Mengen Balancierung

Problem Mengen Balancierung

Gegeben: $A \in \mathbb{F}_2^{n \times m}$
Gesucht: $\mathbf{b} \in \{-1, 1\}^m$, so dass $||A\mathbf{b}||_\infty$ minimal ist.

Satz

Für zufällige $b \in_R \{-1, 1\}^m$ gilt $\Pr(||A\mathbf{b}||_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$.

Beweis:

- Sei $A\mathbf{b} = \mathbf{c}$ und k die Anzahl Einsen in der i-ten Zeile $\mathbf{a}_i$ von A.
- Falls $k \leq \sqrt{4m \ln n}$ dann gilt $c_i \leq \sqrt{4m \ln n}$. Sei also $k > \sqrt{4m \ln n}$.
- Die k Nicht-Null Terme in $\langle \mathbf{a}_i, \mathbf{b} \rangle$ sind unabhängige (± 1)-ZV X_i.
- Es gilt $\Pr(X_i = 1) = \Pr(X_i = (-1)) = \frac{1}{2}$. Mittels Chernoff folgt

$$\Pr(|\langle \mathbf{a}_i, \mathbf{b} \rangle| \geq \sqrt{4m \ln n}) \leq 2e^{-\frac{4m \ln n}{2k}} \leq \frac{2}{n^2}, \text{ wegen } k \leq m. \quad \square$$

Übung: Es existieren A, für die $||A\mathbf{b}|| = \Omega(\sqrt{n})$ für alle $\mathbf{b}$.

Geburtstags-Paradoxon (informal)

Problem Geburtstags-Paradoxon

Gegeben: n mögliche Geburtstage

Gesucht: m Personen, so dass 2 Personen am selben Tag Geburtstag mit Ws $\geq \frac{1}{2}$ haben.

Analyse:

- m Personen haben verschiedene Geburtstage mit Ws
$$(1-\tfrac{1}{n})\cdot(1-\tfrac{2}{n})\cdot\ldots\cdot(1-\tfrac{m-1}{n}) = \prod_{j=1}^{m-1}(1-\tfrac{j}{n}).$$
- Aus der Taylorentwicklung von e^x folgt $1-\frac{k}{n} \approx e^{-\frac{k}{n}}$ für $k \ll n$. D.h.
$$\prod_{j=1}^{m-1}(1-\tfrac{j}{n}) \approx \prod_{j=1}^{m-1} e^{-\frac{j}{n}} = e^{-\sum_{j=1}^{m-1}\frac{j}{n}} \approx e^{-\frac{m^2}{2n}}.$$
- Wir erhalten $e^{-\frac{m^2}{2n}} = \frac{1}{2}$ für $m = \sqrt{2n\ln 2}$.
- Approximation liefert für $n = 365$ den Wert $m \approx 22.49$.

Das Bälle-Urnen Modell

Definition Bälle-Urnen Modell

Im *Bälle-Urnen Modell* werfen wir m Bälle in n Urnen.

Interessante Fragestellungen:

- Wieviele Urnen bleiben leer?
- Wieviele Bälle sind in der vollsten Urne?
- Wann enthält eine Urne mehr als einen Ball? (Geburtstags-Paradoxon)
- Wann enthalten alle Urnen mindestens einen Ball? (Coupon Collector)

Maximale Beladung

Satz Maximale Beladung einer Urne

Für n Bälle in n Urnen und hinreichend große n enthält kein Urne mehr als $3\frac{\ln n}{\ln \ln n}$ Bälle mit Ws höchstens $\frac{1}{n}$.

Beweis:

- Ereignis E_i: Urne i enthält M Bälle.
- Es existieren $\binom{n}{M}$ Mengen mit M Bällen.
- Jede dieser Mengen ist mit Ws $(\frac{1}{n})^M$ komplett in Urne i, d.h.
$$\Pr(E_i) \leq \binom{n}{M}(\tfrac{1}{n})^M \leq \tfrac{1}{M!}.$$
- Es gilt $e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} > \frac{k^k}{k!}$, d.h. $k! > (\frac{k}{e})^k$. Damit $\Pr(E_i) \leq (\frac{e}{M})^M$.
- Für $M \leq 3\frac{\ln n}{\ln \ln n}$ gilt für hinreichend große n

$$\begin{aligned}
\Pr(E) &\leq \Pr(E_1) + \ldots + \Pr(E_n) \leq n\left(\frac{e}{M}\right)^M \leq n\left(\frac{e \ln \ln n}{3 \ln n}\right)^{3\frac{\ln n}{\ln \ln n}} \\
&\leq n\left(\frac{\ln \ln n}{\ln n}\right)^{3\frac{\ln n}{\ln \ln n}} = e^{\ln n}(e^{\ln \ln \ln n - \ln \ln n})^{3\frac{\ln n}{\ln \ln n}} = \frac{1}{n^2} \cdot n^{o(1)} \leq \frac{1}{n}.
\end{aligned}$$

Anwendung BUCKET SORT

Algorithmus BUCKET SORT

EINGABE: $n = 2^m$ Zahlen $x_1, \ldots, x_n \in_R [0, 2^k)$ mit $k \geq m$.

