Class 2: Jan 8'21

SHREE
DATE: / /
PAGE NO.:

Tail inequalities

we look at — 3 ways to estimate the tail prob. of RV. — depending on what we know about RV
— more we know about RV — better estimates

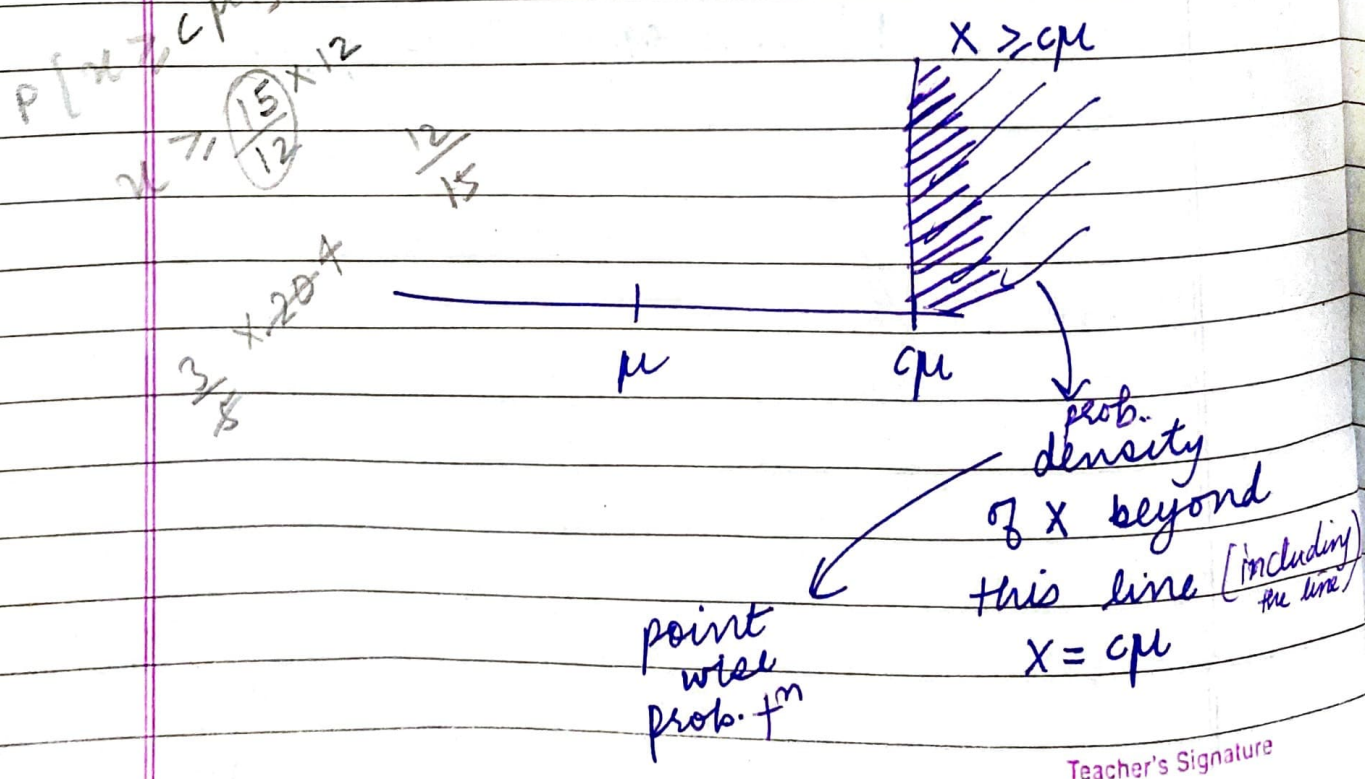
→ what is tail of distribut^n/ set?
└ values far away from mean

① Markov Inequality :

X = non-negative valued RV with Exp(X) = μ (finite expectat^n)
(Exp = mean)

↓
also cl first moment of RV.

we can the following question using this tail inequality &
↳ Pr[X ≥ cμ] = ?
(values c times away from mean)



Teacher's Signature

here c cannot be -ve.

$P[X \geq c\mu] \leq 1/c$

Take $c = 1$

$P[X \geq \mu] \leq 1 \Rightarrow$

if $Exp[X] = \mu$ then RV takes at least one value that is μ or greater

PROOF:

assumptn: X is a discrete RV.
(if not discrete use $\int$ instead of Σ)

we know,

$$\mu = \sum_a a \, Pr(X = a)$$

$$= \sum_{a < c\mu} a \, Pr(X = a) + \sum_{a \geq c\mu} a \, Pr(X = a)$$

we don't need this part

→ since this is +ve/non-ve we replace = with ≥

① if X is non-ve
⇒ every a is at least zero.

$$\mu \geq 0 + \sum_{a \geq c\mu} a \, Pr(X = a)$$

② because a is taking value at least $c\mu$
so, $a = c\mu$ will still hold the inequality.

$$\mu \geq \sum_{a \geq c\mu} c\mu \, P(X = a)$$

$$\mu \geq c\mu \sum_{a \geq c\mu} P(X = a)$$

$$\mu \geq c\mu \cdot Pr(X \geq c\mu)$$

$$\frac{\mu}{c\mu} \geq Pr(X \geq c\mu)$$

$$1/c \geq Pr(X \geq c\mu)$$

Teacher's Signature

So, Markov Inequality states that if a non-negative value RV with an finite expectation of μ, then $Pr[X \geq c\mu] \leq 1/c$

Applying M.I. on randomised quicksort algorithm
⇒ that the algorithm has a runtime of more than twice its expectat^n with Pr. of 1/2.
i.e. $Pr[T(n) \geq 2(2n \ln n)] \leq 1/2$

So, $Pr\left(T(n) \geq 1/2\, n^2\right)$

$\Rightarrow Pr\left(T(n) \geq \frac{1}{2} n^2 \frac{E(T(n))}{E(T(n))}\right)$

$\Rightarrow Pr\left(T(n) \geq \frac{1}{2E(T(n))} n^2 \times E(T(n))\right]$

$\Rightarrow Pr\left(T(n) \geq \frac{1}{2} n^2\right) \leq \frac{2E(T(n))}{n^2}$

$\leq \frac{2(2) n \ln n}{n^2}$

$\leq \frac{\ln n}{n}$

⇒ as n ↑es the prob. that quicksort takes quadratic time is very small — closer to zero

SHREE
DATE / /
PAGE NO.:

(2) Chebychev Inequality:

Here, we use the second moment of RV.
i.e. $E[X^2]$.
and the related quantity is variance.

