

3. Concentration inequalities

Inequalities we saw in our prev lectures:

1. Markov Inequality

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

2. Chebyshев's Inequality

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

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

Some definitions before we prove stronger bounds.

Let X_1, \dots, X_n be 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 $1 \leq i \leq n$, $\Pr[X_i=1] = p_i$ 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

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

$$\left. \begin{array}{l} d > 0, \\ A' \text{ s.t. } \\ \geq 1 - \frac{1}{2^{nd}} \end{array} \right\}$$

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 $\geq \frac{2}{3}$. Then there exists another randomized algorithm A' for T s.t it runs in polynomial time and succeeds w/p $\geq 1 - 2^{-n^{O(1)}}$ (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.

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

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

$$\Pr[X < \frac{S}{2}] = \Pr[X < (1-\delta)\mathbb{E}[X]] \quad \text{where} \\ \downarrow \quad \mathbb{E}[X] = \frac{2S}{3} \quad \text{and} \\ \leq e^{-\frac{\mathbb{E}[X]\cdot\delta^2}{2}}$$

Runtime of A'

$$= S \times \text{Runtime of A} < \exp\left(-\frac{2S}{3} \cdot \frac{\delta^2}{2}\right) \\ = n^d \cdot n^{O(1)} \quad \text{where } d = \Theta(1). \\ = n^{O(1)} \quad = \exp(-\Omega(S)).$$

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

Set S to n^2 and error probability is at most $\frac{1}{2^{cn^2}}$.

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 $\mu = \mathbb{E}[X]$,

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

By applying Markov's ineq;

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

$$\text{Recall that } e^{tx} = \sum_{k \geq 0} t^k \frac{x^k}{k!}; \quad \mathbb{E}[e^{tx}] = \sum_{k \geq 0} t^k \frac{\mathbb{E}[x^k]}{k!}.$$

And $E[X^k]$ is the k^{th} moment. $E[e^{t(\sum_{i=1}^n x_i)}] = \prod_{i=1}^n E[e^{tx_i}]$

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

x_i 's are independent.

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

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

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

$$\Pr[x \geq 5\bar{x}] \leq \frac{1}{n^{\delta}} \stackrel{\delta=4}{\leq} \frac{1}{n^4} \stackrel{\leq \frac{1}{5}}{\leq} \exp(-\mu) \stackrel{\leq \frac{1}{2^{(n+1)}}}{\leq}$$

Thus, $\Pr[x > (1+\delta)\mu] < \frac{\exp(\mu \cdot (e^t - 1))}{\exp(t \cdot (1+\delta) \cdot \mu)}$

$e^{\ln(1+\delta)} = (1+\delta)^{-1}$

$\frac{(1+\delta)^{\delta}}{\delta} \gg e$.

$\delta \geq 4.3$ ✓

This is a great bound if t was some parameter.
Note that this minimizes for $t \approx \ln(1+\delta)$.

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

$$\frac{\exp((1-\delta) \cdot \ln(1-\delta))\mu}{(1-\delta)^{(1-\delta)\mu}}$$

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, \dots, 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 $\delta > 0$,

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

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

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

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

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

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 .

Should not be very large.

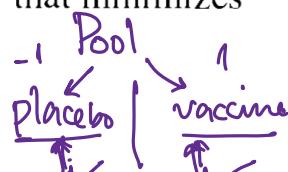
Given an $n \times m$ matrix \mathbf{A} with entries in $\{0, 1\}$, let

$$C_i = \sum_{j=1}^m a_{ij} b_j \quad f_i \begin{pmatrix} t_1 & t_2 \\ a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

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

{ Experiment should assign $|k_{il}|$ to 1 and $|k_{il}|$ to -1. } $\rightarrow C_i = (t_1 - t_2) \sqrt{d_i}$

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



This problem comes from designing statistical experiments.

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

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

$$\begin{aligned} |\bar{x} - \mathbb{E}[x]| &> \delta \mathbb{E}[x] \\ \text{or } \bar{x} &> (1 + \delta) \mathbb{E}[x] \\ e^{\bar{x}} & \end{aligned}$$

Proof: Let $\bar{a}_i = (a_{i1}, \dots, a_{in})$ be the i th 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
 $E[e^{tX_i}] \leq e^{t^2/2}$

Corollary 4.8: Let X_1, \dots, 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 $a > 0$,

$$\Pr(|X| \geq a) \leq 2e^{-a^2/2n}. \quad \boxed{3}$$

$$\begin{aligned} \Pr[Z_i \geq \sqrt{4m \ln n}] &\leq 2e^{-\left(\frac{4m \ln n}{2}\right)} \\ &\leq 2e^{-\frac{L^2}{2m}}. \quad = 2 \cdot n^{-2}. \quad \text{That is } c_i \leq \frac{\sqrt{4m \ln n}}{1 - \frac{2}{n^2}}. \end{aligned}$$

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

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

W. prob of $\geq (1 - \frac{2}{n})$ each c_i takes a value atmost

$$\boxed{\sqrt{4m \ln n}}$$

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

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