Varianz einer geometrischen ZV

Lemma Varianz einer geometrischen ZV

Sei X eine geometrische ZV mit Parameter p. Dann gilt $\mathbf{Var}[X] = \frac{1-p}{p^2}$.

Beweis:

- Sei X ein ZV für die Anzahl Würfe, bis zum 1. Mal Kopf aufritt.
- Sei Y IV für Kopf im 1. Wurf. Dann gilt

$$\begin{aligned}
\mathbb{E}[X^2] &= \Pr(Y=0)\mathbb{E}[X^2 \mid Y=0] + \Pr(Y=1)\mathbb{E}[X^2 \mid Y=1] \\
&= (1-p)\mathbb{E}[(X+1)^2] + p = (1-p)\mathbb{E}[X^2] + 2(1-p)\mathbb{E}[X] + 1.
\end{aligned}$$

- Einsetzen von $\mathbb{E}[X] = \frac{1}{p}$ und Auflösen liefert $\mathbb{E}[X^2] = \frac{2-p}{p^2}$. Es folgt

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Analyse von Coupon Collector

Frage: Ws bei Coupon Collector $\geq 2n\sum_{i=1}^{n}\frac{1}{i} = 2nH_n$-mal zu ziehen?

Vergleich von Markov, Chebyshev und Union Bound:

- Markov liefert
$$\Pr(X \geq 2nH_n) \leq \frac{\mathbb{E}[X]}{2nH_n} = \frac{nH_n}{2nHn} = \frac{1}{2}.$$
- Wir verwenden die (vereinfachte) Ungleichung $\mathbf{Var}[X_i] \leq \frac{1}{p^2}$:
$$\mathbf{Var}[X] = \sum_{i=1}^{n}\mathbf{Var}[X_i] \leq \sum_{i=1}^{n}\left(\frac{n}{n-i+1}\right)^2 = n^2\sum_{i=1}^{n}\frac{1}{i^2} \leq n^2\frac{\pi^2}{6}.$$
- Chebyshev liefert damit
$$\Pr(|X-\mathbb{E}[X]| \geq nH_n) \leq \frac{\mathbf{Var}[X]}{n^2H_n^2} \leq \frac{\pi^2}{6H_n^2} = \mathcal{O}\left(\frac{1}{\ln^2 n}\right).$$
- Ereignis E_i: Element i ist nach $2\ln n$ Schritten nicht gezogen.
$$\Pr(E_i) = \left(1-\frac{1}{n}\right)^{2n\ln n} < e^{-2\ln n} = \frac{1}{n^2}.$$
- Ereignis E: Irgendein Element nach $2\ln n$ Schritten nicht gezogen.
$$\Pr(E) \leq \sum_{i=1}^{n}\Pr(E_i) < n\frac{1}{n^2} = \frac{1}{n}.$$
- D.h. Union Bound ist hier besser als Markov und Chebyshev.

Randomisierter Median

Problem Median-Berechung

Gegeben: $S \subset \mathbb{Z}$ mit $|S| = n$ ungerade

Gesucht: Median m von S, d.h. das $\frac{n+1}{2}$-te Element in Sortierung

Anmerkungen:

- 1. Lösung: Sortiere in $\mathcal{O}(n \log n)$ und gebe $\frac{n+1}{2}$-tes Element aus.
- $\exists$ komplexer deterministischer Algorithmus mit Laufzeit $\mathcal{O}(n)$.
- Idee eines probabilistischen Ansatzes:
  - Sample eine (kleine) Menge R, so dass $\ell, u \in R$ mit $\ell \leq m \leq u$.
  - Betrachte die Menge $C = \{s \in S \mid \ell \leq s \leq u\}$ mit $|C| = o(\frac{n}{\log n})$.
  - Bestimme die Anzahl x_ℓ der Element in S, die kleiner als ℓ sind.
  - Sortiere C und bestimme das $(\frac{n+1}{2} - x_\ell)$-kleinste Element in C.
- Führt zu einfachem $\mathcal{O}(n)$-Algorithmus mit besserer $\mathcal{O}$-Konstante. (im Vergleich zum oben erwähnten deterministischen Algorithmus)

Randomisierter Median

Algorithmus MEDIAN

EINGABE: S mit $|S| = n$ ungerade

1. Wähle Multimenge $R \subset_R S$, $|R| = n^{\frac{3}{4}}$ (Ziehen mit Zurücklegen).
2. Sortiere R und setze ℓ, u auf das $(\frac{1}{2}n^{\frac{3}{4}} \pm \sqrt{n})$-kleinste Element.
3. Berechne $C = \{s \in S \mid \ell \leq s \leq u\}$.
4. Berechne $x_\ell = |\{s \in S \mid s < \ell\}|$ und $x_u = |\{s \in S \mid s > u\}|$.
5. Falls $|C| > 4n^{\frac{3}{4}}$, $x_\ell > \frac{n}{2}$ oder $x_u > \frac{n}{2}$, Ausgabe FAIL.
6. Sortiere C und gib das $(\frac{n+1}{2} - x_\ell)$-kleinste Element in C aus.

AUSGABE: Median m von S

Korrektheit von MEDIAN

Satz Korrektheit von MEDIAN

MEDIAN gibt in $\mathcal{O}(n)$ entweder den Median oder FAIL aus.