Variance & standard variance

RV - X
$E[X] - \mu$
$var(X) = E[(X-\mu)^2]$

(we no longer insist that x be a non-neg. value)

we call it expectat" because when we divide the sum of squares by 'n' — we assume that every one of these is equally likely, that's why $1/n$ comes.
But here every one of these items will be taken by the RV with certain prob. So, ∴ we take a weighted sum of the squares of differences from the expectat". The weights are the probabilities of the differences.

∴ $var(X) = E[[X-\mu]^2]$

and we take squared sum to consider deviat" to the left of the expectat" & right of exp in similar way.

Standard deviat" $\sigma_x = \sqrt{var(X)}$

Now, $var(X) = E[(X-\mu)^2]$
$= E[X^2 - 2X\mu + \mu^2]$
following linearity of expectat"
$= E[X^2] - [E[X]]^2$

Teacher's Signature

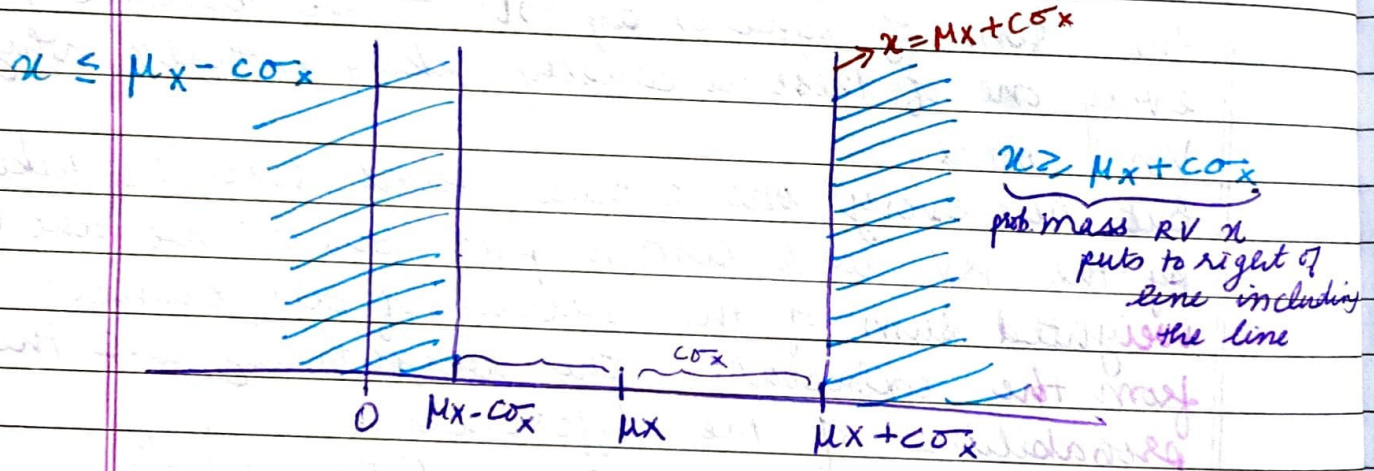
$$= E[X^2] - 2E[X\mu] + E[\mu^2]$$
$$= E[X^2] - 2\mu^2 + \mu^2$$
$$= E[X^2] - \mu^2$$
$$= \begin{pmatrix}\text{second}\\ \text{moment}\\ \text{of } X\end{pmatrix} - (\text{exp. of } X)^2$$

Chebychev Inequality

RV = X

exp = μ_X

St. de = σ_X



So, we are broadly interested in the combined mass of the two sides — and that event is captured by —

$$Pr\left([|X - \mu_X| \geq c\sigma_X]\right)$$

So this the question answered by chebychev inequality

Teacher's Signature

[To be able to apply Chebychev to randomised quicksort, we need the variance of the RV that corresponds to the run time of the algo. & we didn't calculate it. SHREE DATE: / / PAGE NO.: So we wont apply CC directly on that eg.]

Chebychev states that -

$$Pr\left[\,|X-\mu_X| \geq c\sigma_X\,\right] \leq \frac{1}{c^2}$$

Here c can be any real number

PROOF OF CHEBYCHEV

Let RV = $Y = (X-\mu_X)^2$ ← squaring gives an adv. the RV, Y is non-negative now.

So, first we need the exp of RV

$E[Y] = E[(X-\mu_X)^2] = \sigma_X^2$ = variance

$Pr[\,|X-\mu_X| \geq c\,\sigma_X\,]$ = [sq^n on both sides]

$= Pr((X-\mu_X)^2 \geq c^2\sigma_X^2)$

$= Pr(\,Y \geq c^2 E[Y])$

(using markov Ineq.) ∵ Y is non neg.

$Pr[|X-\mu_X| \geq c\sigma_X] = Pr[Y \geq c^2 \underbrace{E[Y]}_{\mu_Y}] \leq 1/c^2$

$\Rightarrow Pr[|X-\mu_X| \geq c\sigma_X] \leq 1/c^2$

HP

As we said earlier, more info — better bounds / tail ineq.

