Lecture 10. Hoeffding Bounds.

Example 2

X: # of wins in a seasons of n games $X \sim Bin(n, \frac{3}{4})$

$\mathbb{E}[X] = \frac{3}{4} n.$

chernoff lower bound. $Pr[X \le \frac{n}{2}] = Pr[X \le (1 - \frac{1}{3}) \frac{3}{4} n]$

$\le e^{-\frac{3n}{4} \cdot \frac{1}{3^2} \cdot \frac{1}{2}}$

$= e^{-\frac{1}{24} n} \simeq 0.96^n$

Example 3

$Y_1 \sim Bin(n, \frac{1}{n})$, $\mathbb{E}[Y_1] = 1$

$Pr[Y_1 > m] = Pr[Y_1 \ge (m+1)]$ $[\delta = m \cancel{?} 1]$

$\le \frac{e^m}{(1+m)^{1+m}} \le \frac{1}{n^2}$ $[(\frac{e^\delta}{(1+\delta)^{1+\delta}})^\mu]$

$m - \cancel{(1+m)^{1+m}} (1+m) \ln (1+m) \le -2 \ln n.$

$(1+m) \ln (1+m) - m \ge 2 \ln n.$

$m \ln m = \Theta(\ln n).$

$m = \Theta(\frac{\ln n}{\ln \ln n})$

Hoeffding Bounds

(upper tail) $Pr[X \ge \mu_0 n + \epsilon n] \le e^{-2n\epsilon^2/(b-a)^2}.$

(lower tail) $Pr[X \le \mu_0 n - \epsilon n] \le e^{-2n\epsilon^2/(b-a)^2}$

Example

p feature. taking n samples.

$X \sim Bin(n, p)$ $X_i = \begin{cases} 0 & o/w \\ 1 & \text{sample } i \text{ has feature} \end{cases}$

$X = \tilde{p} n$ ($\tilde{p}$ empirical average).

$1-\gamma$ confidence interval

$$\Pr[p \in [\tilde{p}-\delta, \tilde{p}+\delta]] \geq 1-\gamma$$

γ, δ, n tradeoff.

~~$\mathbb{E}[Xp]$~~

$\mathbb{E}[X] = np$ $\qquad X = n\tilde{p}$

$p \notin [\tilde{p}-\delta, \tilde{p}+\delta]$.

① $p < \tilde{p}-\delta \Rightarrow \tilde{p} > p+\delta$, $X = n\tilde{p} \geq np+\delta n$.

② $p > \tilde{p}+\delta \Rightarrow \tilde{p} < p+\delta$, $X = n\tilde{p} \leq np+\delta n$.

$$\Pr[X > n(p+\delta)] = \Pr[X > np+n\delta] \leq e^{-2n\delta^2}$$

$$\Pr[X < n(p-\delta)] = \Pr[X < np-n\delta] \leq e^{-2n\delta^2}$$

fail prob: $\boxed{\gamma = 2e^{-2n\delta^2}}$: trade-off. γ, n, δ
from Hoeffding bounds.

fixed error rate. $n = \Theta(\frac{1}{\delta^2})$
$\delta = \Theta(\frac{1}{\sqrt{n}})$

$\gamma = n^{-\epsilon}$, want $\delta = \Theta(\sqrt{\frac{\ln n}{n}})$.

Examples

$$\|(X_1, \ldots X_n)\|_\infty = \max_{i \in [n]} |X_i|$$

$$\underset{0,1}{A} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ & & \ddots & \\ & & & a_{nm} \end{bmatrix}}_{m \text{ subjects}} \Big\} n \text{ features.}$$

Find $b \in \{-1, 1\}^m$, $\min \|Ab\|_\infty$

two groups. how well you balance each feature?

Algorithm

~~For~~ Pick each b_i u.a.r. from $\{-1,1\}$.

$$\Pr[\|A\cdot b\|_\infty \geq \sqrt{4m\ln n}\,] \leq \frac{2}{n}$$

m "people"
n "features"

Proof: consider row i

1 0 1 0 0 0 1 0 1 1

Let k be the number of 1's in row i

If $k \leq \sqrt{4m\ln n}$, then $|\bar{a}_i \cdot b| \leq \sqrt{4m\ln(n)}$.

Suppose $k > \sqrt{4m\ln n}$

$$Z_i = \sum_{j\in[m]} \underbrace{a_{ij} b_j}_{k \text{ nonzero R.V.'s}} \quad \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$$

$\mathbb{E}[Z_i]=0$

$\left(\varepsilon^2 = \frac{4m\ln n}{k^2}\right)$

$(k^2 > 4m\ln n)$.

By Hoeffding:

$$\Pr[\,|Z_i| \geq \varepsilon k\,] \leq 2e^{-\frac{2\varepsilon^2 k}{4}}$$

$$= 2e^{-\frac{2m\ln n}{k}}$$

$$= 2\,\cancel{e}\,n^{-\frac{2m}{k}}$$

Want $\leq \frac{2}{n^2}$ if $\boxed{k \leq m}$

By union bound over all rows:

$$\Pr[\max |Z_i| \geq \ldots] \leq n\cdot\frac{2}{n^2} = \frac{2}{n}.$$