6.046	Lecture 7	Feb 27 2013

RANDOMIZED ALGORITHMS

- what are they?
- Quicksort
- Chernoff bounds

What are they?

- Algorithms with access to random source.
  - e.g. can flip coins / can roll dice / etc.
- Funky: on same input on different executions randomized algorithm may:
  - run for different #steps
  - produce different results

[depending on outcomes of coin flips]

- Flavors:

Monte Carlo	Las Vegas
• always runs in polynomial time • prob[output correct] > high	• Prob[output correct] = 1 • runs in expected polynomial time

expectations & probabilities are with respect to outcomes of coin tosses

- True Randomness?

  ◇ Assumption for this class:

  algorithm has access to subroutine that, given R, outputs a uniformly random number $r \in \{1, \ldots, R\}$

  ◇ In practice: pseudo-random number generator

  ◇ Fundamentally? philosomphical/religious belief

  Einstein: "... I, at any rate, am convinced that He does not throw dice."
  letter to Max Born 1926

Q: Are randomized better than deterministic algs?
↳ new problems? faster? simpler?

- A1: not a fair comparison, as randomized algorithms are allowed to make errors, take longer
- A2: in practice, they are faster and simpler
- A3: in theory, only polynomial gain* in running time
  *under complexity-theoretic assumptions

QUICKSORT [Hoare 1962]

- comparison sort (like Mergesort, unlike Radix Sort)
- divide & conquer
  - ↳ work on divide step
  - ↳ no work on combine step
- in place: $O(1)$ extra space
- practical (with tuning)
- flavors:
  - basic: worst-case $O(n^2)$
    expected $O(n \cdot \log n)$ for random input
  - randomized: expected $O(n \cdot \log n)$ for all inputs
  - deterministic: $O(n \cdot \log n)$ worst-case
    ↳ in practice, slower than randomized

Algorithm

INPUT: Array A of n elements

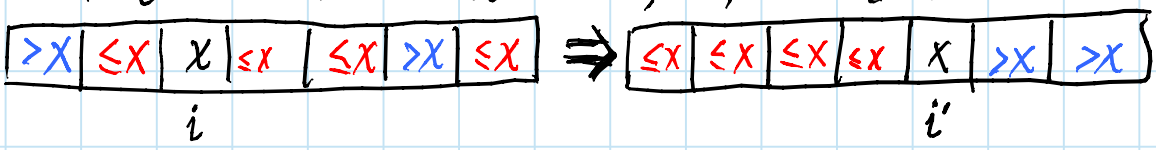
⓪ If $n=1$, stop; A is sorted

① divide:

- pick some $i \in \{1, 2, \ldots, n\}$
  - basic: $i=1$
  - randomized: $i \sim$ uniform$\{1, \ldots, n\}$
  - deterministic: i such that $A[i]$ is median of A

$x = A[i]$ is the pivot element

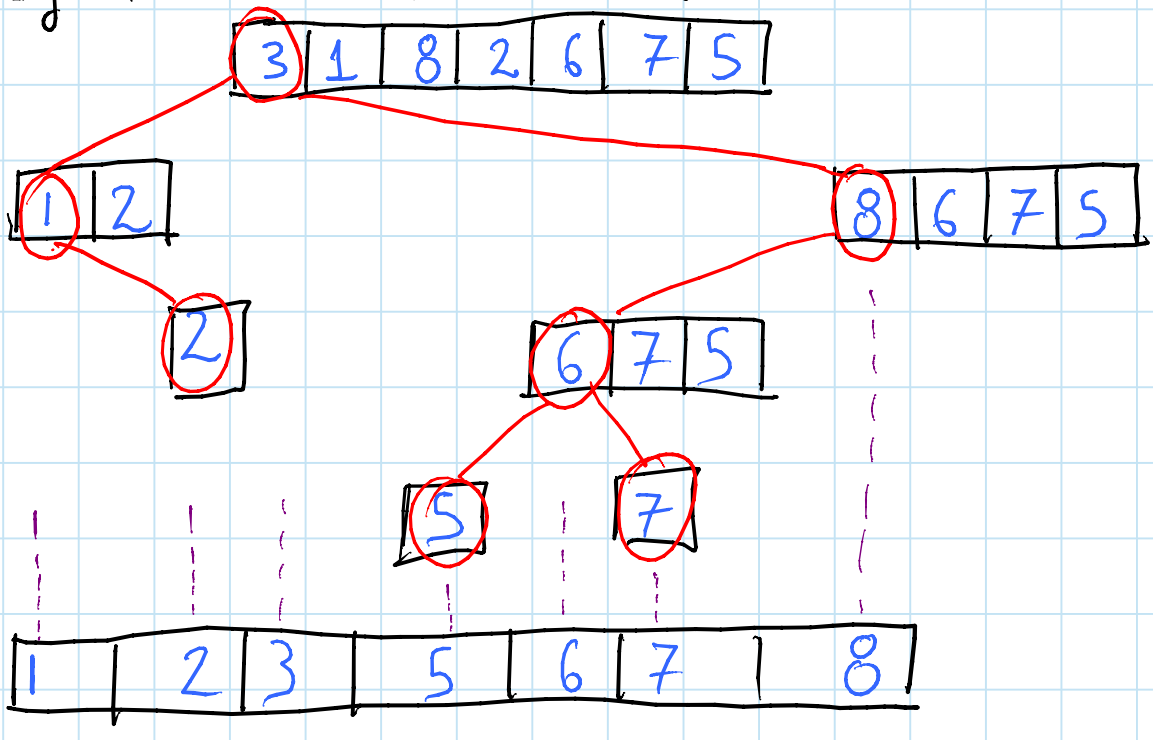
- partition A into elts $\leq x$, x, elts $> x$



