}_0$ mit

$$\{t \in \mathbb{N}_0 \mid p_{jj}^{(t)} > 0\} \subseteq \xi\mathbb{N}_0 = \{i\xi \mid i \in \mathbb{N}_0\}$$

Ein Zustand mit Periode 1 heißt **aperiodisch**. Eine Markov-Kette heißt **aperiodisch**, wenn alle ihre Zustände aperiodisch sind.

Hinreichende Bedingungen, dass ein Zustand j aperiodisch ist:

- $p_{jj} > 0$
- $(\exists t_0)\ (\forall t \ge t_0)\ p_{jj}^{(t)} > 0$
- Es gibt Pfade mit positiver WS von j nach j, deren Längen teilerfremd sind

5.4 Eigenschaften

- Eine Wahrscheinlichkeitsverteilung einer Markov-Kette $\pi \in (\mathbb{R}_0^+)^n$ heißt **stationäre Verteilung**, wenn gilt: $\pi P = \pi$.
- Eine Markov-Kette heißt **irreduzibel**, wenn gilt:

  $$(\forall i, j \in S)\ (\exists k \in \mathbb{N}_0)\ p_{ij}^{(k)} > 0$$

  (d. h., von jedem Zustand kann jeder andere Zustand in endlich vielen Schritten erreicht werden)
- eine irreduzible aperiodische Markov-Kette heißt **ergodisch**.

  Fundamentalsatz für ergodische Markov-Ketten:

  Jede (endliche zeithomogene) ergodische Markov-Kette besitzt eine eindeutige stationäre Verteilung π und es gilt:

  $$\lim_{t \to \infty} q_0 P^t = \pi \quad \text{(unabhängig von } q_0\text{)}$$