So,

first moment = Markov

variance / second = Chebycher

(3) Chernoff Bounds

or Chernoff Hoeffding Bounds

→ its applicability is restricted (reasons marked — R_1, R_2)

RV = X is defined as the sum of n

(R1) independent and identically distributed RVs $X_1, X_2, \ldots, X_n$

(being identically distributed is not a limitation as it can be removed by doing some calc.)

$$X = \Sigma \text{ niid RVs}$$

$$X = \Sigma_i X_i$$

eg. for chernoff → ASSUME, each x_i is a bernoulli random variable i.e. each x_i takes values in $\{0, 1\}$.

let $Pr(X_i = 1) = p$ and $\therefore P(X_i = 0) = 1 - p$

let $E[X] = \mu$

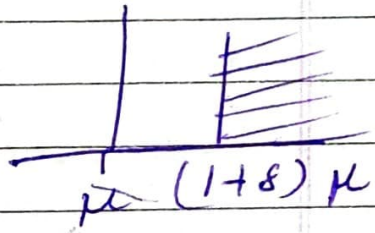
So, $E[\Sigma_i X_i] = \mu$

Teacher's Signature

$$= \sum_i E[X_i]$$
$$= n \cdot E[X_1] \quad (\because \text{they are identically dist})$$
$$= n \cdot (1.p + (1-p).0)$$
$$= np$$

$$\therefore \mu = E[X] = np$$

Now, Chernoff bound theorem (tail ineq. via CB) for the sum of IID Bernoulli variable



Given earlier conditions, for any $\delta > 0$, (R2)

$$Pr(X \geq \mu(1+\delta)) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

To prove this, we use exponential moments (normal strategy to prove tail estimates of sums of Independent RV).

exponential moments are ~~three~~ kinds:

$E[e^X]$ where X is RV

↓

exponential moment of X

Teacher's Signature

SHREE
DATE: / /
PAGE NO.:

PROOF:

For each i $(1 \leq i \leq n)$, we define a RV

$Y_i = e^{tX_i}$ for a real number $t > 0$ (parameter chosen accordingly)

($Y_i = e^{tX_i}$)

Now,

→ Y_i = positive RV or non-neg RV

→ $E[Y_i] = E[e^{tX_i}]$

$$= \sum_x e^{tx} \cdot Pr(X = x)$$

$$= e^{t(0)} \cdot Pr(X=0) + e^{t(1)} Pr(X=1)$$

$$= e^t \cdot p_i + 1 - p_i$$

where $p_i = E[X_i]$

Now, we define another RV, Y such that.

$$Y = Y_1 \cdot Y_2 \cdot \cdots Y_n$$

$$E[Y] = E[Y_1 \cdot Y_2 \cdots Y_n]$$

(∵ they are mutually independent) — we can →

$$= \prod_{i=1}^{n} E[Y_i]$$

(∵ they are ~~independent~~ identically distributed RVs) ←

$$= (p_i e^t + 1 - p_i)^n$$

Teacher's Signature

$$Y = Y_1 . Y_2 . Y_3 \cdots Y_n$$

$$= e^{tX_1} . e^{tX_2} \cdots e^{tX_n}$$

$$= e^{t(X_1 + X_2 + X_3 \cdots X_n)}$$

$$= e^{tX} \quad (\because X = \sum_i X_i)$$

So, $\boxed{Y = e^{tX}}$ where $t > 0$

and Y is non-negative value.
so we can apply markov on Y.
but before that –

$Pr(Y \geq ?)$ for markov

So, for event

$$X \geq \mu(1+\delta)$$

what would be the corresponding event for Y.

$$\frac{\ln Y}{t} \geq \mu(1+\delta)$$

$$\ln Y \geq \mu t(1+\delta)$$

$$\ln Y \geq npt(1+\delta)$$

$$Y \geq e^{npt(1+\delta)}$$

$$Y \geq \underbrace{\frac{e^{\mu t(1+\delta)}}{E[Y]}}_{c} \cdot E[Y]$$

So, now applying markov inequality—

$$Pr(Y \geq cE[Y]) \leq \frac{1}{c}$$

$$Pr\left(Y \geq \frac{e^{\mu t(1+\delta)}}{E[Y]} \cdot E[Y]\right) \leq \frac{E[Y]}{e^{\mu t(1+\delta)}}$$

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

$(\because 1+x \leq e^x)$

$(x = -p_i + p_i e^t)$

$$\leq \frac{\Pi e^{-p_i+p_i e^t}}{e^{t\mu(1+\delta)}}$$

Teacher's Signature

$$\leq \frac{\pi e^{-p_i'(1-e^t)}}{e^{t\mu(1+\delta)}}$$

$$\leq \frac{e^{-p_i(1-e^t)\cdot n}}{e^{t\mu(1+\delta)}}$$

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

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

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

$$\leq e^{-\mu t(1+\delta)-\mu(1-e^t)} \quad -(5)$$

now, take any value for t which pushes the upper bound as small as possible i.e. that minimises the RHS (using differentiation)

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

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

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

$$f'(t) = -\mu \frac{d}{dt}(t+\delta t+1-e^t)$$

$$= -\mu(1+\delta+0-e^t)$$

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

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

Teacher's Signature

SHREE
DATE : / /
PAGE NO. :

So for $f'(t) = 0$

$$0 = e^t - (1+\delta)$$

$$\boxed{\ln(1+\delta) = t}$$

[Now verify this t is minima using double differentiation]

Substituting t in (5)

$$Pr(X \geq \mu(1+\delta)) = Pr\left(Y \geq \frac{e^{\mu t(1+\delta)}}{E[Y]} E[Y]\right) \geq$$

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

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

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

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

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

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

Teacher's Signature

SHREE
DATE. / /
PAGE NO.:

Class 3 : Jan 12'21

Example = Coin tosses — 100 fair coins

Q1. What is expected no. of heads.

$$E[\#\text{Heads}] = 50$$

Consider the event that we witness at least 70 heads.

