Today

1) 3 Common Setups
   - Random Load Balancing
   - Coupon Collector
   - Birthday Paradox

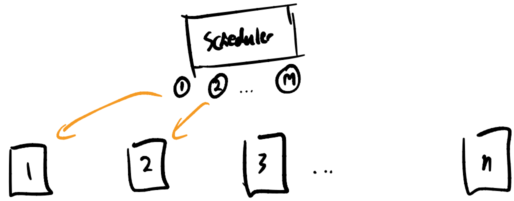
   $\rightarrow$ Balls $\rightarrow$ Bins

2) Concentration Bound Paradigm

3) Solving (1) w/ (2)

3 Scenarios

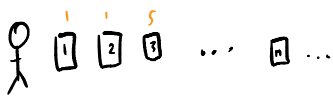
1) Random load balancing



Assume schedule jobs uniformly at random.

What's max # jobs on a machine?

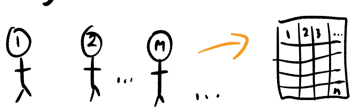
2) Coupons: 1, 2, ..., n



Assume each box contains a uniformly random coupon

How many cereal boxes until collect all coupons w/ good probability?

3) Birthday Paradox



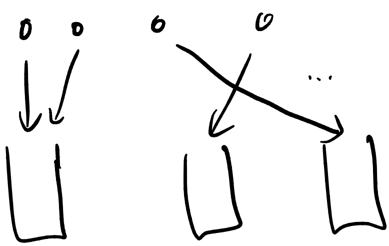
Assume bdays uniform

How many people until 2 people share a bday w/ good probability?

Balls into Bins

Really 3 Questions about 1 Scenario

m balls



n bins

each ind. and uniformly tossed

1) Max # balls in a bin (assuming $m=n$)

$\lesssim \ln n$ w/ $\Pr \geq 1 - \frac{1}{n}$

2) How big must m be until ≥ 1 ball in all bins?

$m \gtrsim n \cdot \ln n$ w/ $\Pr \geq 1 - \frac{1}{n}$

3) How big must m be until ≥ 2 balls in some bin?

$m \gtrsim \sqrt{n}$ gives $\Omega(1)$ probability

How to Solve Your Fave Randomization Question

To show fact (*)

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

b) Concentration: each RV at $\mathbb{E}$ $(\pm \log n)$ w/ good Probability

c) Union bound: all RVs " "

Let $X_{ij} :=$ 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 $m \geq \Omega(n \cdot \log n)$)

3) Call first $\frac{1}{2}$ of balls red, last $\frac{1}{2}$ balls blue

Let $Y_j := \begin{cases} 1 & \text{if } j\text{th blue ball into a bin w/ a red ball} \\ 0 & \text{o/w} \end{cases}$

Let $Y := \sum_j Y_j$

Let $R :=$ $\exists$ bin w/ ≥ 2 red

$$\Pr\left(\substack{\geq 2 \text{ balls} \\ \text{in a bin}}\right) \geq \Pr(R \cup Y \geq 1) = \Pr(R) + \Pr(\bar{R}) \cdot \underbrace{\Pr(Y \geq 1 \mid \bar{R})}_{\text{wts large}}$$
$$\geq \Pr(Y \geq 1 \mid \bar{R})$$

But $\mathbb{E}[Y \mid \bar{R}] = \sum_j \mathbb{E}[Y_j \mid \bar{R}] = \frac{m}{2} \cdot \frac{m/2}{n} = \Omega(1)$ if $m \gtrsim \sqrt{n}$

An ok Fact: Gaussians are "Concentrated"

Let $Z \sim N(0,1)$, then $\Pr(Z \geq |t|) \leq 2 \cdot \exp(-t^2/2)$ $\forall t \geq 0$

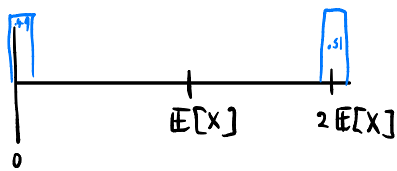
↳ will get similar bounds for general class of RVs

Markov's Inequality

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

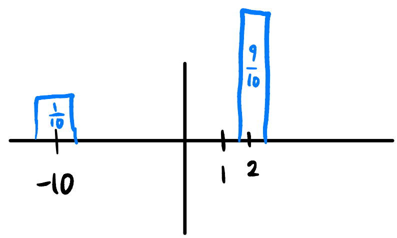
$\Pr(X \geq a \cdot \mathbb{E}[X]) \leq \frac{1}{a}$ ↗ $\geq$ twice $\mathbb{E} \leq \frac{1}{2}$ of the time $\forall a \geq 1$

Intuition



Why Non-Negative?