1. For $i = 1$ to n: Sortiere x_i in Bucket $MSB_m(x_i)$. (m oberste Bits)
2. For $i = 0$ to $n - 1$: Sortiere Bucket i aufsteigend mit INSERTION SORT.

AUSGABE: Zahlen in den Buckets $0, \ldots, n - 1$

Korrektheit:

- Schritt 1: Elemente in Bucket i sind kleiner als die in Bucket $i + 1$.
- Schritt 2: Zusätzliche Sortierung der Elemente pro Bucket.

Analyse BUCKET SORT

Satz Laufzeit BUCKET SORT

BUCKET SORT läuft in erwarteter Zeit $\mathcal{O}(n)$.

Beweis:

- Die Zahlen entsprechen Bällen, die Buckets entsprechen Urnen.
- Schritt 1 läuft in deterministischer Zeit $\mathcal{O}(n)$.
- ZV X_i für die Anzahl Zahlen in Bucket i.
- Die Laufzeit für Bucket i ist höchstens cX_i^2 für eine Konstante c.
- Damit ist die erwartete Laufzeit von Schritt 2 höchstens
$$\mathbb{E}[\sum_{i=0}^{n-1} c(X_i)^2] = c\sum_{i=0}^{n-1}\mathbb{E}[X_i^2] = cn\mathbb{E}[X_0^2].$$
- Da $X_0 \sim B(n, \frac{1}{n})$ wissen wir bereits
$$\mathbb{E}[X_0^2] = n(n-1)p^2 + np = \frac{n(n-1)}{n^2} + 1 < 2.$$
- Damit läuft Schritt in erwarteter Zeit $\mathcal{O}(n)$ $\quad\Box$.

Die Poisson Verteilung

Motivation:

- Wir betrachten den Besetzungsgrad von Urnen.
- ZV $X_i = 1$ gdw die j-te Urne leer ist. D.h. $\mathbb{E}[X_i] = (1 - \frac{1}{n})^m \approx e^{-\frac{m}{n}}$.
- Es folgt für die Anzahl X leerer Urnen
$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n(1 - \tfrac{1}{n})^m \approx ne^{-\frac{m}{n}}.$$
- D.h. der relative Anteil leerer Urnen ist approximativ $e^{-\frac{m}{n}}$.
- Generell: Ws, dass eine feste Urne genau j Bälle enthält, ist
$$p_j = \binom{m}{j}(\tfrac{1}{n})^j(1 - \tfrac{1}{n})^{m-j} = \frac{1}{j!}\frac{m(m-1)\ldots(m-j+1)}{n^j}(1 - \tfrac{1}{n})^{m-j} \approx \frac{e^{-\frac{m}{n}}(\frac{m}{n})^j}{j!}.$$

Die Poisson Verteilung

Definition Poisson Verteilung

Eine ZV X ist *Poisson* verteilt mit Parameter μ, falls für alle $j \geq 0$

$$\Pr(X = j) = \frac{e^{-\mu}\mu^j}{j!}.$$

Anmerkungen:

- Ws-Verteilung: $\sum_{j\geq 0} \Pr(X = j) = e^{-\mu} \sum_{j\geq 0} \frac{\mu^j}{j!} = e^{-\mu} e^{\mu} = 1.$
- Für den Erwartungswert einer Poisson verteilten ZV X gilt
$$\mathbb{E}[X] = \sum_{j\geq 1} j \frac{e^{-\mu}\mu^j}{j!} = \mu \sum_{j\geq 1} \frac{e^{-\mu}\mu^{j-1}}{(j-1)!} = \mu \sum_{j\geq 0} \frac{e^{-\mu}\mu^j}{j!} = \mu.$$
- Bälle-Urnen: Die Verteilung ist approximativ Poisson mit $\mu = \frac{m}{n}$.
- $\mu = \frac{m}{n}$ entspricht der durchschnittlichen Belegung der Urnen.

Summe unabhängiger Poisson ZV

Satz Moment Erzeugendenfunktion einer Poisson ZV

Sei eine ZV X Poisson verteilt mit μ. Dann gilt $M_X(t) = e^{\mu(e^t-1)}$.

Beweis:

$$\mathbb{E}[e^{tX}] = \sum_{j\geq 0} \frac{e^{-\mu}\mu^j}{j!} e^{tj} = e^{-\mu} \sum_{j \geq 0} \frac{(\mu e^t)^j}{j!} = e^{-\mu} e^{\mu e^t} = e^{\mu(e^t-1)} \quad \square$$

Satz Summe unabhängiger Poisson ZV

Die endliche Summe unabhängiger Poisson ZV ist eine Poisson ZV.

Beweis:

- Wir betrachten nur die Summer zweier ZV. Der Satz folgt induktiv.
- Seien X, Y ZV mit Erwartungswerten μ_1, μ_2.
- Wir erhalten $M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{(\mu_1+\mu_2)(e^t-1)}$.
- Dies ist Erzeugendenfkt einer Poisson ZV mit Parameter $\mu_1 + \mu_2$.
- Mit Folie 47 ist $X + Y$ damit eine ZV mit Erwartungswert $\mu_1 + \mu_2$. $\square$