Stochastik für die Informatik, Vorlesung 11

Inhalt

- Gemeinsame Verteilung
- Unabhängigkeit von Zufallsvariablen
- Kovarianz und Korrelation

Lernziele

- Gemeinsame Verteilungen und bedingte Verteilungen von Zufallsvariablen berechnen können
- Zufallsvariablen auf Unabhängigkeit überprüfen können, und einige Implikationen kennen
- Kovarianz und Korrelation kennen und in Beispielen berechnen können
- Eigenschaften der Kovarianz kennen

Vorkenntnisse

Stoff der bisherigen Vorlesungen

Gemeinsame Verteilung

(Def. 4.1) Seien X und Y zwei diskrete Zufallsvariablen, welche auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ definiert sind. Die gemeinsame Verteilung von X und Y ist gegeben durch die Kollektion der Wahrscheinlichkeiten

$$\boxed{\mathbb{P}(X = x, Y = y) := \mathbb{P}(\{X = x\} \cap \{Y = y\}), \quad x \in X(\Omega), y \in Y(\Omega).}$$

- Beispiele: Größe und Gewicht eines Menschen; Anzahl Bestellungen und Bearbeitungszeit in einem Logistikzentrum; Anzahl und Höhe von Schäden bei einer Versicherung...
- Analog wird die gemeinsame Verteilung von mehr als zwei Zufallsvariablen definiert
- (Beispiel: Erdős-Rényi-Zufallsgraph mit $k = 3$)

Gemeinsame Verteilung

(Def. 4.1) Seien X und Y zwei diskrete Zufallsvariablen, welche auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ definiert sind. Die gemeinsame Verteilung von X und Y ist gegeben durch die Kollektion der Wahrscheinlichkeiten

$$\boxed{\mathbb{P}(X = x, Y = y) := \mathbb{P}(\{X = x\} \cap \{Y = y\}), \quad x \in X(\Omega), y \in Y(\Omega).}$$

- Beispiele: Größe und Gewicht eines Menschen; Anzahl Bestellungen und Bearbeitungszeit in einem Logistikzentrum; Anzahl und Höhe von Schäden bei einer Versicherung...
- Analog wird die gemeinsame Verteilung von mehr als zwei Zufallsvariablen definiert
- (Beispiel: Erdős-Rényi-Zufallsgraph mit $k = 3$)

Gemeinsame Verteilung

(Beispiel 4.1: Urnenmodell): Urne mit 4 Kugeln, numeriert von 1 bis 4. Wir ziehen zwei Kugeln mit Zurücklegen. Sei X die Summe der beiden Zahlen, und Y das Minimum der beiden gezogenen Zahlen.

Tabelle der gemeinsamen Verteilung

$Y \backslash X$	2	3	4	5	6	7	8
1	1/16	1/8	1/8	1/8	0	0	0
2	0	0	1/16	1/8	1/8	0	0
3	0	0	0	0	1/16	1/8	0
4	0	0	0	0	0	0	1/16

Gemeinsame Verteilung: Randverteilung

(Def. 4.2) Sei die gemeinsame Verteilung von X und Y gegeben. Die Randverteilungen von X bzw. von Y sind die Verteilungen der einzelnen Zufallsvariablen X bzw. Y, also

$$\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y)$$

$$\mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y).$$

(Beispiel 4.2 Urnenmodell)

Beispiel 4.2: Tabelle

$Y \backslash X$	2	3	4	5	6	7	8	$\mathbb{P}(Y = \cdot)$
1	$1/16$	$1/8$	$1/8$	$1/8$	0	0	0	$7/16$
2	0	0	$1/16$	$1/8$	$1/8$	0	0	$5/16$
3	0	0	0	0	$1/16$	$1/8$	0	$3/16$
4	0	0	0	0	0	0	$1/16$	$1/16$
$\mathbb{P}(X = \cdot)$	$1/16$	$1/8$	$3/16$	$1/4$	$3/16$	$1/8$	$1/16$	1

Bedingte Verteilung

(Def. 4.3) Seien X, Y zwei Zufallsvariablen welche auf demselben Wahrscheinlichkeitsraum definiert sind. Die bedingte Verteilung von X gegeben Y ist definiert als die Kollektion der bedingten Wahrscheinlichkeiten

$$\boxed{\mathbb{P}(X = x \mid Y = y), \qquad x \in X(\Omega), y \in Y(\Omega).}$$

- Die Rollen von X und Y sind wichtig!
- Nach Definition der bedingten Wahrscheinlichkeit und der gemeinsamen Verteilung gilt

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

- Beispiel in den Hausaufgaben

Erinnerung: Unabhängigkeit von Zufallsvariablen

(Def. 4.4) Zwei Zufallsvariablen X und Y definiert auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ heißen unabhängig, falls für alle $x \in X(\Omega), y \in Y(\Omega)$ die Ereignisse $\{X = x\}$ und $\{Y = y\}$ unabhängig sind.

