Today

1) 2 Common Setups

2) Concentration Framework

3) Solving (1) w/ (2)

Admin

- Scribe Note Sign up
- Hw 1 in
- Hw 2 out
- Canvas
- Names

2 Scenarios

Balls $\rightarrow$ bins uniformly at random + indepently

o o o o … o n balls

$\sqcup \quad \sqcup \quad \sqcup \dots \sqcup$ m bins

1) If $m = n$, what is max # balls in a bin?

$\approx \ln n$ (w/ $Pr \geq 1 - \frac{1}{n}$)

2) How big must n be so ≥ 1 ball in each bin?

$\approx n \ln n$ (for $Pr \geq 1 - \frac{1}{n}$)

"Coupon collector"

How to Solve Your Fave Randomization Question

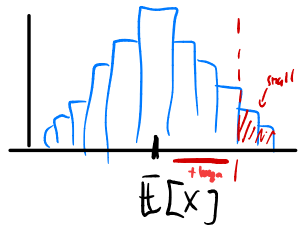
Concentration Framework

a) Show (*) true if all RVs near $\mathbb{E}$

b) Concentration: one RV at $\mathbb{E}$ ($\pm \log n$) w/ high Pr

c) Union Bound: all RVs at $\mathbb{E}$ ($\pm \log n$) w/ high Pr

↳ Reminder: $\Pr(B_1 \cup B_2 \cup \dots) \leq \sum_i \Pr(B_i)$



Let $X_{ij} :=$ indicator of ball $j \rightsquigarrow$ bin i

Let $X_i := \sum_j X_{ij}$ (# balls in bin i)

1) $\mathbb{E}[X_i] = 1$ (when $m = n$)

2) $\mathbb{E}[X_i] = \Omega(\log n)$ (when $m \geq \Omega(n \cdot \log n)$)

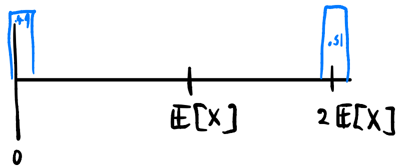
Markov's Inequality

Suppose X is a non-negative RV w/ $\mathbb{E}[X]>0$

$$\Pr(X \geq a\cdot\mathbb{E}[X]) \leq \frac{1}{a}$$

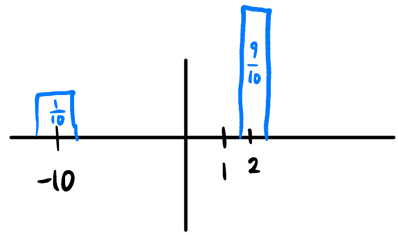
$\geq$ twice $\mathbb{E}$ $\leq \frac{1}{2}$ of the time

Intuition



Why Non-Negative?

$$\mathbb{E}[X] = \frac{4}{5}$$



Skip + comeback

$$a\cdot\mathbb{E}[X]\cdot\Pr(X \geq a\cdot\mathbb{E}[X]) \leq a\cdot\mathbb{E}[X]\cdot\sum_{i \geq a\cdot\mathbb{E}[X]} \Pr(X=i)$$

$$\leq \sum_{i \geq a\cdot\mathbb{E}[X]} i\cdot\Pr(X=i)$$

$$\leq \mathbb{E}[X] \qquad \text{(non-neg)}$$

Linear Scaling, only needs non-negative

Corollary: Chebyshev's Inequality

$$\Pr\left(|X-\mathbb{E}[X]| \geq a\right) \leq \frac{Var(X)}{a^2}$$

Let $Y = (X-\mathbb{E}[X])^2$ so $\mathbb{E}[Y] = Var(X)$

$$\Pr\left(\overbrace{|X-\mathbb{E}[X]| \geq a}^{\sqrt{Y} \geq a}\right) = \Pr\left(Y \geq a^2\right)$$

$$= \Pr\left(Y \geq \frac{a^2}{\mathbb{E}[Y]} \cdot \mathbb{E}[Y]\right)$$

$$\leq \frac{\mathbb{E}[Y]}{a^2} \qquad \text{(Markov)}$$

$$= \frac{Var(Y)}{a^2}$$

skip + come back if time

Quadratic scaling, needs low variance

Simple Chernoff

Let $X_1, X_2, \ldots, X_n$ be independent RVs s.t. $X_i = \begin{cases} 1 & \text{w/ Pr } p_i \\ 0 & \text{o/w} \end{cases}$

Let $X := \sum_i X_i$, $\mu := \mathbb{E}[X]$

Then $\forall\ \delta \geq 0$

$$\Pr\left(X \geq (1+\delta)\cdot\mu\right) \leq \exp\left(-\mu\delta^2/(2+\delta)\right)$$

And $\forall\ \delta \in (0,1)$

$$\Pr\left(X \leq (1-\delta)\cdot\mu\right) \leq \exp\left(-\delta^2\mu/2\right)$$

Exponential scaling, needs indicator $\sum_i$

$\mu\cdot\delta \approx c\cdot\log n \rightarrow \Pr \leq \frac{1}{n^c}$ → "with high probability"

Solving Problems w/ Concentration Framework

1) (a) Fix bin i.

