Stochastik für die Informatik, Vorlesung 12

Inhalt

- Kovarianz und Korrelation
- Verteilungsfunktionen und Dichten
- Rechnen mit Dichten

Lernziele

- Kovarianz und Korrelation und ihre Eigenschaften kennen
- Eigenschaften von Verteilungsfunktionen und Dichten kennen
- Mit Dichten rechnen können

Vorkenntnisse

Stoff der bisherigen Vorlesungen; Integral- und Differentialrechnung

Bemerkung zur Chebyshev-Ungleichung

(Satz 5.5) Sei X eine Zufallsvariable deren Varianz existiert. Dann gilt für jedes $a > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

Diese (leicht stärkere) Aussage gilt auch, mit dem analogen Beweis: (Satz 5.5*) Sei X eine Zufallsvariable deren Varianz existiert. Dann gilt für jedes $a > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}.$$

Erinnerung: Kovarianz

(Def. 5.4). Seien X und Y zwei Zufallsvariablen auf $(\Omega, \mathbb{P})$. Die Kovarianz von X und Y ist definiert als

$$\boxed{\operatorname{cov}(X, Y) := \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]}$$

falls dieser Erwartungswert existiert.
(Def. 5.5) Seien X und Y Zufallsvariablen, deren Kovarianz existiert.

- Ist $\operatorname{cov}(X, Y) > 0$, so heißen X und Y positiv korreliert,
- Ist $\operatorname{cov}(X, Y) < 0$, so heißen X und Y negativ korreliert,
- Ist $\operatorname{cov}(X, Y) = 0$, so heißen X und Y unkorreliert.

(Satz 5.7) Seien X und Y Zufallsvariablen, deren Kovarianz existiert. Es gilt

$$\boxed{\operatorname{cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X]\mathbb{E}[Y].}$$

Varianz der Summe von zwei Zufallsvariablen

(Satz 5.9) Seien X und Y Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann gilt

$$\boxed{\mathbb{V}(X+Y) = \mathbb{V}(X) + 2\mathrm{cov}(X,Y) + \mathbb{V}(Y).}$$

- (Beweis)
- (Spezialfall)
- (Beispiel 5.16: Zufallsgraph)

Korrelation

(Def. 5.5) Seien X, Y zwei Zufallsvariablen mit positiver Varianz auf $(\Omega, \mathbb{P})$. Der Korrelationskoeffizient von X und Y ist definiert als

$$\boxed{\varrho(X,Y) := \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathbb{V}(X)\cdot\mathbb{V}(Y)}}}$$

- Das Vorzeichen der Korrelation stimmt mit dem Vorzeichen der Kovarianz überein.
- $-1 \leq \mathrm{corr}(X,Y) \leq 1$,
- Maß für die Stärke des linearen Zusammenhang zwischen X und Y.
- $|\mathrm{corr}(X,Y)| = 1 \Leftrightarrow \exists a \in \mathbb{R} \setminus \{0\}$ so dass $Y = aX$, und Vorzeichen von $a =$Vorzeichen von $\mathrm{corr}(X,Y)$.

(Beispiel 5.17: Zufallsgraph)

Kapitel 6: Zufallsvariablen mit Dichte

Erinnerung an Kapitel 3: (Diskrete) Zufallsvariablen und Verteilungsfunktion

- Zufallsvariable: $X : \Omega \to \mathbb{R}$, Wertebereich $X(\Omega) \subseteq \mathbb{R}$.
- Diskrete Zufallsvariablen: $X(\Omega)$ endlich oder abzählbar $(= \mathbb{N})$
- Für diskrete Zufallsvariablen enthält die Verteilung

$$p_X(k) = \mathbb{P}(X = k), k \in X(\Omega)$$

im Prinzip alle relevanten Informationen.

- Nicht alle Zufallsvariablen sind diskret. Beispiel: Zeit, Länge, Gewicht...
- Beispiel Wartezeit: $X(\Omega) = [0, \infty)$ oder $[0, T]$ für ein $T \in \mathbb{R}$.
- $\mathbb{P}(X = x)$ kann in solchen Fällen gleich 0 sein.