② Markov : $Pr[X \geq 70] \leq \frac{50}{70} = \frac{5}{7} \simeq 0.7$

① Binomial : $Pr\{X \geq 70\} = \sum_{x=70}^{100} \left(\frac{1}{2}\right)^{100} \binom{100}{x}$

or

$$= 1 - \sum_{x=0}^{69} \left(\frac{1}{2}\right)^{100} \binom{100}{x}$$

③ Chebychev

$$Pr(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

Since in our example,

$x \geq 0$

so

$$Pr(x - \mu \geq c\sigma) \leq \frac{1}{c^2}$$

Teacher's Signature

$$X = \sum_{i=1}^{100} X_i$$

$$\text{variance}(X) = n(\text{var}(X_1))$$

$$\text{var}(X_1) = E[X^2] - (E[X])^2$$
$$= E[X^2] - \left(\frac{1}{2}\right)^2$$
$$= \frac{1}{2} - \frac{1}{4}$$
$$= \frac{1}{4}$$

$\text{var}(X) = npq$
$= 100 \times \frac{1}{2} \times \frac{1}{2}$
$= 25$

So, $\text{var}(X) = \frac{100}{4} = 25$

$\sigma_X = \sqrt{25} = 5$

Chebychev

$$Pr(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

$$Pr\left(X - \mu \geq \left(\frac{5}{1}\right)\left(\frac{70}{5}\right)\right) \leq \frac{1}{c^2}$$

$$\Rightarrow \left(\frac{5}{70}\right)^2$$

$X - \mu \geq 5 \times c$
$X \geq 50 + 5 \times c$
$= 70$

Teacher's Signature

$$Pr\left(X - 50 \geq c\sigma\right) \leq \frac{1}{c^2}$$

$$Pr\left(X - 50 \geq 4 \times 5\right) \leq \frac{1}{c^2}$$

$$c = 4$$

$$\text{So } Pr(X \geq 70) = \frac{1}{16}$$

30 mins

Teacher's Signature

SHREE
DATE : / /
PAGE NO. :

Chernoff bound

$$P_r(X \geq \mu(1+\delta)) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

$P_r(x \geq 70)$

$P_r\left(x \geq 50\left(1+\frac{2}{5}\right)\right)$ $\quad \boxed{\delta = \frac{2}{5}}$

$$\leq \left(\frac{e^{2/5}}{\left(\frac{7}{5}\right)^{(7/5)}}\right)^{50}$$

$$\leq \left(\frac{e^{0.4}}{(1.4)^{1.4}}\right)^{50} \simeq 0.028$$

ANOTHER WAY OF WRITING THE RHS OF Chernoff Bound by simplyfing it

$$P_r(x \geq \mu(1+\delta)) \leq \begin{cases} e^{-\frac{\mu\delta^2}{4}} & \text{if } \delta \leq 1 \\ e^{-\mu\delta\ln\delta} & \text{if } \delta > 1 \end{cases}$$

But this isn't a close bound but a lose & rough however this can be used in some ~~versions~~ questions (complicated)

Teacher's Signature

$2x + y = 1$

$\therefore = \frac{1}{2}$

SHREE
DATE. / /
PAGE NO.:

Applications of Tail Inequalities

- when assuming IID in proving Chernoff bound – it only affects μ

  So if not IID – then μ will not be np – only it will change.

Q will chernoff work for $x > 50$ only?

40:31

App contd.

large dataset — features are ⊕nt

eg. cars — eg. colour, manual/auto etc.

If we want to divide this dataset into 2 parts – eg. test and train

Teacher's Signature

making sure that both are roughly similar

→ also used in drug traile

Candidate	features			
	Age	Diabetic	female	Hypertension

Divide into 2 parts { actually get drug / Placebo

both groups must have ~~&~~ equally distributed people/~~features~~.

i.e no of data items ∈ a feature are same (roughly) across the two division

also called SET BALANCING PROBLEM

So,

→ n data items ~~@~~ ; n features

→ entries = {0, 1}

→ Rows = features
Columns: Data items

Teacher's Signature

A — boolean matrix

label +1 ↓ data set (1)

label −1 ↓ data set (2)

So, we want a vector of size 'n'

1:00

Goal — Minimise diff in no. of people for each feature in two data sets

Maximum absolute entry in Ax indicates how many data items differ at feature i acc to division given by x.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \quad Ax = \begin{bmatrix} -2 \\ 1 \\ -1 \\ -2 \end{bmatrix}$$

MAE = 2

(Vector x has size n)

Teacher's Signature

No good deterministic algo for this problem.
i.e. they don't give soln close to s

Brute force – try all possible vectors = 2^n
& check which gives best soln.

Randomised
↓ each x
choose at random from $\{1, -1\}$ so = prob $\frac{1}{2}$

Now, we will show that the max. absolute entry of Ax in such an x (vector) is bounded by $O((n \ln n)^{1/2})$ with high probability

let $AX = Y$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}_{n \times 1}$$

So, $Y_1 = A_{11} X_1 + A_{12} X_2 + \cdots + A_{1n} X_n$
(matrix vector product)

$$A = \begin{bmatrix} 0 & -1 \\ \text{matrix} \end{bmatrix}_{n \times n} \qquad X = \begin{bmatrix} +1 & -1 \\ \text{vector} \end{bmatrix}_{n \times 1}$$

matrix A is fixed & X is random.
data items — random values.

Teacher's Signature

$$A_{n \times n} \; X_{n \times 1} = Y_{n \times 1}$$

$$Y_i = \sum_{j=1}^{n} A_{ij} \cdot X_j$$

SHREE
DATE : / /
PAGE NO. :

So, Y_i is also a random value because it is a func^n of X_i

X_i = RV $\{-1, 1\}$ ← takes values in

Y_i = RV $\{-n, +n\}$ ← takes values in

$E[X_i] = 0$ because we're picking 1 & -1 with equal prob.

$$E[X_i] = \frac{1}{2} \times 1 + \frac{1}{2}(-1) = 0$$

$E[Y_i] = 0$ by linearity of expectation

$\Rightarrow$ So, $E[MAE] = 0$ ⊛ Proof(top)

So, Y_i can have these A_{ij} values as 0 or 1.

So, if we assume that all non-zero A_{ij} are present at beg. i.e. only first k values in A matrix are non-zero then,

$$Y_i = \sum_{j=1}^{k} A_{ij} X_j$$

under this assumption Y_i is the sum of n IID RVs each of which has expectation zero so, $E[Y_i] = 0$

Apply Chernoff bound on Y

In principle,

$X_i = \{0, 1\}$ with prob 1/2

↑ Bernoulli

$$X = \sum_{i=1}^{n} x_i$$

but we have $X_i = \{-1, 1\}$ so one possibility

Teacher's Signature

(★) So, we want to find how far can the value of Y be from its E[Y] which is 0.