$$X_i = \sum_j^t X_{ij} \quad \begin{cases} 1 & \text{w/ Pr } \frac{1}{n} \\ 0 & \text{o/w} \end{cases}$$

independent

$$\mu = \mathbb{E}[X_i] = 1$$

(b)

$$\Pr\left(X_i \geq \mu(1+\underbrace{2\log n}_{\delta})\right) \leq \exp\left(-4\log^2 n/(2+\log n)\right)$$

$$\leq \exp\left(-4\log^2 n/2\cdot\log n\right)$$

$$= n^{-2}$$

so $\Pr\left(X_i \geq 1+2\cdot\log n\right) \leq n^{-2}$

(c)

$$\Pr\left(X_1 \geq 1+2\cdot\log n \cup X_2 \geq 1+2\cdot\log n \cup \dots\right) \leq \sum_i^t \Pr(X_i \geq 1+2\cdot\log n)$$

$$\leq \frac{1}{n}$$

$\mathbb{E}[X_i] = O(1)$ so $X_i \leq \log n$ whp so all $X_i \leq \log n$

2) (a) Let $m = 16n\cdot\ln n$ and fix i

$$\mu = \mathbb{E}[X_i] = 16\ln n$$

(b)

$$\Pr\left(X_i \leq (1-\underbrace{\tfrac{1}{2}}_{\delta})\cdot 16\ln n\right) \leq \exp\left(-\tfrac{1}{4}16\ln n/2\right)$$

$$= n^{-2}$$

(c)

$$\Pr\left(X_1 \leq 8\ln n \cup X_2 \leq 8\ln n \cup \dots\right) \leq \frac{1}{n}$$

$\mathbb{E}[X_i] = \Theta(\log n)$
so $X_i = \Theta(\log n)$ whp
so all $X_i = \Theta(\log n)$ whp

Full Chernoff (upper Tail)

Let $X_1, X_2, \dots, X_n$ be independent RVs s.t. $X_i = \begin{cases} 1 & \text{w/ Pr } p_i \\ 0 & \text{o/w} \end{cases}$

Let $X := \sum_i X_i$, $\mu := \mathbb{E}[X]$

Then $\forall\, \delta \geq 0$, $\Pr(X \geq (1+\delta)\cdot\mu) \leq \left(\frac{\exp(\delta)}{(1+\delta)^{1+\delta}} \right)^{\mu}$

Let $s = \ln(1+\delta)$ and $a = (1+\delta)\mu$

$$\Pr(X \geq a) = \Pr\left(e^{sX} \geq e^{sa}\right) = \Pr\left(e^{sX} \geq e^{sa} \cdot \frac{\mathbb{E}[e^{sX}]}{\mathbb{E}[e^{sX}]}\right) \underset{\text{markov}}{\leq} \frac{\mathbb{E}[e^{sX}]}{e^{sa}} \underset{\text{dfn.}}{=} \frac{\mathbb{E}\left[\prod_i e^{sX_i}\right]}{e^{sa}}$$

$$\underset{\text{independence}}{=} \frac{\prod_i \mathbb{E}\left[e^{sX_i}\right]}{e^{sa}} = \frac{\prod_i p_i e^{s} + (1-p_i)}{e^{sa}}$$

$$e^{sX_i} = \begin{cases} e^{s} & \text{w/ Pr } p_i \\ 1 & \text{w/ Pr } 1-p_i \end{cases}$$

$$= \prod_i \frac{1 + p_i(e^{s} - 1)}{e^{sa}} \underset{1+x \leq \exp(x)}{\leq} \prod_i \frac{\exp\left(p_i(e^{s} - 1)\right)}{e^{sa}}$$

$$= \frac{\exp\left(\sum_i p_i (e^{s} - 1)\right)}{e^{sa}} = \frac{\exp\left(\mu(e^{s} - 1)\right)}{e^{sa}}$$

$$= \left(\frac{\exp(\delta)}{(1+\delta)^{1+\delta}} \right)^{\mu}$$

$s = \ln(1+\delta)$

$a = (1+\delta)\mu$

Proof of Simple Chernoff via Full Chernoff

Have

$$\frac{2\delta}{2+\delta} \le \log(1+\delta) \quad \forall \delta \ge 0 \quad \overset{\delta=\delta-1}{\Longleftrightarrow} \quad 2\delta \le \underbrace{(1+\delta)\cdot\log\delta + 2}_{\text{Convex and LHS tangent at } \delta=1 \text{ to RHS}} \quad \forall \delta \ge 1$$

So

$$\left(\frac{\exp(\delta)}{(1+\delta)^{1+\delta}}\right)^{\mu} = \left(\frac{\exp(\delta)}{\exp\left((1+\delta)\cdot\log(1+\delta)\right)}\right)^{\mu} \le \left(\frac{\exp(\delta)}{\exp\left((1+\delta)\cdot\frac{2\delta}{2+\delta}\right)}\right)^{\mu}$$

$$= \left(\exp\left(\delta - (1+\delta)\frac{2\delta}{2+\delta}\right)\right)^{\mu} = \exp\left(-\frac{\delta^2\mu}{2+\delta}\right)$$

Hoeffding's Inequality

Let $X_1, X_2, \ldots, X_n$ be independent RVs s.t. $X_i \in [a_i, b_i]$

Let $X := \sum_i X_i$, $\mu := \mathbb{E}[X]$

$$\Pr(|X-\mu| \geq t) \leq 2 \cdot \exp\left(-\frac{2t^2}{\sum_i (b_i - a_i)^2}\right)$$

Like Chernoff but additive/doesn't assume $X_i \in \{0,1\}$