Chernoff Schranke für Poisson ZV

Satz Chernoff Schranke für Poisson ZV

Sei X eine Poisson ZV mit Parameter μ. Dann gilt

1. für $x > \mu$: $\Pr(X \geq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}$.
2. für $x < \mu$: $\Pr(X \leq x) \leq \frac{e^{-\mu}(e\mu)^x}{x^x}$.

Beweis:

- Wir zeigen nur die 1. Ungleichung, die 2. folgt analog. Es gilt
$$\Pr(X \geq x) = \Pr(e^{tX} \geq e^{tx}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} \leq e^{\mu(e^t-1)-xt}.$$
- Für die Wahl $t = \ln(\frac{x}{\mu}) > 0$ folgt
$$\Pr(X \geq x) \leq e^{x-\mu-x\ln(\frac{x}{\mu})} = e^{x-\mu-x\ln x+x\ln\mu} = \frac{e^{-\mu}(e\mu)^x}{x^x}. \qquad \square$$

Poisson als Grenzwert der Binomial-Verteilung

Satz Poisson ist Grenzwert der Binomial-Verteilung für kleine p

Sei $X_n \sim B(n,p)$, wobei $\lim_{n\to\infty} np = \lambda$ konstant ist. Dann gilt

$$\lim_{n\to\infty} \Pr(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!} \text{ für alle festen } k.$$

Beweisskizze:

- Es gilt unter Verwendung von $1 + x \leq e^x$ für $|x| \leq 1$

$$\Pr(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \leq \frac{n^k}{k!} p^k \frac{(1-p)^n}{(1-p)^k} \leq \frac{(np)^k}{k!} \frac{e^{-pn}}{1-pk}.$$

- Wegen $\lim_{n\to\infty} np = \lambda$ folgt $\lim_{n\to\infty} p = \frac{\lambda}{n} = 0$. Damit gilt

$$\lim_{n\to\infty} \Pr(X_n = k) \leq \frac{e^{-pn}(np)^k}{k!} = \frac{e^{-\lambda}\lambda^k}{k!}.$$

- Ähnlich kann man $\lim_{n\to\infty} \Pr(X_n = k) \geq \frac{e^{-\lambda}\lambda^k}{k!}$ zeigen. $\square$

Poisson Approximation für Balle und Urnen

Bälle-Urnen Modell:

- ZV $X_i^{(m)}$, $i = 1, \ldots, n$, der Bälle pro Urne sind nicht unabhängig.
- Z.B. gilt offenbar $X_n^{(m)} = m - \sum_{i=1}^{n-1} X_i^{(m)}$.
- Wir würden gerne die $X_i^{(m)}$ als **unabhängige** Poisson-ZV $Y_i^{(m)}$, $i = 1, \ldots, n$, mit $\mu = \frac{m}{n}$ behandeln (*Poisson-Fall*).

Lemma Poisson versus exakt

Die Verteilung $(Y_1^{(m)}, \ldots, Y_n^{(m)})$ eingeschränkt auf $\sum_i Y_i^{(m)} = k$ ist identisch zur Verteilung $(X_1^{(k)}, \ldots, X_n^{(k)})$, unabhängig von m.

Beweis:

- Für $\sum_i k_i = k$: $p_1 = \Pr((X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n)) = \frac{\binom{k}{k_1, \ldots, k_n}}{n^k}$.
- Für $p_2 = \Pr((Y_1^{(m)}, \ldots, Y_n^{(m)}) = (k_1, \ldots, k_n) \mid \sum_i Y_i^{(m)} = k)$ gilt

$$p_2 = \frac{\Pr((Y_1^{(m)} = k_1) \cap \ldots \cap (Y_1^{(m)} = k_n) \cap (\sum_i k_i = k))}{\sum_{i=1}^{n} Y_i^{(m)} = k} = \frac{\prod_{i=1}^{n} e^{-\frac{m}{n}} (\frac{m}{n})^{k_i} / (k_i!)}{e^{-m} m^k / k!} = p_1 . \Box$$

Poisson versus exakt

Satz Poisson versus exakt

Sei $f(x_1, \ldots, x_n)$ eine nicht-negative Funktion. Dann gilt

$$\mathbb{E}[f(X_1^{(m)}, \ldots, X_n^{(m)})] \leq e\sqrt{m} \cdot \mathbb{E}[f(Y_1^{(m)}, \ldots, Y_n^{(m)})].$$

