Stochastik für die Informatik, Vorlesung 9

Inhalt

- Erwartungswert: Beispiele
- Varianz
- Chebyshev-Ungleichung

Lernziele

- Das Bildchensammelproblem und den randomisierten Quicksort-Algorithmus kennen und die jeweiligen Erwartungswerte berechnen knnen
- Den Begriff der Varianz kennen, und Varianzen in Beispielen berechnen können

Vorkenntnisse

Stoff der bisherigen Vorlesungen,

Erinnerung: Erwartungswert

(Def. 5.1) Sei X eine diskrete Zufallsvariable. Der Erwartungswert von X ist definiert als

$$\boxed{\mathbb{E}[X] := \sum_{k \in X(\Omega)} k \cdot \mathbb{P}(X = k)}$$

"Mittelwert" der Zufallsvariable.

Eigenschaften des Erwartungswerts

(Satz 5.1) Seien X und Y diskrete Zufallsvariablen auf $(\Omega, \mathbb{P})$. Dann gelten

(a) $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

(b) $\mathbb{E}[a \cdot X] = a \cdot \mathbb{E}[X]$ für jedes $a \in \mathbb{R}$

(c) Falls $X(\omega) \geq 0 \forall \omega$, so ist $\mathbb{E}[X] \geq 0$.

(d) Falls $X = c$ konstant ist, so ist $\mathbb{E}[X] = c$.

(e) Sei $f : \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann gilt

$$\mathbb{E}[f(X)] = \sum_{k \in X(\Omega)} f(k)\mathbb{P}(X = k).$$

- (Beispiel 5.3/5.8: Geometrische Verteilung und coupon collector problem)

Beispiel: Randomisiertes Quicksort

- zufälliger Sortieralgorithmus
- rekursives Vorgehen

Algorithmus

- Eingabe: Eine Menge $S = \{x_1, ..., x_n\}$ von paarweise verschiedenen Zahlen
- Ausgabe: Die Elemente von S aufsteigend sortiert

1. Wähle zufällig ein Element Y von S
2. Bestimme die Mengen $S_<$ und $S_>$ bestehend aus den Elementen von S welche kleiner bzw. größer sind als Y, indem jedes Element von S mit Y verglichen wird.
3. Ordne rekursiv die Mengen $S_<$ und $S_>$ durch Wiederholen von Schritt 2.
4. Gebe die sortierte Menge $S_<$ gefolgt von Y gefolgt von der sortierten Menge $S_>$ aus

Randomisiertes Quicksort

Laufzeitanalyse

- Definition 12.2: Maß für die Laufzeit: X_S :=Anzahl durchgeführter Paarvergleiche bei Eingabe einer Liste S.
- X_S ist eine Zufallsvariable

Worst-case Laufzeit: Wählt man in jedem Schritt das aktuell kleinste (bzw. größte Element) der Liste, so ist die Laufzeit

$$\frac{(n-1)n}{2} \approx \frac{n^2}{2}.$$

Randomisiertes Quicksort

Beim randomisierten Quicksort ist $X_S = X_n$, d.h. die Laufzeit hängt nur von der Länge n der Liste ab

Satz 12.1: Erwartete Laufzeit. Im randomisierten Quicksort gilt

$$\mathbb{E}[X_n] \leq 2nH_n \in O(n \log n),$$

wobei $H_n = \sum_{i=1}^{n} \frac{1}{i} \in O(\log n)$ ist.

- Skizze: Baumstruktur
- Beweis

Streuung

Beispiel (Glücksspiel)

Glücksspiel mit $\mathbb{P}(\text{Gewinn}) = \frac{1}{2}$. Im Falle eines Gewinns erhält der Spieler 1 Euro, im Falle des Verlusts verliert er denselben Betrag. Sei X die Gewinnsumme eines Spielers nach 5 Spielen.

Erwartungswert: $\mathbb{E}[X] = 0$.

Nun wird der Einsatz auf 1000 Euro erhöht.

Erwartungswert: $\mathbb{E}[X] = 0$ bleibt gleich.

Streuung

Beispiel (Glücksspiel)

Glücksspiel mit $\mathbb{P}(\text{Gewinn}) = \frac{1}{2}$. Im Falle eines Gewinns erhält der Spieler 1 Euro, im Falle des Verlusts verliert er denselben Betrag. Sei X die Gewinnsumme eines Spielers nach 5 Spielen.

Erwartungswert: $\mathbb{E}[X] = 0$.

Nun wird der Einsatz auf 1000 Euro erhöht.

Erwartungswert: $\mathbb{E}[X] = 0$ bleibt gleich.

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: $\mathbb{V}(X) = \mathrm{var}(X)$

Varianz

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

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

- Beachte die Reihenfolge von Quadrieren und Erwartungswert!
- Man kann zeigen dass die Varianz von X genau dann existiert, wenn der Erwartungswert von X^2 existiert. In diesem Fall existiert auch der Erwartungswert von X.
- (Beispiel 5.9: Bernoulli-Verteilung)

Varianz und Standardabweichung

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

(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)

Varianz und Standardabweichung

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

(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)

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)