② conquer: recursively sort elts $\leq x$ ($A[1 \ldots i'-1]$)
recursively sort elts $> x$ ($A[i'+1 \ldots n]$)

③ combine: do nothing :)

e.g. execution of basic Quicksort



(assuming stable partition algorithm)

sorted array:

exercise: basic Quicksort may take $\Omega(n^2)$

exercise 2: no matter what strategy for picking pivot is used: output of Quicksort is sorted array and runtime is $O(n^2)$

exercise 3 if median is used as pivot Quicksort runs in $O(n \cdot \log n)$ time

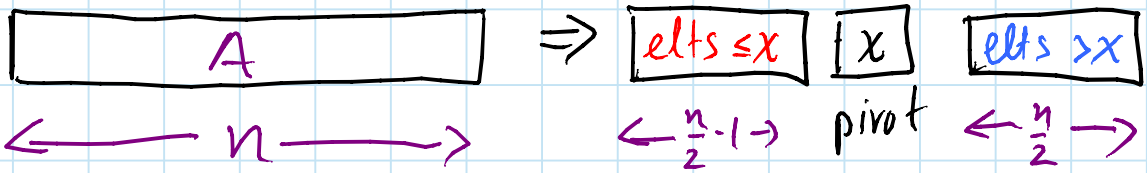
Analysis of RANDOMIZED QUICKSORT

- want to show expected running time is $O(n \cdot \log n)$
- will show something stronger, namely:

$$\Pr\left[\underbrace{\text{Quicksort takes longer than } c \cdot n \cdot \log n}_{\text{the Bad event}}\right] \leq \frac{1}{n}$$

$$\Rightarrow \mathbb{E}[\text{running time}] \leq \left(1 - \frac{1}{n}\right) \cdot c \cdot n \cdot \log n + \frac{1}{n} \cdot \underbrace{c \cdot n^2}_{\text{even under bad event runtime is bounded by } c \cdot n^2} = O(n \cdot \log n)$$

- Intuition: (i) suppose we're extremely lucky and every choice of pivot results in a 50-50 split of the elements

