3. Concentration inequalities

Inequalities we saw in our prev lectures:

1. Markov Inequality

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$ First moment bound

2. Chebyshev's Inequality

$$\Pr\big[|X - \mathbb{E}[X]| \geq t\big] \leq \frac{\mathrm{Var}[X]}{t^2}.$$ Second central moment bound.

$\Pr[X = i] = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$ | $\sum_{i=0}^{n} \Pr[X=i] = 1.$

Some definitions before we prove stronger bounds.

Let $X_1, \dots, X_n$ are independent Bernoulli trials with $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1-p$ (for all $i \in [1,n]$). Let $X = \sum_{i=1}^{n} X_i$. Then X is said to have the binomial distribution.

Ex: $X_1, \dots, X_n$ could be indicators of HEADS in n independent coin tosses. X is a r.v corresponding to the no. of heads.

More generally, let $X_1, \dots, X_n$ be independent trials s.t $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$ (for all $i \in [n]$). Such trials are called Poisson trials.

Ex: $X = \sum_{i=1}^{n} X_i$ is a "sum of Poisson trials".

3.1 Chernoff Bound

Theorem: Let $X_1, \dots, X_n$ be independent Poisson trials s.t for all $1 \le i \le n$; $\Pr[X_i = 1] = p_i$ where $0 < p_i < 1$. Then, for $X = \sum X_i$ and $\delta > 0$,

$$\Pr[X > (1+\delta) E[X]] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{E[X]}$$

and

$$\Pr[X < (1-\delta) E[X]] < \left[\frac{e^{\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu} \le e^{-\frac{E[X] \cdot \delta^2}{2}}.$$

$d > 0$, A'_d s.t $\left.\begin{matrix} \\ \ge 1 - \frac{1}{2^{n^d}} \end{matrix}\right\}$

3.1.1 Application: Error reduction.

Theorem: Let A be a randomized algorithm for task T such that A runs in polynomial time and succeeds w.p $\ge \frac{2}{3}$. Then there exists another randomized algorithm A' for T s.t it runs in polynomial time and succeeds w/p $\ge 1 - 2^{-n^{O(1)}}$ (where n is input size to A).

Proof: For some parameter s that we shall soon fix; algo A' runs A s many times (independently) and outputs the solution of the majority.

Let X_i be an indicator random variable for success of algorithm A in i^{th} trial.

$$E[X_i] = \Pr[X_i = 1] \ge \frac{2}{3}.$$

Let $X = \sum x_i$ indicate the no. of successful trials. Thus, we are interested in $\Pr[X < \frac{s}{2}]$ being small.

$$\Pr\left[X < \frac{s}{2}\right] = \Pr\left[X < (1-\delta)\mathbb{E}[X]\right] \quad \text{where } \mathbb{E}[X] = \frac{2s}{3} \text{ and}$$

$$\leq e^{-\frac{\mathbb{E}[X]\cdot\delta^2}{2}}$$

$$< \exp\left(-\frac{2s}{3}\cdot\frac{\delta^2}{2}\right)$$

$$= \exp(-\Omega(s)).$$

Runtime of A'

$= s \times$ runtime of A

$= n^d \cdot n^{O(1)}$ where $d = \Theta(1)$.

$= n^{O(1)}$

Ex: X could be the no. of heads in n independent trials.

Set s to n^2 and error probability is at most $\frac{1}{2^{\Omega(n^2)}}$.

Remark: This is called Error reduction of randomized algorithms.

Proof of Chernoff bound:

$X = \sum_{i=1}^{n} x_i$ where x_i's are Poisson trials.

For any $t \in \mathbb{R}_{\geq 0}$ and $\mu = \mathbb{E}[X]$,

$$\Pr[X > (1+\delta)\mu] = \Pr\left[e^{tX} > e^{t(1+\delta)\mu}\right].$$

By applying Markov's ineq:

$$\Pr[X > (1+\delta)\mu] < \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}}.$$

Recall that $e^{tX} = \sum_{k\geq 0} \frac{t^k X^k}{k!}$; $\mathbb{E}[e^{tX}] = \sum_{k\geq 0} \frac{t^k}{k!}\cdot \mathbb{E}[X^k]$

And $\mathbb{E}[X^k]$ is the k^{th} moment. $\mathbb{E}\left[e^{t\left(\sum_{i=1}^{n} X_i\right)}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}]$

$$\mathbb{E}[e^{tX}] = \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] \text{ and } \mathbb{E}[e^{tX_i}] = p_i \cdot e^t + (1-p_i).$$

X_i's are independent.

$$= \prod_{i=1}^{n} \left(1 + p_i(e^t - 1)\right)$$

$$< \prod_{i=1}^{n} \exp\left(p_i(e^t-1)\right) \quad // \ 1+x < e^x$$

$$= \exp\left(\mu \cdot (e^t - 1)\right) \quad // \ \sum p_i = \mathbb{E}[X] = \mu.$$