Class 4 : Jan 15' 21

SHREE
DATE : / /
PAGE NO. :

One possibility : derive Chernoff bound again for the sum of n {-1, +1}

Prob ($X >= k$) for some integral k.

$$E[X] = n \cdot E[X_i] = n \cdot 0 = 0$$

Other possibility :

Define new variable T_i (← this T_i is diff than $Y_i = A_i X_i$)

$T_i = (1 + X_i)/2$

now, T_i is {0, 1} valued.

Define as sum of T_i

$$T = \sum_{i=1}^{n} T_i$$

$$E[T_i] = \frac{1}{2} + E[X_i] = \frac{1}{2} = 0 = \frac{1}{2}$$

$$E[T] = n E[T_i] = n \times \frac{1}{2} = \frac{n}{2}$$

$$\boxed{T = \Sigma \frac{(1 + X_i)}{2} = \frac{n}{2} + \frac{X}{2}}$$

Also, $X >= k$ if & only if $T >= \frac{n}{2} + \frac{k}{2}$

So event ↑ is same as ↑

So, $Pr(X >= k) = Pr(T >= n/2 + k/2)$

Teacher's Signature

Choose $k = 8\sqrt{n \ln n}$ Why? we want

SHREE
DATE : / /
PAGE NO. :

Constraint on k

In general, k can be at most n
as n RV each -1 or 1
so max — n
min — -n
} for k

not like original X

So, if $X \geq k \Rightarrow T \geq n/2 + k/2$

(?) X is same as Y_1 — (2)

$T \geq E(T)\ (c)$ $(c = 1+\delta)$

$T \geq n/2 \left(1 + \frac{k}{n}\right)$ $\left(\frac{k}{n} = \delta\right)$

on solving this we get

$\leq \exp\left(-\frac{E(T)\delta^2}{4}\right)$

$\leq \exp(-k^2/8n)$ — (1) with $\delta = k/n < 1$

Coming back to set balancing problem

where $Y_i = A_{i1}X_1 + A_{i2}X_2 \cdots$

So, we get

$Pr(Y_1 >= 8\sqrt{n \ln n}) \leq \exp\left(\frac{-64 n \ln n}{8n}\right)$ using (1) — solved next page

by (2) $k = 8\sqrt{n \ln n}$

$= \exp(-8 \ln n)$

$= n^{-8}$

Teacher's Signature

SHREE
DATE: / /
PAGE NO.:

So, we know

$$Pr(Y_1 \geq 8\sqrt{n \ln n}) \leq \frac{1}{n^8}$$

two sided bound

But, we were interested in Max. absolute value of Y_i so, we also need —

$Pr(Y_1 \leq -\delta)$ which is

$Pr(Y_1 \leq -8\sqrt{n \ln n})$

which will also be $\leq \frac{1}{n^8}$

(by symmetry) or by deriving left tail*

So,

$$Pr(|Y_1| \geq 8\sqrt{n \ln n}) \leq \frac{2}{n^8}$$

* LEFT TAIL

also Q3 of HW2

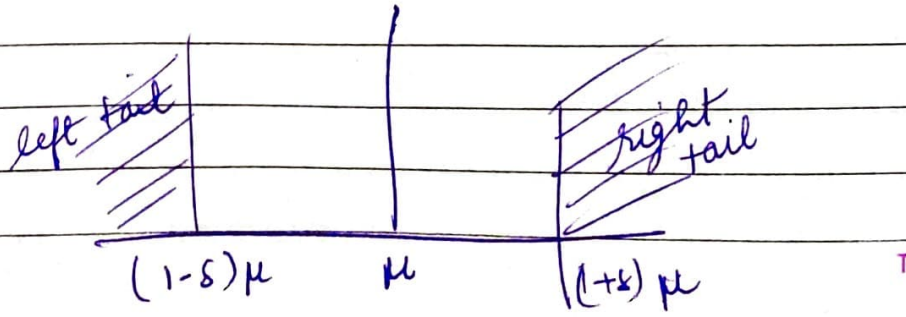
Let X_i be IID Bernoulli RV with success prob. p.

Let

$X = \sum_i X_i$ let $E[X] = \mu$

For any $0 < \delta < 1$,

$Pr(X < (1-\delta)\mu) \leq \exp(-\mu\delta^2/2)$



Teacher's Signature

~~So, we can say this for each Y_i~~

~~$Pr(|Y_i| \geq 8\sqrt{n \ln n}) \leq \frac{2}{n^8}$~~

But what about Y_2, Y_3

So, using

BOOLE'S INEQUALITY

for any n events, $E_1, E_2 \cdots E_n$

$Pr(E_1 \cup E_2 \cup \cdots \cup E_n) \leq Pr(E_1) + Pr(E_2) + \cdots + P(E_n)$

define $E_i = \{ |Y_i| \geq 8\sqrt{n \ln n} \}$

$$E = \cup E_i$$

$$Pr(E) = Pr(\cup E_i)$$

$$\leq n \cdot \frac{2}{n^8}$$

$$\leq \frac{2}{n^7}$$

Complement of $\cup E_i$ will be that none of E_i happen at all.

So. $\sim \cup E_i = \sim E_1 \wedge \sim E_2 \wedge \sim E_3 \wedge \cdots \wedge \sim E_n$

$\Rightarrow$ every Y_i has absolute value within $8\sqrt{n \ln n}$

$\geq 1 - \frac{2}{n^7}$ } $\geq$ prob at least $1 - \frac{2}{n^7}$

Teacher's Signature

SHREE
DATE: / /
PAGE NO.:

Another Application of Tail Inequalities

Randomised Rounding

Let A be a $n \times n$ $\{0,1\}$ matrix and X be a column vector of size 'n'. The elements of X are real numbers ≥ 0.