Beweis: Unter Verwendung der Abschätzung $m! \leq e\sqrt{m}(\frac{m}{e})^m$ gilt

$$\begin{aligned}
\mathbb{E}[f(Y_1^{(m)}, \ldots, Y_n^{(m)})] &= \sum_{k=0}^{\infty} \mathbb{E}[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) \mid \sum_{i=1}^{n} Y_i^{(m)} = k] \cdot \Pr(\sum_{i=1}^{n} Y_i^{(m)} = k) \\
&\geq \mathbb{E}[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) \mid \sum_{i=1}^{n} Y_i^{(m)} = m] \cdot \Pr(\sum_{i=1}^{n} Y_i^{(m)} = m) \\
&= \mathbb{E}[f(X_1^{(m)}, \ldots, X_n^{(m)})] \cdot \frac{e^{-m} m^m}{m!} \\
&\geq \frac{1}{e\sqrt{m}} \cdot \mathbb{E}[f(X_1^{(m)}, \ldots, X_n^{(m)})]. \quad \square
\end{aligned}$$

Poisson versus exakt

Korollar Poisson versus exakt

Jedes Ereignis E, dass Ws p im Poisson-Fall besitzt,
besitzt Ws $\leq e\sqrt{m}p$ im exakten Bälle/Urnen-Fall.

Beweis: Sei f die Indikatorfunktion von E. Dann ist $\mathbb{E}[f] = \Pr(E)$. $\square$

Übung: Für spezielle f sind noch bessere Schranken möglich.

Untere Schranke für maximale Beladung

Satz Untere Schranke für maximale Beladung

Wir werfen n Bälle in n Urnen. Dann besitzt für hinreichend große n eine Urne mindestens $\frac{\ln n}{\ln \ln n}$ Bälle mit Ws $\geq 1 - \frac{1}{n}$.

Beweis:

- Modelliere Anzahl X_i der Bälle in Urne i als Poisson-ZV mit $\mu = 1$.
- Es gilt $\Pr(X_i \geq M) \geq \frac{e^{-\mu}\mu^M}{M!} = \frac{1}{eM!}$.
- D.h. alle Urnen haben weniger als M Bälle mit Ws höchstens
$$\left(1 - \frac{1}{eM!}\right)^n \leq e^{-\frac{n}{eM!}}.$$
- Falls $e^{-\frac{n}{eM!}} \leq \frac{1}{n^2}$, ist diese Gegenws im exakten Fall $\leq \frac{e\sqrt{n}}{n^2} < \frac{1}{n}$.
- Für die Wahl $M = \frac{\ln n}{\ln \ln n}$ folgt $e^{-\frac{n}{eM!}} \leq \frac{1}{n^2}$ (nach etwas Rechnen). $\square$

Anwendung Hash-Ketten

Szenario: Wörterbuchsuche

- Besitzen sortierte Black-List mit n nicht erlaubten Passworten.
- Überprüfen eines einzelnen Passworts benötigt $\Theta(\log n)$ Schritte.
- Frage: Effizientere Datenstruktur für die Wörterbuchsuche?

Datenstruktur Hash-Kette:

- Hashfunktion $f : U \to [1, n]$ mit $\Pr(f(x) = j) = \frac{1}{n}$ für alle $x \in U$.
- Hashkollisionen werden mittels verketteter Listen behandelt.
- D.h. die n Passwörter entsprechen Bällen, die n Hashwerte Urnen.
- Suche eines Wortes: erwartet $\Theta(1)$ und $\mathcal{O}(\frac{\ln n}{\ln \ln n})$ mit hoher Ws.

Scheduling und Leader Election

Szenario: Scheduling

- n Nutzer $u_i \in U$ wollen gleichzeitig einen Rechner nutzen.
- Müssen Reihenfolge festlegen, d.h. eine Permutation wählen.

Lösung mittels Hashing:

- Wähle eine Hashfunktion $f : U \to [1, n^3]$ mit $\Pr(x = j) = \frac{1}{n^3}$.
- Nutzer u_i mit kleinstem Hashwert $f(u_i)$ kommt zuerst, usw.
- Benötigen: Keine zwei Nutzer erhalten denselben Hashwert.
- Für ein festes u_i gilt $f(u_i) = f(u_j)$ für ein $j \neq i$ mit Ws
$$1 - (1 - \frac{1}{n^3})^{n-1} \leq \frac{n-1}{n^3} < \frac{1}{n^2}.$$
- Union Bound: 2 Nutzer besitzen denselben Hashwert mit Ws $< \frac{1}{n}$.

Leader Election: Wähle Leader u_i mit kleinstem Hashwert $f(u_i)$.