(Satz 4.1). Zwei Zufallsvariablen X und Y sind genau dann unabhängig, wenn für alle $x \in X(\Omega), y \in Y(\Omega)$ gilt

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

- (Letztes Mal: Beispiel 4.4 Urnenmodell)
- (Beispiel: 4.5 Gemeinsame Verteilung)

Erinnerung: Unabhängigkeit von Zufallsvariablen

Beispiel: X, Y mit gemeinsamer Verteilung aus Tabelle:

$Y \backslash X$	0	1	$\mathbb{P}(Y = y)$
0	1/4	1/4	1/2
1	1/12	0	1/12
2	1/6	1/4	5/12
$\mathbb{P}(X = x)$	1/2	1/2	1

$$\mathbb{P}(X = 0, Y = 0) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 0).$$

Bedeutet das, dass X und Y unabhängig sind?

Nein, denn (z.B).

$$\mathbb{P}(X = 0, Y = 1) = \frac{1}{12} \neq \frac{1}{2} \cdot \frac{1}{12} = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1).$$

Somit gilt die Gleichung $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ nicht für alle Paare x und y.

Erinnerung: Unabhängigkeit von Zufallsvariablen

Beispiel: X, Y mit gemeinsamer Verteilung aus Tabelle:

$Y \backslash X$	0	1	$\mathbb{P}(Y=y)$
0	$1/4$	$1/4$	$1/2$
1	$1/12$	0	$1/12$
2	$1/6$	$1/4$	$5/12$
$\mathbb{P}(X=x)$	$1/2$	$1/2$	1

$$\mathbb{P}(X=0, Y=0) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X=0) \cdot \mathbb{P}(Y=0).$$

Bedeutet das, dass X und Y unabhängig sind?

Nein, denn (z.B).

$$\mathbb{P}(X=0, Y=1) = \frac{1}{12} \neq \frac{1}{2} \cdot \frac{1}{12} = \mathbb{P}(X=0) \cdot \mathbb{P}(Y=1).$$

Somit gilt die Gleichung $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ nicht für alle Paare x und y.

Erinnerung: Unabhängigkeit von Zufallsvariablen

Beispiel: X, Y mit gemeinsamer Verteilung aus Tabelle:

$Y \backslash X$	0	1	$\mathbb{P}(Y=y)$
0	$1/4$	$1/4$	$1/2$
1	$1/12$	0	$1/12$
2	$1/6$	$1/4$	$5/12$
$\mathbb{P}(X=x)$	$1/2$	$1/2$	1

$$\mathbb{P}(X=0, Y=0) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X=0) \cdot \mathbb{P}(Y=0).$$

Bedeutet das, dass X und Y unabhängig sind?

Nein, denn (z.B).

$$\mathbb{P}(X=0, Y=1) = \frac{1}{12} \neq \frac{1}{2} \cdot \frac{1}{12} = \mathbb{P}(X=0) \cdot \mathbb{P}(Y=1).$$

Somit gilt die Gleichung $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ nicht für alle Paare x und y.

Verteilung der Summe unabhängiger Zufallsvariablen

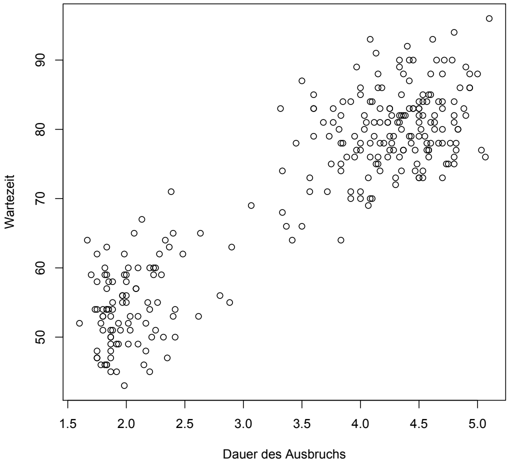
(Satz 4.4: Faltungsformel) Seien X und Y zwei unabhängige diskrete Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann hat die Zufallsvariable $X + Y$ die Verteilung

$$\mathbb{P}(X + Y = k) = \sum_{m \in X(\Omega)} \mathbb{P}(X = m) \cdot \mathbb{P}(Y = k - m).$$

- Beachte dass $(X + Y)(\Omega) = \{m + l : m \in X(\Omega), l \in Y(\Omega)\}$
- Falls X und Y nicht unabhängig sind ist diese Formel falsch
- (Beweis)
- (Beispiel 4.7: Poisson-Verteilung)

Linearer Zusammenhang zwische zwei Zufallsvariablen

Beispiel: Geysir-Ausbruch: Dauer und Wartezeit (R: Datensatz "faithful")



Offensichtlich: Je größer X, desto größer Y, und umgekehrt

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.