Verteilungsfunktion

Erinnerung (Def. 3.6): Die Verteilungsfunktion einer Zufallsvariablen X ist definiert durch $F_X : \mathbb{R} \to [0,1]$

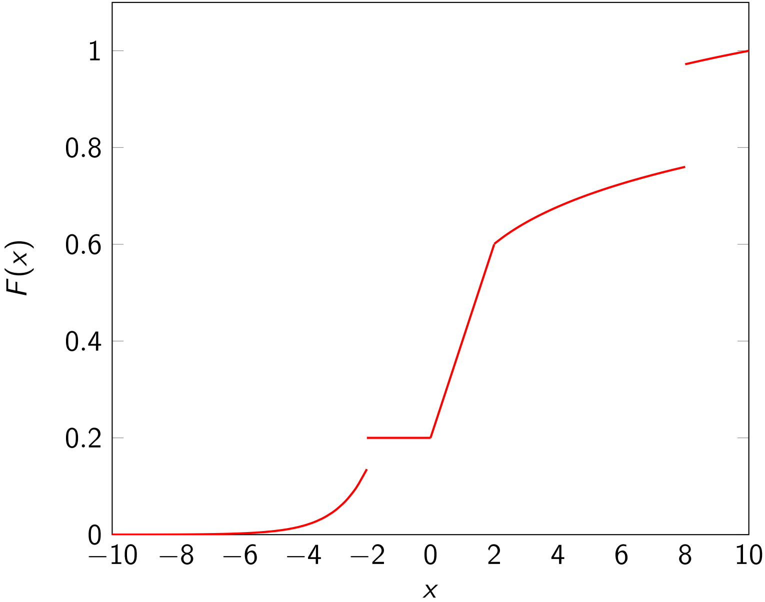
$$F_X(x) = \mathbb{P}(X \leq x).$$

Eigenschaften der Verteilungsfunktion:

- $F_X : \mathbb{R} \to [0,1]$ ist monoton nichtfallend, d.h. $x \geq y$ impliziert $F_X(x) \geq F_X(y)$, und rechtsstetig
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$

Die Verteilungsfunktion einer Zufallsvariablen kann auch für nicht-diskrete Zufallsvariablen definiert werden.
Jede Funktion mit diesen drei Eigenschaften ist die Verteilungsfunktion einer Zufallsvariablen

Verteilungsfunktion



Zufallsvariablen mit Dichten

(Def. 6.1) Sei X eine Zufallsvariable mit Verteilungsfunktion F_X. Dann besitzt X bzw. F_X eine Dichte, falls eine Funktion $f_X : \mathbb{R} \to \mathbb{R}$ existiert, so dass für alle $x \in \mathbb{R}$ gilt

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Die Funktion f_X heißt dann Dichte von X bzw. von F_X.

(Satz 6.2) Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ ist genau dann die Dichte einer Zufallsvariablen, wenn sie die folgenden beiden Eigenschaften hat:

(i) $f(t) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(t)dt = 1$

- (Beweis)
- (Bem. isolierte Punkte)
- (Bem. Hauptsatz der Differential- und Integralrechnung)

Zufallsvariablen mit Dichten

(Def. 6.1) Sei X eine Zufallsvariable mit Verteilungsfunktion F_X. Dann besitzt X bzw. F_X eine Dichte, falls eine Funktion $f_X : \mathbb{R} \to \mathbb{R}$ existiert, so dass für alle $x \in \mathbb{R}$ gilt

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Die Funktion f_X heißt dann Dichte von X bzw. von F_X.

(Satz 6.2) Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ ist genau dann die Dichte einer Zufallsvariablen, wenn sie die folgenden beiden Eigenschaften hat:

(i) $f(t) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(t)dt = 1$

- (Beweis)
- (Bem. isolierte Punkte)
- (Bem. Hauptsatz der Differential- und Integralrechnung)

Beispiel 6.3: Gleichverteilung

Seien $a < b$, $a, b \in \mathbb{R}$. Die Funktion $F : \mathbb{R} \to \mathbb{R}$,

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

erfüllt die Eigenschaften einer Verteilungsfunktion.

Eine Zufallsvariable X mit dieser Verteilungsfunktion heißt (stetig) gleichverteilt oder uniform verteilt auf $[a, b]$.

