Chernoff Bounds

$X_1, X_2 \ldots X_n$ are identical independent Bernoulli r.v. w/ param p.

for every $\varepsilon > 0$, $P[|\bar{X}_n - p| > \varepsilon p] \le e^{-\frac{\varepsilon^2 np}{3}}$

note that bound is on a multiplicative factor of p, not an absolute error

ex. n computers sending to m routers

we need to distribute packets as evenly as possible btwn routers.

If each packet chooses randomly, with very high probability no router will see more than $(1+\varepsilon)\frac{n}{m}$ packets.

let X = # packets going through first router

$Pr[X > (1+\varepsilon)\frac{n}{m}] \le Pr[|X - \frac{n}{m}| < \varepsilon \frac{n}{m}]$

one sided — two sided w/ abs value, so inequality must be true

Union bound

$Pr[\bigcup_{i=1}^{m} A_i] \le \sum_{i=1}^{m} Pr[A_i]$

the probability of a union of events can be no more than the sum of each indiv probability.

ex. algorithm $A(x)$ receives some input, trying to compute some function $f(x) = y$ (input)

$Pr[A(x)$ outputs correct answer$] > \frac{2}{3}$

algorithm A uses randomness to help solve problem (can't be solved analytically, efficiently)

but 2/3 sucess seems not good enough

however, we can amplify the sucess of $A(x) \rightarrow$ reduce error

$\bar{A}(x)$: run $A(x)$ n times (n is large) independently, w/ diff randomness

take the majority of results $\rightarrow P[$sucess$]$ is now huge

we invest more running time, but we gain in accuracy ex. Algorithm continued..

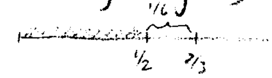
when will $\bar{A}_n(x)$ output wrong answer?

define $Y_i = \begin{cases} 1 & \text{if } i^{th} \text{ run of } A \text{ gives correct answer} \\ 0 & \text{otherwise} \end{cases}$ Bernoulli

$Y = \sum_{i=1}^{n} Y_i$ = number of correct answers for n trials

$\therefore$ wrong answer when $\sum_{i=1}^{n} Y_i < \frac{1}{2} n$ since taking majority

$\frac{Y}{n} = \frac{\sum Y_i}{n} < \frac{1}{2}$



$Pr[|\bar{Y}_n - \frac{1}{2}| > \frac{1}{6}] = Pr[|\bar{Y}_n - \frac{1}{2}| > \frac{1}{4} \cdot \frac{2}{3}]$

$Pr[|\bar{Y}_n - \frac{1}{2}| > \frac{1}{4} \cdot \frac{2}{3}] < e^{-\frac{0.25^2 \cdot n \cdot 2/3}{3}}$ very small

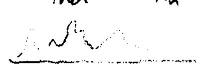
we are looking at probability you are more than $1/6$ away from $2/3$

which approximates the probability $\bar{Y}_n$ is less than $\frac{1}{2}$

Central Limit Theorem

Informal: $X_1\ X_2 \dots X_n$ are identical independent r.v.

then $\bar{X}_n = \frac{\sum X_i}{n}$ tends to a normal distribution



no matter how crazy X_i's distribution, if you take enough samples, avg will tend to smooth out

$X_1\ X_2 \dots X_n$ are independent identically distributed r.v w/ μ and σ^2

$Z_n = \frac{\bar{X}_n - \mu}{\sigma \cdot \frac{1}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ transformation of $\bar{X}_n$

if $\bar{X}_n$ normal, Z_n is standard normal

$= \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{Var[\bar{X}_n]}}$ b/c $Var[\bar{X}_n] = \frac{\sigma^2}{n}$

Def $X_1\ X_2 \dots X_n$ are i.i.d. r.v w/ expectation μ variance σ^2

Let $\bar{X}_n = \frac{\sum X_i}{n}$ $\quad Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$

for every $-\infty < a < \infty$

$\lim_{n \to \infty} F_Z(a) = \Phi(a)$ ← dist function of a std normal variable

↑ distribution function of Z

ex. 100 questions on exam, each has 4 possible answers

let X = # correct answers Binomial w/ $n=100$ $p=\frac{1}{4}$

$P[X > 30]$? difficult to compute directly even though we know it's binomial

instead, $X = R_1 + R_2 + \ldots R_{100}$ $R_i = \begin{cases} 1 & \text{if } i^{th} \text{ answer correct} \\ 0 & \text{otherwise} \end{cases}$

$P[X > 30] = Pr[\sum_{i}^{100} R_i > 30] = Pr[\sum_{i=1}^{100} \frac{R_i}{n} > \frac{30}{n}]$

avg of R_is $= \bar{R}$.

Standardize R: $Pr\left[\frac{\sqrt{100}(\bar{R} - 0.25)}{\sqrt{3/16}} > \frac{\sqrt{100}(30/n - 0.25)}{\sqrt{3/16}}\right]$

note: $\mu\ R_i = \frac{1}{4}$ $\sigma^2 = 3/16$ for bernoulli r.v.

$P[X>30] \approx Pr[Z > 1.154]$ if $n=100$ is large enough to make $\bar{R}$ normal

$P[X>30] \approx 1 - \Phi[1.154]$

this is much better than Chebychev or Chernoff b/c not a bound

provided n is large enough (100 is good rule of thumb)