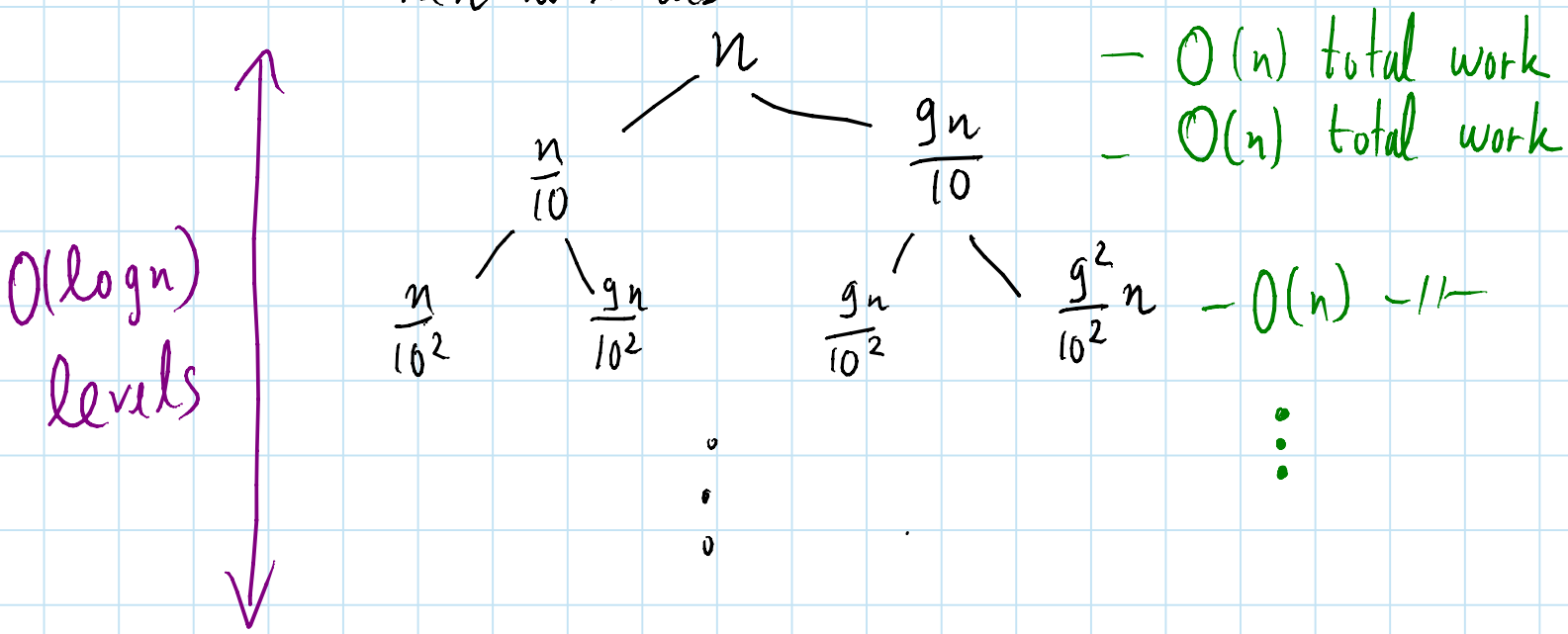


then $T(n) \leq 2 \cdot T\left(\frac{n}{2}\right) + O(n) = O(n \cdot \log n)$

probability [single split 50%-50%] $= O\left(\frac{1}{n}\right)$ (too unlikely)

(ii) why exact split? what if we always get a 10% - 90% (or more balanced) split?

then worst case:



Total work: $O(\log n) \times O(n) = O(n \cdot \log n)$

$\Pr[\text{single split is 10\%-90\% or better}] = \frac{8}{10}$:)

problem? $\Pr[\text{all splits are 10\%-90\% or better}] = \left(\frac{8}{10}\right)^{O(n)} \approx 0$

- Idea: It is not crucial that all splits are 10%-90%.

let's formalize this idea...

For every element α of the array and recursive depth t of quicksort, define random variable

$$X_{\alpha,t} = \begin{cases} 0, & \text{if in the subarray where } \alpha \text{ belongs in depth } t \text{ (*), quicksort chose a good pivot} \\ 1, & \text{o.w.} \end{cases}$$

good pivot: results in a 10%-90% or better split of the subarray

(*) note: if Quicksort finished before reaching rec. depth t, or if α does not appear in any sub-array at depth t (because in some earlier depth the subarray of α had only α in it), we still define $X_{\alpha,t}$ and take it to be $X_{\alpha,t} = \begin{cases} 0, & \text{wpr } 0.8 \\ 1, & \text{wpr } 0.2 \end{cases}$

- so for all α, for all t: $\Pr[X_{\alpha,t} = 1] = 0.2$
- observe also that $\forall \alpha$, $X_{\alpha,1}, X_{\alpha,2}, \ldots$ are independent random variables

(in contrast the variables $\{X_{\alpha,t}\}_{\alpha}$ for a fixed t are not independent)

◇ Let $T = c \cdot \log n$ (with log I mean $\log_e$)

c : some constant to be set later

$$\mathbb{E}\left[\sum_{t=1}^{T} X_{\alpha,t}\right] = 0.2 \cdot T$$

Q: $$\Pr\left[\sum_{t=1}^{T} X_{\alpha,t} \geq \gamma \cdot \mathbb{E}\left[\sum_{t=1}^{T} X_{\alpha,t}\right]\right]$$

γ : constant to be set later

$$\equiv \Pr[\,\mathrm{Bin}(T, 0.2) \geq \gamma \cdot 0.2 \cdot T\,] \quad ?$$

A1: $$\sum_{\lambda=\lceil 0.2\gamma T\rceil}^{T} \binom{T}{\lambda} \cdot 0.2^{\lambda} \cdot 0.8^{T-\lambda}$$

