{ csci 3|6907 | Lecture 4 }

Hoeteck Wee · hoeteck@gwu.edu

Announcements

- Homework 2 is out, due next Wed

  feel free to discuss in groups

  homework must be written up **individually**

PART I | tail bounds

Tail bounds, III

Chernoff Bounds

Let $X_1, \ldots, X_n$ be *independent* r.v.'s assuming values in $\{0, 1\}$. Let $X = X_1 + X_2 + \cdots + X_n$ and $\mu = \mathrm{E}[X]$. Then,

1. For all $0 < \delta < 1$,

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

- PROOF IDEA. apply Markov's to non-negative r.v. e^{tX}.

$$\mathrm{E}[e^{tX}] = \prod_{i=1}^{n} \mathrm{E}[e^{tX_i}]$$

- EXAMPLE. toss n fair coins...

Chernoff Bounds

Let $X_1, \ldots, X_n$ be *independent* r.v.'s assuming values in $\{0, 1\}$. Let $X = X_1 + X_2 + \cdots + X_n$ and $\mu = \mathrm{E}[X]$. Then,

1. For all $0 < \delta < 1$,
$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$
2. For all $0 < \delta < 1$,
$$\Pr[X \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}$$
3. For all $\delta > 0$,
$$\Pr[X \geq (1 + \delta)\mu] \leq e^{-\frac{\mu\delta^2}{2+\delta}}$$

Comparison of tail bounds

- GENERALITY: Markov's $\gg$ Chebyshev's $\gg$ Chernoff

  (non-negative · bounded variance · independence)

- "ERROR": Markov's $\ll$ Chebyshev's $\ll$ Chernoff

  (constant · $1/\text{poly}$ · exponential)

- "DEVIATION": Markov's $\ll$ Chebyshev's $=$ Chernoff

  (one-sided · two-sided · two-sided)

PART 2 | birthday paradox & balls-and-bins

Birthday "Paradox"

QUESTION. What is the probability that amongst 30 people in a room, two share the same birthday?

MODEL. Everyone's birthday is independently and uniformly chosen at random amongst 365 days.

ANALYSIS. Pr[all birthdays are distinct] is

$$(1 - \tfrac{1}{365}) \cdot (1 - \tfrac{2}{365}) \cdot (1 - \tfrac{3}{365}) \cdots (1 - \tfrac{29}{365}) \approx 0.2937$$

MORE GENERALLY... For m people and n "birthdays", it's

$$\begin{aligned} & (1 - \tfrac{1}{n}) \cdot (1 - \tfrac{2}{n}) \cdot (1 - \tfrac{3}{n}) \cdots (1 - \tfrac{m-1}{n}) \\ \approx \quad & \prod_{j=1}^{m-1} e^{-j/n} = e^{-m(m-1)/2n} \approx e^{-m^2/2n} \end{aligned}$$

$\Rightarrow$ constant prob of "collision" whenever $m \gtrsim \sqrt{2n \ln 2}$

Interlude: Union Bound

Chernoff Bound

For any events E_1, E_2 not necessarily independent,

$$\Pr[E_1 \cup E_2] \leq \Pr[E_1] + \Pr[E_2]$$

- EXAMPLE. two types of errors: first w.p. ≤ 0.1, second w.p. ≤ 0.2.
- QUESTION. $\Pr[\text{no errors}] \geq \ldots$?
- GENERALIZATION.

  $$\Pr[E_1 \cup E_2 \cup E_3 \cdots] \leq \Pr[E_1] + \Pr[E_2] + \Pr[E_3] + \cdots$$

Balls-and-Bins Model

- m balls thrown into n bins
  - location of each ball independent and random
- Example: job scheduling
  - balls = tasks, bins = processors
- Quantities of interest
  - average load = expected number of balls in each bin
  - maximum load = number of balls in fullest bin
  - number of empty bins (= number of idle processors)
- L_i be r.v. for # balls in Bin i
  - $L_i \sim B(m, \frac{1}{n})$, so $\mathrm{E}[L_i] = \frac{m}{n}$, $\mathrm{Var}[L_i] = \frac{m}{n}(1 - \frac{1}{n})$

Average/Maximum Load

Chernoff Bound

Let $X_1, \ldots, X_n$ be *independent* $\{0,1\}$-r.v.'s. Let $X = X_1 + \cdots + X_n$ and $\mu = \mathrm{E}[X]$.
Then, for all $\delta > 0$, $\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2+\delta}}$