- (Skizze)
- (Dichte)
- (Spezialfall: $[0, 1]$)

Dichten: Beispiele

Beispiel 6.1: Welche der folgenden Funktionen sind Dichten von Zufallsvariablen?

$$f(t) = \begin{cases} t^3 & 0 \leq t \leq 1 \\ 0 & \text{sonst} \end{cases}$$

$$f(t) = \begin{cases} \frac{1}{4}t^3 & 0 \leq t \leq 2 \\ 0 & \text{sonst} \end{cases}$$

$$f(t) = \begin{cases} -4t^3 & -1 \leq t \leq 0 \\ 0 & \text{sonst} \end{cases}$$

$$f(t) = \begin{cases} \frac{4}{15}t^3 & -1 \leq t \leq 2 \\ 0 & \text{sonst} \end{cases}$$

Rechnen mit Dichten

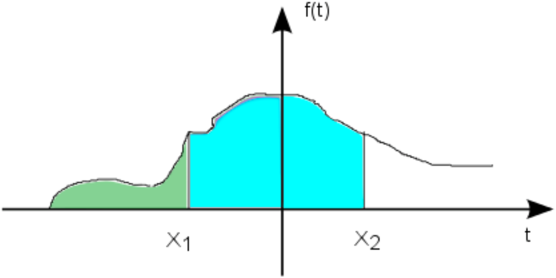
(Satz 6.3) Sei X eine Zufallsvariable mit Dichte f_X. Seien $a, b \in \mathbb{R}$ mit $a < b$.

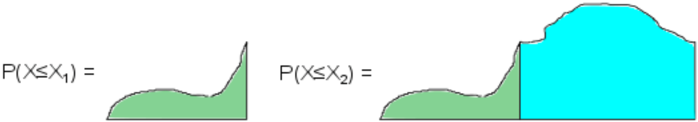
- $\mathbb{P}(X \leq b) = \mathbb{P}(X < b) = \int_{-\infty}^{b} f_X(t)dt$
- $\mathbb{P}(X \geq a) = \mathbb{P}(X > a) = \int_{a}^{\infty} f_X(t)dt$
- $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b[) = \int_{a}^{b} f_X(t)dt$

- Beispiel
- (Beweis)
- Graphische Darstellung
- $\mathbb{P}(X = a) = 0$

Rechnen mit Dichten

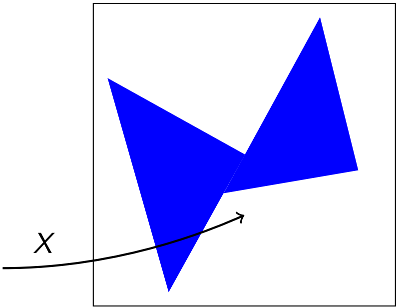
$\mathbb{P}(x_1 \leq X \leq x_2)$ ist die Fläche unter der Dichte:





$\mathbb{P}(X = a) = 0$

Die Wahrscheinlichkeit, einen einzelnen Punkt zu treffen, ist 0



Erwartungswert und Varianz

(Def. 6.2) Sei X eine Zufallsvariable mit Dichte f_X. Der Erwartungswert von X ist definiert als

$$\boxed{\mathbb{E}[X] := \int_{-\infty}^{\infty} t \cdot f_X(t) dt,}$$

falls das Integral exisitiert. Für eine Funktion $g : \mathbb{R} \to \mathbb{R}$ sei

$$\boxed{\mathbb{E}[g(X)] := \int_{-\infty}^{\infty} g(t) \cdot f_X(t) dt,}$$

falls das Integral existiert. Die Varianz von X ist definiert als

$$\boxed{\mathbb{V}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2],}$$

falls der Erwartungswert existiert.

- (Bem. Satz 6.4 $\mathbb{E}[g(X)]$)
- (Beispiele)

Erwartungswert und Varianz

(Satz 6.4) Sei X eine Zufallsvariable mit Dichte f_X. Für die Varianz (falls sie existiert) gilt

$$\begin{aligned}\mathbb{V}(X) &= \int_{-\infty}^{\infty}(t-\mathbb{E}[X])^2 \cdot f_X(t)dt \\ &= \int_{-\infty}^{\infty} t^2 \cdot f_X(t)dt - (\int_{-\infty}^{\infty} t \cdot f_X(t)dt)^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

- (Beweis)
- (Beispiel 6.4)