Chernoff Bounds

CSC 473 Advanced Algorithms

Computer Science
UNIVERSITY OF TORONTO

Variance and Chebyshev

- Let $X_1, \dots, X_n \in \{0,1\}$ be independent random variables
  - Not necessarily uniform or identically distributed
- Remember, for $X = \sum_{i=1}^{n} X_i$:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] \qquad Var(X) = \sum_{i=1}^{n} Var(X_i) \leq \mathbb{E}[X]$$

- By Chebyshev's inequality:

$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) \leq \frac{Var(X)}{\delta^2 \mathbb{E}[X]^2} \leq \frac{1}{\delta^2 \mathbb{E}[X]}$$

The Chernoff Bound

- Let $X_1, \ldots, X_n \in \{0,1\}$ be independent random variables
  - Not necessarily uniform or identically distributed
- Chernoff Bound: if $\mathrm{X} = \sum_{i=1}^{n} X_i$ and $\mathbb{E}[X] \leq \mu$

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)}\right)^{(1+\delta)\mu}$$

- For $0 \leq \delta \leq 1$, the right hand side is $\leq e^{-\delta^2 \mu/3}$
  - Compare with $\frac{1}{\delta^2 \mu}$ from Chebyshev.

Proof Idea

- "Chernoff trick": for any $t \geq 0$, by Markov's inequality

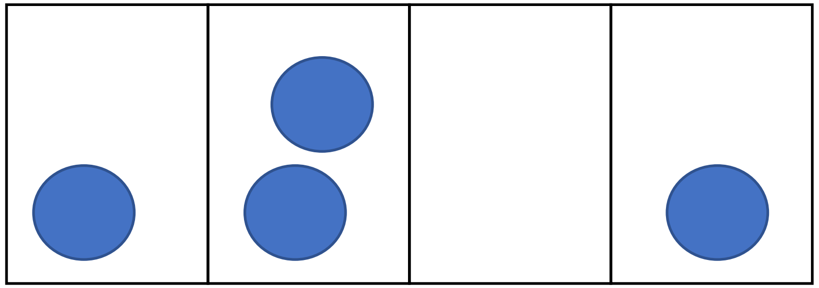
$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}[X]) = \mathbb{P}\left(e^{tX} \geq e^{t(1+\delta)\mathbb{E}[X]}\right) \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mathbb{E}[X]}}$$

- By independence of $X_1, \dots, X_n$

$$\mathbb{E}[e^{tX}] = \mathbb{E}\left[e^{t(X_1+\cdots+X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}]$$

- Using $1 + z \leq e^z$, $\mathbb{E}[e^{tX_i}] \leq e^{\mathbb{E}[X_i]\left(e^t - 1\right)}$, so $\mathbb{E}[e^{tX}] \leq e^{\mathbb{E}[X]\left(e^t-1\right)}$
    - Choosing $t = \ln(1+\delta)$ gives the best bound.

Balls and Bins



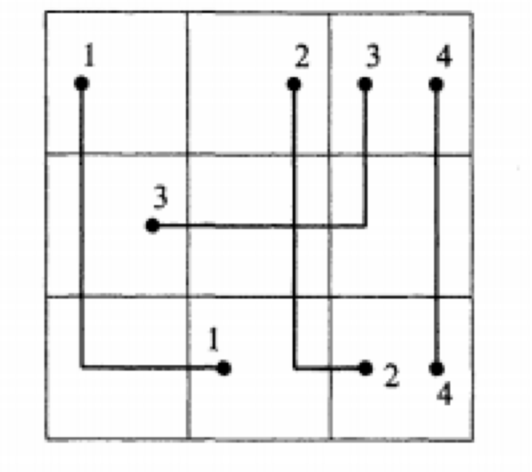
- Suppose we throw n balls into n bins
  - Each ball lands in a uniformly random bin, independently from the others
- **Theorem** With prob. $\geq \frac{1}{2}$, no bin has more than $O\left(\frac{\log n}{\log\log n}\right)$ balls
- $X_{ij} = 1 \Leftrightarrow$ ball j lands in bin i. $X_i = \sum_{j=1}^{n} X_{ij}$ number of balls in bin i.
  - $\mathbb{E}\left[X_{ij}\right] = \mathbb{P}\left(X_{ij} = 1\right) = \frac{1}{n}$, so $\mathbb{E}[X_i] = 1$.
- *Chernoff*: $\mathbb{P}\left(X_i \geq \frac{c \ln n}{\ln \ln n}\right) \leq \frac{1}{2\text{n}}$ for all large enough c, n
  - Use $\mu = 1$, $1 + \delta = \frac{c \ln n}{\ln \ln n}$. Then $\left(\frac{e^{\delta}}{(1+\delta)}\right)^{(1+\delta)} \approx e^{-c \ln n} = n^{-c}$
- *Union bound*: $\mathbb{P}\left(\exists i: X_i \geq \frac{c \ln n}{\ln \ln n}\right) \leq n \cdot \frac{1}{2n} = \frac{1}{2}$

