Stochastik für die Informatik, Vorlesung 10

Inhalt

- Unabhängigkeit von Zufallsvariablen
- Varianz
- Chebyshev-Ungleichung

Lernziele

- Unabhängigkeit von Zufallsvariablen kennen
- Entscheiden können, ob zwei Zufallsvariablen unabhängig sind
- Den Begriff der Varianz kennen, und Varianzen in Beispielen berechnen können
- Die Chebyshev-Ungleichung und das Gesetz der großen Zahlen kennen

Vorkenntnisse

Stoff der bisherigen Vorlesungen,

Erinnerung: Varianz

(Def. 5.2) Sei X eine Zufallsvariable. Die Varianz von X ist definiert als

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

(falls dieser Erwartungswert existiert, in diesem Fall sagen wir, dass die Varianz existiert).

- Mittlere quadratische Abweichung der Zufallsvariable vom Erwartungswert
- Andere Notation: $\mathbb{V}(X) = \mathrm{var}(X)$

(Satz 5.3) Sei X eine Zufallsvariable, deren Varianz existiert. Es gilt

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

(Def. 5.3) Sei X eine Zufallsvariable, deren Varianz existiert. Dann heißt $\sigma(X) := \sqrt{\mathbb{V}(X)}$ Standardabweichung von X.

Einschub: Unabhängigkeit von Zufallsvariablen

Erinnerung: Unabhängigkeit von Ereignissen

- A, B unabhängig $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- A, B unabhängig $\Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$
- Unabhängigkeit: Wahrscheinlichkeit für das eine Ereignis wird nicht durch das Eintreten (oder Nicht-Eintreten) des anderen beeinflusst

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

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

- (Beispiel 4.4 Urnenmodell)

Unabhängigkeit: Weitere Beispiele

(Satz 4.2) Seien $Y_1, ..., Y_n$ unabhängige Bernoulli-verteilte Zufallsvariablen mit Erfolgswahrscheinlichkeit p. Dann ist $X := \sum_{i=1}^{n} Y_i$ binomialverteilt mit Parametern n und p.

(Satz 4.3) Seien X und Y unabhängige (diskrete) Zufallsvariablen, und seien $f : X(\Omega) \to \mathbb{R}$ und $g : Y(\Omega) \to \mathbb{R}$ zwei Funktionen. Dann sind die Zufallsvariablen $f(X)$ und $g(Y)$ unabhängig.

Eigenschaften der Varianz

(Satz 5.4) Seien X und Y zwei Zufallsvariablen, deren Varianzen existieren.

(a) Für $a, b \in \mathbb{R}$ gilt

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$$

(b) Falls X und Y unabhängig sind, so gilt

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

(c) Falls X konstant ist, so gilt $\mathbb{V}(X) = 0$.

- Wichtig: Falls X und Y nicht unabhängig sind, so gilt im allgemeinen nicht $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.
- (Beispiel 5.11: Varianz der Binomialverteilung)
- (Beispiel 5.12: Varianz im Glücksspiel)
- (Beispiel: Varianz im coupon collector problem)

Tabelle: Erwartungswert und Varianz wichtiger Verteilungen

Verteilung	Parameter	Erwartungswert	Varianz
Bernoulli	p	p	$p(1-p)$
Binomial	n, p	np	$np(1-p)$
Geometrisch	p	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	λ	λ	λ

Zipf: Der Erwartungswert exisistiert falls $a > 2$ ist, die Varianz existiert falls $a > 3$ ist.

Chebyshev-Ungleichung

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

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

- Varianz als Maß für die Abweichung vom Erwartungswert
- Kleine Varianz: Große Abweichungen vom Erwartungswert sind unwahrscheinlich. Umformung: Mit $\sigma = \sqrt{\mathbb{V}(X)}$ lautet die Chebyshev-Ungleichung auch

$$\mathbb{P}(|X - \mathbb{E}[X]| > K \cdot \sigma) \leq \frac{1}{K^2}$$

  Die Wahrscheinlichkeit, um mehr als das K-Fache der Standardabweichung vom Erwartungswert abzuweichen, ist höchstens $1/K^2$.
- (Beweis)
- (Beispiel 5.13: Coupon collector problem)

