Concentration of Measure: Markov's Inequality, Chebyshev's Ineq, Hoeffding's Inequality

① Flip a coin n times, each with heads prob. p.
$\mathbb{E}[\#\text{heads}] = pn$, but how likely $\approx pn$? $[pn \pm \sqrt{n}]$

② Throw n balls in n bins, each random.
Max # balls in a bin? $[O(\log n)]$

③ How many unit vectors in $\mathbb{R}^n$ s.t. $|\langle v_i, v_j\rangle| \leq \varepsilon$ $\forall i,j$?

Law of Large Numbers says ① $\to pn$ as $n \to \infty$, but want

Non-Asymptotic Convergence: bound $\Pr[X \geq \mu + \lambda]$ $(\mu = \mathbb{E}X)$
$\Pr[X \leq \mu + \lambda]$
as function of n.

Markov's Inequality: If X is a nonnegative random variable, and $\lambda > 0$, then

$$\Pr[X \geq \lambda] \leq \frac{\mathbb{E}X}{\lambda}.$$

Proof: $$\mathbb{E}X = \underbrace{\mathbb{E}[X \mid X < \lambda]}_{\geq 0}\Pr[X<\lambda] + \underbrace{\mathbb{E}[X \mid X \geq \lambda]}_{\geq \lambda}\Pr[X \geq \lambda]$$

$$\geq \lambda \Pr[X \geq \lambda]$$

Def (Variance): For r.v. X with mean μ, the variance
$$\sigma^2 = \mathbb{E}(X-\mu)^2.$$
If X, Y indep. then

$\sigma^2 = \mathbb{E}(X-\mu)$.

If X, Y indep, then

$$\sigma^2_{X+Y} = \mathbb{E}\left[\left((X+Y)-(\mu_X+\mu_Y)\right)^2\right]$$
$$= \mathbb{E}\left[\left((X-\mu_X)^2+(Y-\mu_Y)\right)^2\right]$$
$$= \mathbb{E}(X-\mu_X)^2 + \mathbb{E}(Y-\mu_Y)^2 + \underbrace{\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]}_{= \mathbb{E}[X-\mu_X]\,\mathbb{E}[Y-\mu_Y]}$$
$$= \sigma_X^2 + \sigma_Y^2.$$

$$\mathbb{E}[X-\mu_X]\,\mathbb{E}[Y-\mu_Y] = \underbrace{\mathbb{E}[X-\mu_X]}_{=0}\;\underbrace{\mathbb{E}[Y-\mu_Y]}_{=0}$$

Chebyshev's Ineq: $\Pr[|X-\mu| \geq \lambda] \leq \frac{\sigma^2}{\lambda^2}$.

Proof: $\Pr[|X-\mu|\geq\lambda] = \Pr[(X-\mu)^2 \geq \lambda^2]$

Markov on $Y=(X-\mu)^2$ and λ^2:

$$\Pr[Y \geq \lambda^2] \leq \frac{\mathbb{E}Y}{\lambda^2} = \frac{\sigma^2}{\lambda^2}.$$

① Let $S_n = X_1 + \cdots + X_n$, $X_i = \mathbb{1}(i^{th}\text{ flip heads})$.

$\sigma_i = \mathbb{E}(X_i - \mu_i) = \text{constant}$

Since X_i are indep., $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 = O(n)$

Want $\Pr[|S_n - pn| \geq \lambda] \leq \varepsilon$: $\quad \frac{\sigma^2}{\lambda^2} = \varepsilon \Rightarrow \lambda \sim \sqrt{n/\varepsilon}$.

$\varepsilon = \frac{1}{\text{poly}(n)}$? No good...

<u>Hoeffding's Inequality</u>: Let $X_1, \ldots, X_n$ be n indep. r.v. taking values in $[0,1]$.

Let $S_n = X_1 + \cdots + X_n$ with mean $\mu = \mathbb{E}S_n = \sum \mathbb{E}X_i$. Then for any

$\varepsilon \geq 0$ we have

(Upper tail) $\Pr[S_n \geq (1+\delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right)$

(Lower tail) $\Pr[S_n \leq (1-\delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{3}\right)$.