We want to find a boolean vector Y such that the largest absolute entry in $A(Y-X)$ is minimised
↳ $\| |A(Y-X)| \|$ (L1-norm)

In other words, we need Y such that

$$AX \cong AY$$

i.e. X is very very similar to Y or that Y pretends to be X

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad X = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.5 \\ 0.6 \end{bmatrix} \quad Y = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$Y - X = \begin{bmatrix} -0.3 \\ -0.2 \\ 0.5 \\ 0.4 \end{bmatrix}$$

minimise this ↓

$$A(Y-X) = \begin{bmatrix} 0.2 \\ 0.6 \\ -0.2 \\ -0.5 \end{bmatrix}$$

max entry = 0.6

Teacher's Signature

* $X \in \{0,1\}$ else scale them

- Again, deterministic algo are inefficient
So, we use randomised algorithm

- We won't worry about A.

→ Let each Y_i be value 1 with prob. X_i
In our eg.,
$Pr[Y_1 = 1] = 0.3$

Now, we show that the Y prepared above has the property that the MAE of $A(Y-X)$ is $O(\sqrt{n \ln n})$.

Proof: Technique:
show take one of Y_i i.e. Y_1
and one of $A(Y-X)$
and repeat over all entries

~~$Y_1 \in \{0,1\}$~~

Y_1 is a RV

$$Y_1 = \begin{cases} 0 \rightarrow \text{prob } 1-X_1 \\ 1 \rightarrow \text{prob } X_1 \end{cases}$$

Hence $E[Y_1] = 1 * X_1 = X_1$

(X_1 is not a RV, but a scalar given as input)

Teacher's Signature

→ Let $Z = A(Y-X)$. First entry in Z

$$Z_1 = A_{11}(Y_1 - X_1) + A_{12}(Y_2 - X_2) + \dots$$

→ Z_1 is also a RV (∵ Y_i is RV)

$$E[Z_1] = E[A_{11}(Y-X_1) + A_{12}(Y_2 - X_2) \dots]$$

∵ we are looking at only the maximum absolute entry in Z, we can take all A_{ii} as 1 without loss of generality.

So, $E[Z_1] \leq E[(Y_1 - X_1) + (Y_2 - X_2) + \dots (Y_n - X_n)]$

(linearity of exp) $= E[Y_1 - X_1] + E[Y_2 - X_2] + \dots + E[Y_n - X_n]$

∵ $E[Y_i] = X_i$ and X_i is not a RV, so,

$$E[Y_i - X_i] = 0$$

So, $E[Z_1] = 0$. This implies that the expected value of largest absolute entry is zero. But, we require a tail bound — a parameter 'd' such that $|Z_i|$ does not exceed d ≥ high prob.

Now, $Z_i = \sum_i Y_i - X_i$

$= \sum_i T_i \quad (Y_i - X_i = T_i)$

and if we consider Y_i as independent, then T_i is indep. and hence Z_i is independent RVs.

(if we carry the calculations of Chernoff bound on independent and not necessarily identical RVs, we end up with same result.)

maybe HW2 Q2 i establish expo. moment

Y_i	T_i
0	$-X_i$
1	$1-X_i$

So, $T_i \in \{-X_i, 1-X_i\}$

Teacher's Signature

if $X_i = 1$
$T_i \in \{0, 1\}$

this constant can be any thing

Consider $d = 4\sqrt{n \ln n}$

$$Pr(Z_i \geq d) \leq \exp(-d^2/2n)$$
$$= \exp\left(-\frac{16 n \ln n}{2n}\right)$$
$$\leq \frac{1}{n^8}$$

By, symmetry
$$Pr(|Z_i| \geq d) \leq \frac{2}{n^8}$$

Using Boole's inequality —
$$Pr(|Z_i| \geq d) \leq 2/n^7$$

Hence, every entry of Z has an absolute value of at most d.

Chernoff bound gives $1/n^c$ → this constant will depend upon what deviation is used & what constant is used in deviatn(d).

Lecture 5 : Jan 19' 21

SHREE
DATE. / /
PAGE NO.:

Applications of Randomisat^n

└ clubbed ε algebraic techniques
└ applicat^n : verifficat^n of identities

Technique : fingerprinting

U – universe of objects
x, y – two elements from U

Q. ⇒ Is $x = y$?

A. (i) deterministic manner = using
{ log |U| bits at least (~~every obje~~
↳ TRICY

every object ← compare these bit by bit
is given ID of log |U| bits
but size of U can be vv large.

Are there better methods?

(ii) Smaller universe V. – map ele. of U to ele of V.

$$f : U \longrightarrow V$$

elements x & y are identical if & only if $f(x) = f(y)$
i.e. their images on V are identical. But this may not ALWAYS work as
$|V| < |U|$

|V| = no. of unique fingerprints

So, we will see what chance do we have ε this approach

Teacher's Signature

SHREE
DATE : / /
PAGE NO. :

Example =

Here U will be a set of square matrices — B

Let F be a field and A & B are two matrices with entries from F

(Properties of Field (Read by yourself ← Algebra)

Claim = C = A · B

Method 1 = A · B ← compute

compare ≤ C → each element of A·B & C.

Time Complexity =

→ Strassen's recursive algo. takes $O(n^{\log_2 7})$ — more practical algo.

$\simeq O(n^{2.7\ldots})$

→ Current best algo runs in $O(n^{2.376})$ — but difficult

So we can say to verify our claim it takes time equal to multiplying two matrices.

(V is the set of vectors of 0,1 elements)

So $|V| \leq |U|$

Here, $|V| = 2^n$ $|U| = 2^{n^2}$

Teacher's Signature

(read comments)

We will look at a randomised approach:-

* Remember 0 & 1 are additive & multiplicative identities. (assume 0 & 1 are add/mul identities of field F)

Let 'r' be any vector ~~&~~ entries 0 ~~and~~ or 1
$n \times 1$

and each element of 'r' is chosen independently and uniformly at random.

Now, $y = A \times B \times r$
$(n \times 1) \quad (n \times n) \quad (n \times n) \quad (n \times 1)$

To improve time complexity of this multiplicat^n:-
i.e. to avoid doing A×B (same as deterministic)

we, use <u>associativity</u> —

$X = B * r \rightarrow O(n^2)$

$y = A * X$

Similarly, $z = C * r \rightarrow O(n^2)$

Now if $A * B = C$ is true,
then, $y = z$ for any r

and computations take $O(n^2)$. So, time complexity is established.

Teacher's Signature

So, are we really safe to say that
$C = A \cdot B$ if $y = z$?