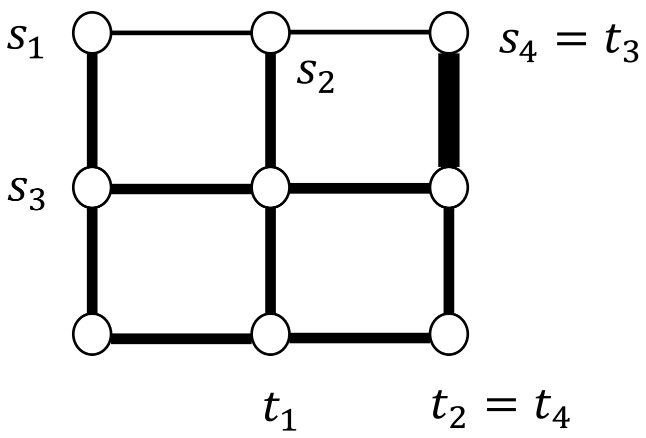
Multicommodity Flow Problem

- Motivation: given a chip with "wire channels", connect locations with wires, so that no channel is overloaded
- **Multicommodity Flow**: Given an undirected graph $G = (V, E)$, and vertices $s_1\ t_1, \dots, s_k, t_k$, find paths P_i in G connecting s_i and t_i so that the maximum number of paths going through any edge is minimized.

Squares -> vertices

Boundaries -> edges





LP Relaxation

- $\mathcal{P}_i$ = all paths between s_i and t_i. $\mathcal{P} = \bigcup_{i=1}^{k} \mathcal{P}_i$
- Exponential size relaxation: introduce a variable x_P for every $P \in \mathcal{P}$

No edge used by more than W paths

One path between s_i and t_i for each i.

$$\min W$$

s.t.

$$\sum_{P \in \mathcal{P}: e \in P} x_P \leq W \quad \forall e \in E$$

$$\sum_{P \in \mathcal{P}_i} x_P = 1 \quad \forall i \in [k]$$

$$x_P \in \{0,1\} \quad \forall P \in \mathcal{P}$$

$$\min W$$

s.t.

$$\sum_{P \in \mathcal{P}: e \in P} y_P \leq W \quad \forall e \in E$$

$$\sum_{P \in \mathcal{P}_i} y_P = 1 \quad \forall i \in [k]$$

$$y_P \geq 0 \quad \forall P \in \mathcal{P}$$

- Can be solved in polynomial time, and only poly-many y_P are not 0

Randomized Rounding

$$\min W$$

s.t.

$$\sum_{P \in \mathcal{P}: e \in P} y_P \leq W \quad \forall e \in E$$

$$\sum_{P \in \mathcal{P}_i} y_P = 1 \quad \forall i \in [k]$$

$$y_P \geq 0 \quad \forall P \in \mathcal{P}$$

- Solve the LP to get optimal y_P, with value $LP = W$
- $\{y_P : P \in \mathcal{P}_i\}$ give a probability distribution over $\mathcal{P}_i$
- Independently for each $i \in [k]$:
  - Sample $P_i \in \mathcal{P}_i$ with probability y_P
- $Z_{e,i} = 1 \Leftrightarrow e \in P_i$. $Z_e = \sum_{i=1}^{k} Z_{e,i}$ is the load on edge e
- $\mathbb{E}[Z_e] = \sum_{P \in \mathcal{P}: e \in P} y_P \leq LP \leq OPT$
- **Theorem.** With prob. $\geq \frac{1}{2}$, $\max_{e \in E} Z_e = O\left(\frac{\log n}{\log \log n}\right) \cdot OPT$
  - Same calculation as Balls & Bins, with $\mu = OPT \geq 1$.

More on Multicommodity Flow

- Much better approximation if LP (or OPT) is large
- E.g., if $LP \geq 10 \ln n$, then, with prob. $\geq 1/2$, randomized rounding finds a solution with value $\leq LP + \sqrt{10\ LP \ln n} < 2\ LP$.

- Under an assumption slightly stronger than P $\neq$ NP (NP doesn't have randomized algorithms running in expected time $n^{\log^{O(1)} n}$), the $O\left(\frac{\log n}{\log \log n}\right)$ approximation is best possible in the worst case.