- Werte zwischen $-\infty$ und ∞
- $\operatorname{cov}(X, Y) > 0$ ist genau dann der Fall, wenn im Mittel beide Abweichungen vom jeweiligen Mittelwert gleiches Vorzeichen haben.

Kovarianz

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

- Ist $\operatorname{cov}(X, Y) > 0$, so heißen X und Y positiv korreliert: Im Mittel sind X und Y entweder beide größer als ihr jeweiliger Erwartungswert, oder beide kleiner.
- Ist $\operatorname{cov}(X, Y) < 0$, so heißen X und Y negativ korreliert: Nimmt X einen besonders großen Wert an, so nimmt Y eher einen besonders kleinen Wert an.
- 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

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

- Für Berechnungen oft nützlicher als die Definition!

Kovarianz

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

- Ist $\mathrm{cov}(X, Y) > 0$, so heißen X und Y positiv korreliert: Im Mittel sind X und Y entweder beide größer als ihr jeweiliger Erwartungswert, oder beide kleiner.
- Ist $\mathrm{cov}(X, Y) < 0$, so heißen X und Y negativ korreliert: Nimmt X einen besonders großen Wert an, so nimmt Y eher einen besonders kleinen Wert an.
- Ist $\mathrm{cov}(X, Y) = 0$, so heißen X und Y unkorreliert.

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

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

- Für Berechnungen oft nützlicher als die Definition!

Kovarianz: Beispiel

Beispiel 5.14: Gemeinsame Verteilung von X und Y als Tabelle:

$Y \setminus X$	0	1	2	Σ
0	$1/16$	$1/4$	$1/8$	$7/16$
1	$3/16$	$1/4$	$1/8$	$9/16$
Σ	$1/4$	$1/2$	$1/4$	1

Bestimmung von $\operatorname{cov}(X, Y)$: Mit Hilfe von Satz 5.7.
Man findet

$$\mathbb{E}[X] = 1, \quad \mathbb{E}[Y] = \frac{9}{16}, \quad \mathbb{E}[XY] = \frac{1}{2}$$

und somit

$$\operatorname{cov}(X, Y) = \frac{-1}{16}$$

Eigenschaften der Kovarianz

(Satz 5.8) Seien X, Y, Z Zufallsvariablen auf $(\Omega, \mathbb{P})$ mit positiver Varianz, und seien $a, b \in \mathbb{R}$.

(a) $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$,

(b) $\operatorname{cov}(X, X) = \operatorname{var}(X)$,

(c) $\operatorname{cov}(aX + bY, Z) = a \cdot \operatorname{cov}(X, Z) + b \cdot \operatorname{cov}(Y, Z)$,

(d) Falls X und Y unabhägig sind, so gilt $\operatorname{cov}(X, Y) = 0$.

In Worten:

(a) Die Kovarianz ist symmetrisch

(b) Die Kovarianz einer Zufallsvariablen mit sich selbst ist die Varianz

(c) Die Kovarianz ist bilinear

(d) Unabhängig impliziert unkorreliert

- Wichtig: Die Umkehrung von (d) gilt im Allgemeinen nicht!
- (Beweis von (d))

Eigenschaften der Kovarianz

(Satz 5.8) Seien X, Y, Z Zufallsvariablen auf $(\Omega, \mathbb{P})$ mit positiver Varianz, und seien $a, b \in \mathbb{R}$.

(a) $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$,

(b) $\mathrm{cov}(X, X) = \mathrm{var}(X)$,

(c) $\mathrm{cov}(aX + bY, Z) = a \cdot \mathrm{cov}(X, Z) + b \cdot \mathrm{cov}(Y, Z)$,

(d) Falls X und Y unabhägig sind, so gilt $\mathrm{cov}(X, Y) = 0$.

In Worten:

(a) Die Kovarianz ist symmetrisch

(b) Die Kovarianz einer Zufallsvariablen mit sich selbst ist die Varianz

(c) Die Kovarianz ist bilinear

(d) Unabhängig impliziert unkorreliert

- Wichtig: Die Umkehrung von (d) gilt im Allgemeinen nicht!
- (Beweis von (d))

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)

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

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

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

(Beispiel 5.17: Zufallsgraph)