$\Pr[X \geq 5\,\mathbb{E}[X]] \leq \frac{1}{5}$, $\leq \frac{1}{n^2_{3n}}$
$\delta = 4$
$\leq \exp(-\mu)$
$\leq \frac{1}{2^{\Omega(n \log n)}}$

Thus, $\Pr[X > (1+\delta)\mu] < \dfrac{\exp\left(\mu \cdot (e^{t} - 1)\right)}{\exp\left(t \cdot (1+\delta) \cdot \mu\right)}$ $\quad e^{\ln(1+\delta)} = (1+\delta)$, -1

This is a great bound if $(1+\delta)^{\delta} \gg e^{\delta}$. $\delta \geq 4.3$ ✓

t was some parameter. Note that this minimizes for $t \approx \ln(1+\delta)$.

$$< \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}. \checkmark$$

$\exp\left((1+\delta)\cdot \ln(1+\delta)\mu\right) \to (1+\delta)^{(1+\delta)\mu}$

Proof for $\Pr[X < (1-\delta)\mu]$ is similar.

Remark: • We studied r.v e^{tX} instead of X

• Independence of X_i's was used.

• Optimized for the best value of t wrt δ.

More specific bounds: [Mitzenmacher-Upfal] Section 4.3

Theorem 4.4: *Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr(X_i) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbf{E}[X]$. Then the following Chernoff bounds hold:*

1. *for any $\delta > 0$,*

$$\Pr(X \geq (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}; \tag{4.1}$$

2. *for $0 < \delta \leq 1$,*

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}; \tag{4.2}$$

3. *for $R \geq 6\mu$,*

$$\Pr(X \geq R) \leq 2^{-R}. \tag{4.3}$$

[Dubashi Panaconi]

3.1.2 Set balancing

c_i is a measure of difference of allocation of those w/ feature i

$|c_i|$ = absolute value of c_i.

$\therefore$ should not be very large.

Given an $n \times m$ matrix $\mathbf{A}$ with entries in $\{0, 1\}$, let

$$c_i = \sum_{j=1}^{m} a_{ij} b_j = \sum_{j \in K_i} a_{ij} b_j$$

$t_1 = 1$, $t_2 = -1$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

Suppose that we are looking for a vector $\bar{b}$ with entries in $\{-1, 1\}$ that minimizes

$$\|\mathbf{A}\bar{b}\|_\infty = \max_{i=1,\ldots,n} |c_i|.$$

Experiment should assign $\frac{|K_i|}{2}$ to 1 and $\frac{|K_i|}{2}$ to -1. $\rightarrow$ $c_i = (t_1 - t_2) = d_i$

Pool: -1 placebo (T_1) | 1 vaccine (T_2)

This problem comes from designing statistical experi-ments.

Theorem 4.11: *For a random vector $\bar{b}$ with entries chosen independently and with equal probability from the set $\{-1, 1\}$,*

$$\Pr(\|\mathbf{A}\bar{b}\|_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}.$$

$|X - E[X]| > \delta E[X]$

or $X > (1+\delta)E[X]$

e^{tX}

Proof: Let $\bar{a}_i = (a_{i1}, \ldots, a_{in})$ be the ith row. If $\bar{a}_i$ has at most $\sqrt{4m \ln n}$ 1's; $|\langle \bar{a}_i, \bar{b} \rangle|$ can never be more than $\sqrt{4m \ln n}$. Else, $\bar{a}_i$ has at least $\sqrt{4m \ln n}$ 1's.

Let Z_i be a r.v defined to be $\langle \bar{a}_i, \bar{b} \rangle$; $Z_i = \sum_{j \in K_i} a_{ij} \cdot b_j$.

$K_i = \{ j \mid a_{ij} \neq 0 \}$

Uses the fact that $\mathbb{E}[e^{tX_i}] \leq e^{t^2/2}$

Corollary 4.8: *Let $X_1, \ldots, X_n$ be independent random variables with*

$$\Pr(X_i = 1) = \Pr(X_i = -1) = \frac{1}{2}.$$

Let $X = \sum_{i=1}^{n} X_i$. Then, for any $a > 0$,

$$\Pr(|X| \geq a) \leq 2e^{-a^2/2n}.$$

$$\Pr[Z_i \geq \sqrt{4m \ln n}] \leq 2e^{-\left(\frac{4\cancel{m} \ln n}{2\cancel{m}}\right)}$$

$$\leq 2e^{-\frac{L^2}{2m}}. \qquad = 2 \cdot n^{-2}.$$

$c_i \geq \sqrt{4m \ln n}$ w prob at most $\frac{2}{n^2}$

That is $\underline{c_i} \leq \sqrt{4m \ln n}$ w.p $1 - \frac{2}{n^2}$.

By union bound, any row fails to attain the bound is with a probability of at most $\frac{2}{n}$. $\left(1 - \frac{2}{n}\right)$

W. prob of $\geq \left(1 - \frac{2}{n}\right)$ each c_i takes a value atmost $\sqrt{4m \ln n}$

$\left(1 - 2n \cdot e^{-\frac{L^2}{2m}}\right)$ each c_i takes a value of at most L.

$$\| Ab \|_\infty \leq L.$$