~~if y=z?~~

→ it may be possible that
$A \times B \neq C$ but
$A \times B \times r = C \times r$

When will this happen?

→ Let $D = AB - C$ and $AB \neq C$
D is not the matrix of all zeros.

We are interested in the event that
$Dr = 0$

eg: $A \times B = \begin{bmatrix} 1 & 0 \\ 2 & 4 \end{bmatrix}$ $C = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$

$D = A \times B - C = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}$

for $r = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, $Dr = 0$

$C \times r = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

So here,
$A \times B \neq C$
but,
$A \times B \times r = Cr$

Teacher's Signature

So, what is the prob. of this happening?
↓
$A \times B \neq C$ but $A \times B \times r = Cr$

Lemma: Let A, B & C be $n \times n$ matrices from F such that $A \times B \neq C$. Then, for r chosen uniformly at random from $\{0,1\}^n$, $\boxed{Pr(ABr = Cr) \leq 0.5}$

Pr('r' happens in null space of AB-C) ≤ 0.5

Proof: Consider
$D = AB - C$.
$\because$ $AB \neq C$, D is not all zeros

So, $\{ABr = Cr\} = \{Dr = 0\}$

Assumption: Assume without loss of generality that the first row of D has a non-zero entry and all non-zero entries in that row are before any zero entry.
(so we want all zeros to be on one side & all non-zeros to be on one side of the row.
And even if it doesn't look like that we can re-arrange rows/columns. And these rearrangements/transformations done on D — corresponding transformations must be done on 'r' as well)

Teacher's Signature

So,

$$D \times r = \begin{bmatrix} \neq 0 & \cdots & 0s \\ \leftarrow k \rightarrow & & \end{bmatrix} \times \begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Consider first row of $D \times r$ —

$= D_{11} \times r_1 + D_{12} \times r_2 + \cdots D_{1n} \times r_n$

(using our assumption)

$= D_{11} \times r_1 + D_{12} \times r_2 + \cdots D_{1k} \times r_k$

where $1 \leq k \leq n$

if $D_{11} \times r_1 + D_{12} \times r_2 + \cdots D_{1k} \times r_k = 0$

$$D_{11} \times r_1 = -\sum_{i=2}^{k} D_{1i}\, r_i$$

$$r_1 = \frac{-\sum_{i=2}^{k} D_{1i}\, r_i}{D_{11}} \quad \text{(1)}$$

$D_{11} \neq$ non zero because of our assumption

So if r_1 ↑

then first element of $D \times r$ is zero.

Remember, entries of r are independent & chosen VAR. So r_i is a RV and Hence, an expression containing them is also a RV.

So, we can find $P_r((Dr)_1 = 0)$

Teacher's Signature

$(Dr_0)_1 = (D \cdot r)_1$ — first element of $D \cdot r$

But $\{(Dr)_1 = 0\}$ is a super event of $\{Dr_0 = 0\}$. Because it is possible that $\{(r)_1 = 0\}$ is true but $\{Dr = 0\}$ is not but vice versa follows. Hence $\{r_1 = 0\}$ is a SUPER EVENT.

So, $$\boxed{Pr(Dr = 0) \leq Pr((Dr_0) = 0)}$$

* Remember,
  $Pr((r_1) = 0)$ is not same as picking r_1 as zero from the field because r_1 is related to other RVs that we are choosing (eq. ①)

Using Principle of deferred choices, imagine that all choices $r_2, \dots r_k$ have been made. (to compute $Pr(\{(D \cdot r)_1 = 0\})$)

→ In that case, RHS i.e. s is a scalar from the field F. (└ in eq ①)

→ The LHS is a value chosen at random amongst (at least) two values in F. in eq ①
The req. prob. therfore cannot exceed 1/2.
Because $|F| \geq 2$

So, $$P((D \cdot r)_1 = 0) = Pr\left(r_1 = \frac{-\Sigma}{D_{11}}\right) \leq \frac{1}{2}$$

So, $Pr(Dr = 0) = Pr(ABr = Cr) \leq 1/2$

Teacher's Signature

if ~~F~~ $F = \{0, 1\}$

$r_1 = -\Sigma$ ——
D_{11}

$r_1 \in \{0, 1\}$
because $\times / \div / + / -$ on $\{0, 1\}$ will give $\{0, 1\}$

~~if~~ if F is large —
So, if $S \notin \{0, 1\}$, then,
$r_1 \neq S$ because choice of r_1 is limited to $\{0, 1\}$

then prob $(D_r = 0)$ will be much precise / zero in this case

So, let's say $S = 1$
So, we want $P_r(r_1 = 1)$
which is $1/2$.

But in general any field will have more elements, but 0, 1 will be there as they are additive / mul identity.

(?) No matter the size of field
$r_1 \in \{0, 1\}$

Teacher's Signature

SHREE
DATE / /
PAGE NO.:

We need to improve the verification efficiency of the procedure; so we can use repeated independent trials.
So, we will do an experiment 't' times —
i.e. t independent trials of above procedure

for $AB \neq C$, $P(\text{test fails in each trial}) \leq 1/2$
So, in 't' trials, —
$P(\text{all t trials fail}) \leq (1/2)^t$

A failure is when indeed $AB \neq C$ & the chosen 'r' s.t. $ABr = Cr$

eg.

t:	1	2	3	4	← trials (t)
$ABr_t = Cr_t$	✓	✓	✓	...	X

if at some r_t $ABr \neq Cr$
⇒ there exists a r in the null space of $AB - C$
⇒ $AB \neq C$

But if

					t
	✓	✓	✓	✓	✓

then $P(AB = C) \leq (1/2)^t$

Teacher's Signature

Proof: When $t = O(\log n)$ where n is the size of matrices -
the failure prob. will be $\leq \frac{1}{n^c}$

which is very low!

examples of field:

→ $\mathbb{Z}_p$ = set of residuals % p ← prime
$\{0, 1 \ldots p-1\}$

→ Real Numbers

Teacher's Signature

Lecture 6: Jan 22'19

SHREE
DATE / /
PAGE NO.:

Another applicat^n of fingerprinting — verifying polynomial identities

For instance, let $P_1(x)$ and $P_2(x)$ be two polynomials in a field F.
The polynomial product verification:

$$P_1(x) \cdot P_2(x) = P_3(x)$$

for a given $P_3(x)$

Note
Multiplicat^n of two polynomials: takes $O(n \log n)$ time where n is the $\max(\deg(P1), \deg(P2))$.

