Algorithmen, Wahrscheinlichkeit und Information

Chebyshev-Ungleichung

Theorem: Für ZV X und a > 0 gilt:

$$P[\mid X - E[X] \mid \geq a] \leq \frac{Var[X]}{a^2}$$

Beweis (mittels Markov-Ungleichung):

$$\text{Ereignis } \mid X - E[X] \mid \geq a$$

$$\Leftrightarrow (X - E[X])^2 \geq a^2$$

$$P[\mid X - E[X] \mid \geq a] = P[(X - E[X])^2 \geq a^2] \leq \frac{(X - E[X])^2}{a^2} = \frac{Var[X]}{a^2}$$

$*$ Nützliche Abschätzung, aber häufig zu schwach
Beispiel: Faire Münze, n Mal

$$X = \text{Anzahl Kopf} \qquad E[X] = \frac{n}{2}$$

$$Var[X] = E[X^2] - E[X]^2$$

$$Var[X] = \frac{n}{4}$$

$$P[X \geq \frac{7}{8} \cdot n] = P[\mid X - \frac{n}{2} \mid \geq \frac{3}{8} \cdot n] \leq \frac{Var[X]}{(\frac{3}{8} \cdot n)^2} = \frac{\frac{n}{4}}{\frac{9}{64} \cdot n^2} = \frac{16}{9 \cdot n}$$

Zum Vergleich:

$$P[X \geq \frac{7}{8} \cdot n] \leq \frac{4}{7} \text{ (Markov)}$$

$$P[X \geq \frac{7}{8} \cdot n] \leq 2 \cdot e^{-\frac{49}{374} \cdot n} < 2 \cdot e^{-\frac{5}{8} \cdot n} \text{ (Chernoff)}$$

Varianz der geometrischen ZV

X-Anzahl Versuche bis zum ersten Erfolg. Jedes Experiment hat Erfolg mit WSK p.

$$P[X = n] = (1-p)^{n-1} \cdot p$$

$$E[X] = \frac{1}{p}$$

Ereignis A: erste Experiment hat Erfolg

$$E[X^2] = P[\overline{A}] \cdot E[X^2 \mid \overline{A}] + P[A] \cdot E[X^2 \mid A]$$

$$E[X^2 \mid A] = 1$$

$$P[A] = p$$

$$E[X^2] = (1-p) \cdot E[X^2 \mid \overline{A}] + p$$

$$Z := \text{Anzahl Versuch nach dem ersten Versuch gegeben } \overline{A}$$

$$P_Z(z) = P_{X|\overline{A}}(z+1) \text{ für } z \geq 1 \qquad (\text{X = Z + 1})$$

$$E[X^2] = (1-p)\cdot E[(Z+1)^2]+p = (1-p)\cdot E[Z^2]+2\cdot(1-p)\cdot E[Z]+1-p+p = (1-p)\cdot E[Z^2]+\frac{2\cdot(1-p)}{p}+1$$

$$,\text{da } E[Z] = \frac{1}{p}$$

$$E[Z^2] = (1-p)\cdot E[Z^2] + \frac{2\cdot(1-p)}{p} + 1 = \frac{1}{p}\left(\frac{2\cdot(1-p)}{p}\cdot\frac{p}{p})\right) = \frac{2-p}{p^2}$$

$$E[Z]^2 = \left(\frac{1}{p}\right)^2$$

$$Var[Z] = E[Z^2] - E[Z]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = Var[X]$$

Bälle in Töpfen ("Balls into Bins")

m Bälle "fallen" in n Töpfe
X_i: Anzahl Bälle in Topf T_i
$E[X_i] = \frac{m}{n}$
mit welcher WSK ist X_i "nicht viel größer" als $\frac{m}{n}$
Sei $X_i = \sum_{j=1}^m X_{ij}$, $X-ij$ ist Indikator für j-ten Ball in Topf i.

$$P[X_{ij}=1] = \frac{1}{n}$$

$$P[X_{ij}=0] = \frac{n-1}{n}$$

$$E[X_{ij}] = \frac{1}{n}$$

$$Var[X_{ij}] = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}$$

$$E[X_i] = \frac{m}{n} \qquad Var[X_i] = \sum_{i=1}^{m} Var[X_{ij}] = \frac{m}{n} - \frac{m}{n^2} = \frac{m\cdot(n-1)}{n^2}$$

WSK für k oder mehr Bälle in T_i?

$$P[X_i \geq \frac{m}{n}+k] \leq P[\mid X_i - \frac{m}{n} \mid \geq k] \overset{(Chebyshev)}{\leq} \frac{Var[X_i]}{k^2} < \frac{m}{n\cdot k^2} =: \varepsilon$$

$$\frac{m}{n\cdot k^2} = \varepsilon \qquad k = \sqrt{\frac{m}{n\cdot\varepsilon}}$$

$$P[X_i \geq \frac{m}{n} + \sqrt{\frac{m}{n\cdot\varepsilon}}] \leq \varepsilon$$

$$P[\exists i : X_i \geq \frac{m}{n}+k] \leq \sum_{i=1}^{n} P[X_i \geq \frac{m}{n}+k] = n\cdot P[X_i \geq \frac{m}{n}+k]$$

$$*P[\exists i : X_i \geq \frac{m}{n}+k] \leq \frac{m}{k^2}$$

$$*P[\exists i : X_i \geq \frac{m}{n}+\sqrt{\frac{m}{\delta}}] \leq \delta \qquad \delta > 0$$

$$\Rightarrow \text{ max. Abweichung ist höchstens } \sqrt{\frac{m}{\delta}} \text{ mit WSK } \geq 1-\delta$$

Die Probabilistische Methode

- Oft können Probleme nicht analytisch gelöst werden
- Probabilistische Analysen können einfacher sein
- Zeige Existenz von bestimmten Objekten, dadurch dass sie mit positiver Wahrscheinlichkeit existieren

Über den Erwartungswert:
IDEE: ZV X hat Werte ≤ E[X] und ≥ E[X]
Theorem: Reelle ZV X mit E[X]=μ

$$P[X \geq \mu] > 0 \land P[X \leq \mu] > 0$$

Beweis:
Angenommen:

$$P[X \geq \mu] = 0$$

$$\mu = \sum_{X} x \cdot P_X(x) \leq \sum_{x<\mu} \mu \cdot P_X(x) \leq \mu \cdot \sum_{x<\mu} P_X(x) \leq \mu \cdot \sum_{x \in X} P_X(x) = \mu$$

Beispiel: SAT: Satisfiability-Problem
$\Psi(x_1, x_2, \ldots, x_n)$ Boolescher Ausdruck in n Variablen, finde eine Belegung von $x_1, \ldots, x_n$, so dass $\Psi(x_1, x_2, \ldots, x_n) = TRUE$
Beispiel: Konjunktive Normalform

$$\underbrace{\underbrace{(\overline{x_1} \lor x_3)}_{\text{Term}} \land (\overline{x_2} \lor x_1) \land (x_2 \lor \overline{x_3} \lor x_4) ...}_{\text{n Terme}}$$

- SAT-Entscheidungsproblem ist NP-vollständig
- Optimierungsproblem: finde Belegung, so dass möglichst viele Terme erfüllt