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



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

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

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

Aside: A function $f:\mathbb{R}\to\mathbb{R}$ is <u>non-decreasing</u> if $a \geq b \to f(a) \geq b$

For RV X have $\Pr(X \geq a) \leq \Pr(f(X) \geq f(a))$ for mono. inc. f

Corollary: Chebyshev's Inequality

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

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

$$\Pr(|X - \mathbb{E}[X]| \geq a) \overset{\text{mono}}{\leq} \Pr(Y \geq a^2)$$

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

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

All skipped in class

Chernoff Bound

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

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

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

$(1+\delta)\frac{1}{\delta} + \log\delta$

Chernoff Simplified

$$\Pr(X \geq (1+\delta)\cdot\mu) \leq \exp\left(-\mu\delta^2/(2+\delta)\right) \quad \forall\, \delta \geq 0$$

Have $\dfrac{2\delta}{2+\delta} \leq \log(1+\delta) \quad \forall\, \delta \geq 0 \overset{\delta = \delta - 1}{\longleftrightarrow} 2\delta \leq (1+\delta)\cdot\log\delta + 2 \quad \forall\, \delta \geq 1$

Convex and LHS tangent at $\delta = 1$ to RHS

So $$\frac{\exp(\delta)}{(1+\delta)^{1+\delta}} = \frac{\exp(\delta)}{\exp((1+\delta)\cdot\log(1+\delta))} \leq \frac{\exp(\delta)}{\exp\left((1+\delta)\cdot\frac{2\delta}{2+\delta}\right)}$$
$$= \exp\left(\delta - (1+\delta)\frac{2\delta}{2+\delta}\right) = \exp\left(-\frac{2\delta^2}{2+\delta}\right)$$

$\rightarrow$ If $\mu\cdot\delta \approx c\cdot\log n$, get at most $\approx \frac{1}{n^c}$ $\rightarrow$ "with high probability"

Lower Tail

$$\Pr(X \leq (1-\delta)\cdot\mu) \leq \exp(-\delta^2\mu/2) \quad \forall\, \delta \in (0,1)$$

Only did simplified in class

Solving Problems w/ Chernoff + Union Bound

1) Fix bin i.

independent

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

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

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

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

$$= n^{-2}$$

$\rightarrow$ So $X_i \geq 1+2\cdot\log n$ w/ $\Pr \leq n^{-2}$

$$\Pr(X_1 \geq 1+2\cdot\log n \cup X_2 \geq 1+2\cdot\log n \cup \ldots) \underset{\text{UB}}{\leq} \sum_i \Pr(X_i \geq 3\cdot\log n)$$

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

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

skipped

2) Let $m = 16n\cdot\ln n$, Fix i.

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

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

$$= n^{-2}$$

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

$E[X_i] = \Theta(\log n)$

so $X_i = \Theta(\log n)$ whp

so all $X_i = \Theta(\log n)$ whp

3) For blue ball j

Skipped

Let $M = 8\sqrt{n}$

Condition on $\bar{R}$

$$Y_j = \begin{cases} 1 & \text{w/ Pr } \frac{M/2}{n} = \frac{4}{\sqrt{n}} \\ 0 & \text{o/w} \end{cases}$$

All independent

Have $\mu = \mathbb{E}[Y \mid \bar{R}] = \frac{M}{2} \cdot \frac{4}{\sqrt{n}} = 16$

$$\Pr(Y=0 \mid \bar{R}) \le \Pr(Y \le \underbrace{(1-\tfrac{1}{2})}_{\delta} \cdot 16 \mid \bar{R})$$

$$\le \exp\left(-\frac{1}{4} \cdot 16/2\right)$$

$$\le \frac{1}{e^2}$$

$$\Pr\left(\substack{\ge 2 \text{ balls} \\ \text{in a bin}}\right) \ge \Pr(Y \ge 1 \mid \bar{R}) = 1 - \frac{1}{e^2}$$

$\mathbb{E}[Y \mid \bar{R}] = O(1)$ so $Y \ge 1$ w/ $\Omega(1)$ probability

Hoeffding's Inequality

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

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

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

→ Additive

→ Doesn't assume $X_i \in \{0,1\}$

Skipped

Chernoff Proof

← skipped in class

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

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

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

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

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

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

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

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

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