How quickly can we find value of P(x) given a x?
↳ Horner's Method

Time $\propto$ degree of P(x)
↳ $O(n)$

So, we design a faster verification procedure —
if $P_3(x) = P_1(x) \cdot P_2(x)$
then, $P_3(r) = P_1(r) \cdot P_2(r)$
for 'r' chosen uniformly at random from 'S', where $S \subset F$
i.e. S is a subset of the entire field &
$|S| \geq 2n+1$ ⇐ why? (PTO)

Teacher's Signature

→ And we already know that evaluating a polynomial at a given input can be done in $O(n)$ time.

→ So, we can declare that

$$P_3(x) = P_1(x) \cdot P_2(x)$$

unless $P_3(r) \neq P_1(r) \cdot P_2(r)$

→ This algo fails only when

$$P_3(r) = P_1(r) \cdot P_2(r)$$

but $P_3(x) \neq P_1(x) \cdot P_2(x)$

→ So what is the prob. of this happening when we pick a 'r' uniformly at random from set 'S'.

$|S| \geq 2n+1$ why?

P_1, P_2 - degree $= n$

P_3 - degree $\leq 2n$

or

degree $= 2n$ ($\because P_1, P_2$ degree $= n$)

Now,

$$Q(x) = \underbrace{P_3(x)}_{\substack{2n \\ \text{(degree)}}} - \underbrace{P_1(x) \cdot P_2(x)}_{\substack{2n \\ \text{(degree)}}}$$

So, degree$(Q(x)) \leq 2n$

Teacher's Signature

SHREE
DATE / /
PAGE NO.:

→ In cases where
$$P_3(x) = P_1(x) \cdot P_2(x)$$
this holds

→ When $P_3(x) \neq P_1(x) \cdot P_2(x)$
then,
$Q(x)$ is non-zero polynomial

Now, in our test (in this case)
$$P_3(r) = P_1(r) \cdot P_2(r)$$
$\Rightarrow Q(r) = 0$
but $P_3(x) \neq P_1(x) \cdot P_2(x)$
} test fails

∵ $Q(x)$ can have at most $2n$ roots (∵ degree $(Q(x)) = 2n$)

$\Rightarrow$ $2n$ points where polynomial Q will evaluate to zero.

$\Rightarrow Q(r) = 0$ when any of these $2n$ points are picked out of S.

$$\Rightarrow P(Q(r) = 0) = \frac{2n}{|S|}$$

this is also the prob. of error.

So, $|S|$ is at least $2n+1$.
Now, $\frac{2n}{2n+1}$ is a large value, so, to reduce error prob. we ↑se $|S|$. OR use repeated trials OR do both

Teacher's Signature

Practical applicatn of this technique - when you do not have access to these polynomials.
Because, here in this approach we need not know what the polynomial is - we just need the value of the polynomial.

eg. polynomial corresponding to determi-nant of a matrix.
In this case, computing coefficients of polynomial is difficult.

Finger-printing (Conclusion)

These two algos we studied, have the property that for inputs that are identical the algo does not make any error.

But when they are not - then algo makes an error that is upper bounded by at least a constant.

To reduce this error prob. - we can do repeated trials to catch the error & make it small. ——

Consider 'A' to be a fingerprinting algo
let's run A on input x, y for 't' iterations
The 't' outputs are $O_1, O_2 \ldots O_t$
Each of these outputs is a yes/no.

Teacher's Signature

If any of these are 'no', then 'NO' is the answere.
If all are YES, then YES is the answer.

So, given that $x = y$

$$Pr(A(x,y) = YES) = 1 \quad (\leftarrow PTO)$$

given $x \neq y$

$$Pr[A(x,y) = YES) \leq (1/2)^t$$

So if $t = O(\log n)$
then error prob. is $O(1/n^c)$

$$\left(\frac{1}{2^t} = \frac{1}{2^{c\log_2 n}} = \frac{1}{n^c}\right)$$

Because of the nature of these algos, the algos we just studied are called co-RP algorithms. (RP = randomised polynomial)

→ co-RP is complement of RP

→ RP – The class RP consists of languages L s.t there exists a ① randomised algo running in ② worst case polynomial time s.t. for any input x:

③ . $x \in L \Rightarrow Pr(A \text{ accepts } x) \geq 1/2$
. $x \notin L \Rightarrow Pr(A \text{ accepts } x) = 0$

Teacher's Signature

SHREE
DATE: / /
PAGE NO.:

⇒ given x is part of L, your algo should output "yes x is part of L" w prob. at least $1/2$.

⇒ if x isn't part of L, algo should ~~also~~ always reject "x".

co-RP

① ✓
② ✓
③ changes — get reversed

if $x \in L \Rightarrow Pr(A \text{ accepts } x) = 1$

if $x \notin L \Rightarrow \underbrace{Pr(A \text{ accep } x)}_{\text{algo makes an error w prob. at most } 1/2} \leq 1/2$

→ RP do not make errors of the "false negative" kind whereas co-RP does.

→ But any RP or co-RP can err only on one side — either for x in L or for x not in L.

Quest^n: what is randomised quicksort RP or co-RP?

Neither. Because ~~Quest~~ quicksort takes a random pivot — but the output i.e. the

Teacher's Signature

Sort order is never incorrect. It's just the complexity that may be ~~best-case or~~ worst-case or not.
So, the correctness of algo is not an but how long it takes is.

Such algorithms are called ZP algo.

Zero-error Expected ~~algorithm~~ polynomial time.

The class zp consists of languages L s.t. there is a randomised algorithm A that always outputs the correct answer while running in expected polynomial time.

Another name: Las Vegas algorithm.

(2) → Monte-Carlo algorithms

Teacher's Signature