Chebyshev-Ungleichung

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

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

- Varianz als Maß für die Abweichung vom Erwartungswert
- Kleine Varianz: Große Abweichungen vom Erwartungswert sind unwahrscheinlich. Umformung: Mit $\sigma = \sqrt{\mathbb{V}(X)}$ lautet die Chebyshev-Ungleichung auch

  $$\mathbb{P}(|X - \mathbb{E}[X]| > K \cdot \sigma) \leq \frac{1}{K^2}$$

  Die Wahrscheinlichkeit, um mehr als das K-Fache der Standardabweichung vom Erwartungswert abzuweichen, ist höchstens $1/K^2$.
- (Beweis)
- (Beispiel 5.13: Coupon collector problem)

Markov-Ungleichung

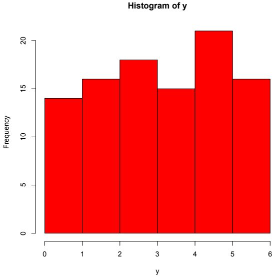
(Satz 5.5) Sei X eine Zufallsvariable, und sei $f : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende Funktion. Falls $\mathbb{E}[f(X)]$ existiert, so gilt für jedes $a \in \mathbb{R}$

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[f(X)]}{f(a)}.$$

- Chebyshev-Ungleichung ist ein Spezialfall der Markov-Ungleichung
- Beweis der Markov-Ungleichung analog zum Beweis der Chebyshev-Ungleichung

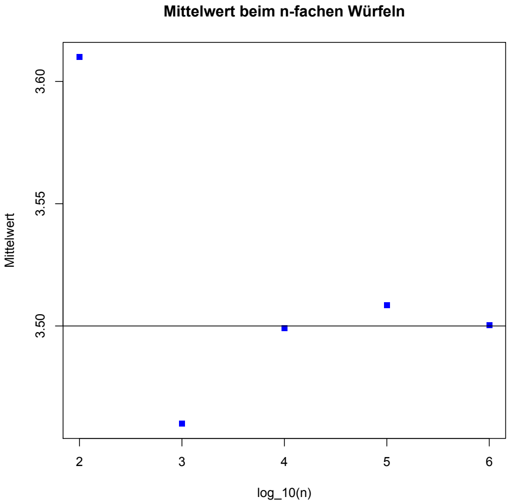
Gesetz der großen Zahlen (Ausblick auf Kapitel 9)

Erinnerung: Beispiel 5.1: Fairer Würfel, $X =$Ergebnis eines Wurfs.
Simulation: 100x Würfeln, y_i Ergebnis des $i-$ten Wurfs.



$$\frac{1}{100}\sum_{i=1}^{100} y_i = 3.61 \approx 3.5 = \mathbb{E}[X]$$

Mittelwert beim n-fachen Würfeln



Für große n nähert sich der beobachtete Mittelwert $\frac{1}{n}\sum_{i=1}^{n} y_i$ dem Erwartungswert an (vgl. später Kap. 9: Gesetz der großen Zahlen).

Gesetz der großen Zahlen

(Satz 9.2: Gesetz der großen Zahlen). Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen auf $(\Omega, \mathbb{P})$ mit $\mathbb{V}(X_i) < \infty$. Sei

$$S_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Die Folge von Zufallsvariablen $(S_n)_{n\in\mathbb{N}}$ konvergiert gegen den Erwartungswert $\mathbb{E}[X_1]$ in dem Sinne, dass für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} \mathbb{P}(|S_n - \mathbb{E}[X_1]| > \varepsilon) = 0.$$

- Bei vielen Versuchen tritt im Durchschnitt der Erwartungswert auf; für n groß ist $S_n \approx \mathbb{E}[X]$.
- Anwendung: Schätzung von Parametern (später)

Zum Beweis des Gesetzes der großen Zahlen

Sei $Y = S_n - \mathbb{E}[S_n]$. Es gilt nach Satz 5.21 (a) und (b)

$$\mathbb{V}(Y) = \mathbb{V}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[S_n]\Big) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n^2} n \cdot \mathbb{V}(X_1) = \frac{\mathbb{V}(X_1)}{n}.$$

Somit folgt aus der Chebyshev-Ungleichung für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|S_n - \mathbb{E}[S_n]| > \varepsilon) \le \lim_{n\to\infty} \frac{\mathbb{V}(Y)}{\varepsilon^2} = \lim_{n\to\infty} \frac{\mathbb{V}(X_1)}{n\varepsilon^2} = 0.$$