- APPLICATION. bounding $\Pr[L_i \geq 2\ln n + 1]$ for $m = n$
  - set $\mu = 1, \delta = 2\ln n$, so $\frac{\mu\delta^2}{2+\delta} \geq 2\ln n$
    $\Rightarrow \Pr[L_i \geq 2\ln n + 1] \leq e^{-2\ln n} = \frac{1}{n^2}$
  - By union bound, $\Pr[\bigvee_{i=1}^{n}(L_i \geq 2\ln n + 1)] \leq \frac{1}{n}$
  - Hence, $\Pr[\text{maximum load} \leq 2\ln n + 1] \geq 1 - \frac{1}{n}$.
  - e.g. $n = 1$ million, max load is at most 30 w.h.p.

Maximum Load for $m = n$

- **BETTER ANALYSIS.**

$$
\begin{aligned}
\Pr[L_i \geq k] &= \Pr[\exists \text{ subset of } k \text{ balls all of which fall into bin } i] \\
&\leq \binom{n}{k} \cdot (1/n)^k \\
&\leq (ne/k)^k \cdot (1/n)^k = (e/k)^k \\
&\leq 1/n^2 \qquad \text{for } k \geq \tfrac{3 \ln n}{\ln \ln n}
\end{aligned}
$$

- **BETTER BOUND.**
  - obtain a bound of $O(\frac{\log n}{\log \log n})$ instead of $O(\log n)$ for the maximum load.
  - e.g. $n = 1$ million, max load is at most 16 w.h.p.

Empty Bins

- Let X be random variable for # empty bins.
- Let X_i be r.v. indicating whether Bin i is empty.
- $\Pr[X_i = 1] = (1 - \frac{1}{n})^m$ and $\mathrm{E}[X] = n(1 - \frac{1}{n})^m$.
- NOTE. X_i and X_j are *not* independent, e.g.

  $\Pr[X_i = 1 \wedge X_j = 1] = (1 - \frac{2}{n})^m \neq \Pr[X_i = 1] \cdot \Pr[X_j = 1]$

Empty Bins: Variance

- Recall $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$
  - $\mathrm{E}[X^2] = \mathrm{E}[(X_1 + \cdots + X_n)^2] = \sum_{i=1}^{n} \mathrm{E}[X_i^2] + \sum_{i \neq j} \mathrm{E}[X_i X_j]$
  - If $X_i \in \{0, 1\}$, then $\mathrm{E}[X_i^2] = \mathrm{E}[X_i]$
- Computing $\mathrm{E}[X_i X_j]$
  - $\mathrm{E}[X_i X_j] = \Pr[X_i X_j = 1] = \Pr[X_i = 1 \wedge X_j = 1] = (1 - \frac{2}{n})^m$
- Computing $\mathrm{Var}[X]$
  - $\mathrm{E}[X^2] = n(1 - \frac{1}{n})^m + n(n-1)(1 - \frac{2}{n})^m$
  - $\mathrm{Var}[X] = n(1 - \frac{1}{n})^m + n(n-1)(1 - \frac{2}{n})^m - n^2(1 - \frac{1}{n})^{2m}$

PART 4 | random graphs

Random Graphs

- Random graph model $\mathcal{G}_{n,p}$
  - Distribution over undirected graphs on n vertices
  - Every edge occurs with probability p
  - Graph with given set of m edges has probability

$$p^m(1-p)^{\binom{n}{2}-m}$$

- Basic properties
  - Expected number of edges is $p\binom{n}{2}$
  - Each vertex has expected degree $p(n-1)$

Threshold behavior for triangles

- NEXT WEEK: show that for random graph model $\mathcal{G}_{n,p}$:

$$\Pr[G \text{ contains a triangle}] \overset{n\to\infty}{\longrightarrow} \begin{cases} 1 & \text{if } p = \omega(\frac{1}{n}) \\ 0 & \text{if } p = o(\frac{1}{n}) \end{cases}$$

  - If p grows faster than $\frac{1}{n}$, almost *every* graph contains a triangle
  - If p grows slower than $\frac{1}{n}$, almost *no* graph contains a triangle

- THRESHOLD BEHAVIOR: holds for many properties, e.g. "is connected", "contains a clique of size 4", with difference choices of "$\frac{1}{n}$"