Beweis:

- Der Median ist in C gdw $x_\ell \leq \frac{n}{2}$ und $x_u \leq \frac{n}{2}$.
- Die Laufzeit in Schritt 3 und 4 ist jeweils $\mathcal{O}(n)$.
- $|C| \leq 4n^{\frac{3}{4}}$ sichert, dass Schritt 6 nur Zeit $o(n)$ benötigt. □

Wir erhalten FAIL für eines der folgenden Ereignisse:

1. $E_1 : Y_1 = |\{r \in R \mid r \leq m\}| < \frac{1}{2}n^{\frac{3}{4}} - \sqrt{n}$
   Für dieses Ereignis gilt $\ell > m$ und damit $x_\ell > \frac{n}{2}$.
2. $E_2 : Y_2 = |\{r \in R \mid r \geq m\}| < \frac{1}{2}n^{\frac{3}{4}} - \sqrt{n}$
   Für dieses Ereignis gilt $u < m$ und damit $x_u > \frac{n}{2}$.
3. $E_3 : |C| > 4n^{\frac{3}{4}}$

Ws von FAIL

Lemma Ws von E_1 (E_2 analog)

$\Pr(E_1) \leq \frac{1}{4} n^{-\frac{1}{4}}$

Beweis:

- IV $X_i = 1$, falls das i-te Element in R kleiner oder gleich m ist.
$$\Pr(X_i = 1) = \frac{\frac{n+1}{2}}{n} = \frac{1}{2} + \frac{1}{2n}$$
- ZV $Y_1 = \sum_{i=1}^{n^{\frac{3}{4}}} X_i$ ist verteilt gemäß $B(n', p) = B(n^{\frac{3}{4}}, \frac{1}{2} + \frac{1}{2n})$. D.h.
$$\textbf{Var}[Y_1] = n'p(1-p) = n^{\frac{3}{4}}(\tfrac{1}{2} + \tfrac{1}{2n})(\tfrac{1}{2} - \tfrac{1}{2n}) < \tfrac{1}{4} n^{\frac{3}{4}}.$$
- Chebyshev liefert
$$\begin{aligned} \Pr(E_1) &= \Pr(Y_1 < \frac{1}{2} n^{\frac{3}{4}} - \sqrt{n}) \\ &\leq \Pr(|Y_1 - \mathbb{E}[Y_1]| > \sqrt{n}) \leq \frac{\textbf{Var}[Y_1]}{n} < \frac{1}{4} n^{-\frac{1}{4}}. \square \end{aligned}$$

Ws von FAIL

Lemma Ws von E_3

$\Pr(E_3) \leq \frac{1}{2} n^{-\frac{1}{4}}$

Beweis:

- Falls $|C| > 4n^{\frac{3}{4}}$, sind $> 2n^{\frac{3}{4}}$ Elemente in C größer/kleiner als m.
- Ereignis $E_>$: Mehr als $2n^{\frac{3}{4}}$ Elemente in C sind größer als m.
- D.h. die Ordnung von u in S's Sortierung ist mindestens $\frac{1}{2}n + 2n^{\frac{3}{4}}$, bzw. $\geq \frac{1}{2}n^{\frac{3}{4}} - \sqrt{n}$ Elem. in R sind unter den $\frac{1}{2}n - 2n^{\frac{3}{4}}$ größten in S.
- Sei X ZV für die $(\frac{1}{2}n - 2n^{\frac{3}{4}})$-größten Elemente von S in R.
- X ist gemäß $B(n^{\frac{3}{4}}, \frac{1}{2} - 2n^{-\frac{1}{4}})$ verteilt mit $\mathbb{E}[X] = \frac{1}{2}n^{\frac{3}{4}} - 2\sqrt{n}$ und
$$\mathbf{Var}[X] = n^{\frac{3}{4}}(\tfrac{1}{2} - 2n^{-\frac{1}{4}})(\tfrac{1}{2} + 2n^{-\frac{1}{4}}) < \tfrac{1}{4}n^{\frac{3}{4}}.$$
- Chebyshev liefert $\Pr(E_>) =$
$$\Pr(X \geq \tfrac{1}{2}n^{\frac{3}{4}} - \sqrt{n}) \leq \Pr(|X - \mathbb{E}[X]| \geq \sqrt{n}) \leq \frac{\mathbf{Var}[X]}{n} < \tfrac{1}{4}n^{-\frac{1}{4}}.$$
- Analog $\Pr(E_<) < \frac{1}{4}n^{-\frac{1}{4}}$ und $\Pr(E_3) \leq \Pr(E_>) + \Pr(E_<) < \frac{1}{2}n^{-\frac{1}{4}}$.

Monte Carlo und Las Vegas

Korollar Ws von FAIL

MEDIAN gibt FAIL mit Ws $\geq n^{-\frac{1}{4}}$ aus.

Definition Monte Carlo Algorithmus

Ein *Monte Carlo Algorithmus* ist ein probabilistischer Algorithmus, der FAIL oder eine inkorrekte Ausgabe liefern kann.

Definition Las-Vegas Algorithmus

Ein *Las Vegas Algorithmus* ist ein probabilistischer Algorithmus, der stets die korrekte Antwort liefert.

Anmerkung: Beim Monte Carlo Alg. ist die Laufzeit im Gegensatz zum Las Vegas Alg. typischerweise keine Zufallsgröße.

Übung: Wandeln Sie MEDIAN von einem Monte-Carlo Alg. mit Laufzeit $\mathcal{O}(n)$ in einen Las-Vegas Alg. mit erwarteter Laufzeit $\mathcal{O}(n)$.