Applizations to our problems:

① $\frac{\delta^2\mu}{3} \sim \log n \iff \delta \sim \sqrt{\log n / n}$

So $S_n \in \mu \pm \delta\mu = pn \pm O(\sqrt{n \log n})$ with prob $1 - \frac{1}{\mathrm{poly}(n)}$

② Each bin is Binomial$(n, 1/n)$, $\mu = 1$.

$\Pr[\#\text{balls} \geq 1+\delta] \leq \exp\left(-\frac{\delta^2}{2+\delta}\right)$, take $\beta = \Theta(\log n) \rightarrow$ prob $1 - \frac{1}{\mathrm{poly}(n)}$

Correct answer $\Theta\left(\frac{\log n}{\log\log n}\right)$.

Proof: Idea: Apply Markov to appropriate function of S_n.

Consider e^{tS_n} for t to be optimized later.

$$\Pr[S_n \geq (1+\delta)\mu] = \Pr\left[e^{tS_n} \geq e^{t(1+\delta)\mu}\right]$$

$$\leq \frac{\mathbb{E}\, e^{tS_n}}{e^{t(1+\delta)\mu}} \quad \text{(Markov)}$$

$$= \frac{\prod_i \mathbb{E}\, e^{tX_i}}{e^{t(1+\delta)\mu}} \quad \text{(indep.)}$$

Suppose first $X_i \in \{0,1\}$. Then

$$\mathbb{E}\, e^{tX_i} = (1-\mu)\cdot 1 + \mu \cdot e^t = 1 + \mu_i(e^t - 1)$$

$$\leq e^{\mu_i(e^t-1)}$$

$$= \quad \mu_i(e^t - 1)$$

$$= \frac{\prod_i e^{\mu_i(e^t-1)}}{e^{t(1+\delta)\mu}}$$

$$= \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}}$$

$$= \exp\left(\mu(e^t-1) - t(1+\delta)\mu\right).$$

Optimize for t: set derivative of $=0$: $t = \ln(1+\delta)$

$$= \exp\left(\mu\delta - \ln(1+\delta)(1+\delta)\mu\right)$$

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

Next, use $\frac{\delta}{1+\delta/2} \le \ln(1+\delta)$

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

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

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

Removing the assumption:

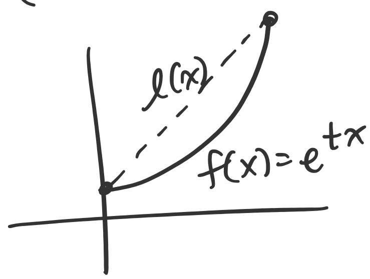
Define $Y_i \sim \text{Bernoulli}(\mu_i)$: $\mathbb{E} X_i = \mathbb{E} Y_i$

Removing [illegible]

Define $Y_i \sim \text{Bernoulli}(\mu_i)$: $\quad \mathbb{E} X_i = \mathbb{E} Y_i$

Claim: $\mathbb{E}\, e^{tX_i} \leq \mathbb{E}\, e^{tY_i}$.

Proof (convexity of e^{tx})



$f(x) \leq \ell(x)$ for $x \in [0,1]$

$$
\begin{aligned}
\mathbb{E}\, e^{tX_i} = \mathbb{E} f(X_i) &\leq \mathbb{E}\, \ell(X_i) \\
&= \ell(\mathbb{E} X_i) \quad \text{(linearity of } \mathbb{E}\text{)} \\
&= \ell(\mathbb{E} Y_i) \\
&= \mathbb{E}\, \ell(Y_i) \\
&= \mathbb{E}\, e^{tY_i} \quad (\ell(x) = f(x) \text{ on support of } Y_i : \{0,1\})
\end{aligned}
$$

Summary:

1. Markov's Ineq on "optimized" function e^{tX}
2. $\mathbb{E}\, e^{tS_n} = \prod e^{tX_i}$ by indep.
3. Replace X_i by Bernoulli Y_i "worst-case"
4. algebra, optimize for t