↳ accurate but not usable

A2: Central Limit Theorem:

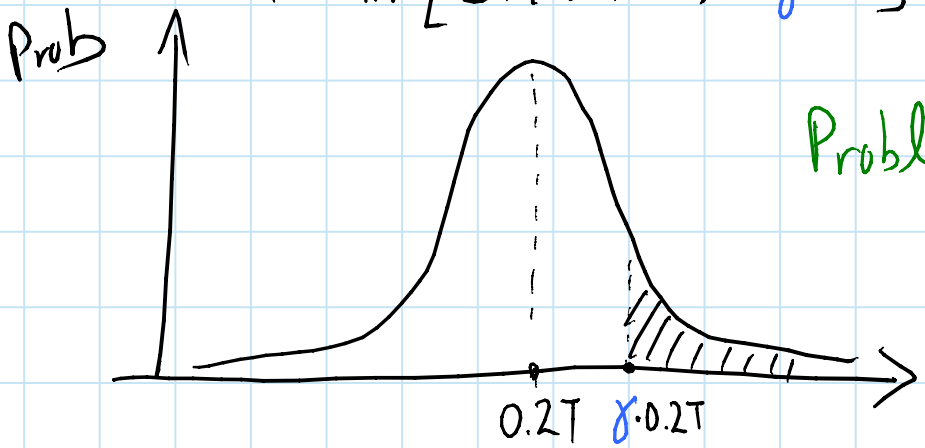
Gaussian distribution

as $T \to \infty$: $\mathrm{Bin}(T,p) \to \mathrm{Normal}(T\cdot p,\ T\cdot p(1-p))$

Binomial distribution

mean, variance of $\mathrm{Bin}(T,p)$

$$\Rightarrow \Pr[\mathrm{Bin}(T,0.2) \geq \gamma \cdot 0.2T] \overset{\circledast}{\approx} \Pr[N(0.2T,\ 0.16T) \geq \gamma \cdot 0.2T]$$



Problem: $\circledast$ holds in the limit as $T \to \infty$

so not usable directly either

A3: Chernoff Bounds

Theorem (Chernoff): Suppose $Y_1, Y_2, \ldots, Y_n \in [0,1]$ are independent random variables. Then $\forall\, \delta \in (0,1]$:

$$\Pr\left[\sum Y_i > (1+\delta)\,\mathbb{E}\left[\sum Y_i\right]\right] \leq e^{-\delta^2 \mathbb{E}\left[\sum Y_i\right]/3}$$

and

$$\Pr\left[\sum Y_i < (1-\delta)\,\mathbb{E}\left[\sum Y_i\right]\right] \leq e^{-\delta^2 \mathbb{E}\left[\sum Y_i\right]/2}$$

Proof: later

Using Chernoff Bound:

$$\Pr\left[\sum_{t=1}^{T} X_{\alpha,t} \geq \gamma \cdot \mathbb{E}\left[\sum_{t=1}^{T} X_{\alpha,t}\right]\right] \leq e^{-\frac{1}{3}\mathbb{E}\left[\sum_t X_{\alpha,t}\right]\cdot(\gamma-1)}$$

$$\leq \underbrace{e^{-\frac{1}{15}\cdot T\cdot(\gamma-1)}}_{\| \; n^{-\frac{1}{15}c\cdot(\gamma-1)}}$$

if we chose $c\cdot(\gamma-1) \geq 30$ e.g. $c=30, \gamma=2$ or $c=60, \gamma=1.5$ etc. $\longrightarrow \leq \frac{1}{n^2}$ $(*)$

Union bound: If $E_1, E_2, \ldots, E_n$ are arbitrary events then

$$\Pr[E_1 \text{ OR } E_2 \text{ OR } \ldots \text{ OR } E_n] \le \sum_{i=1}^{n} \Pr[E_i]$$

Since (*) holds for all α, the union bound gives

$$\Pr\left[\exists \alpha \text{ s.t. } \sum_{t=1}^{T} X_{\alpha,t} \ge \gamma \cdot 0.2 \cdot T\right] \le n \cdot \frac{1}{n^2} = \frac{1}{n}$$

$$\Rightarrow \Pr\left[\forall \alpha : \sum_{t=1}^{T} X_{\alpha,t} < \gamma \cdot 0.2 \cdot T\right] \ge 1 - \frac{1}{n}$$

good event

→ under this event: $\forall \alpha$, if we trace the sub-arrays containing α inside the recursion tree of the quicksort execution, at least $T - \gamma \cdot 0.2 \cdot T = (1 - \frac{\gamma}{5})T$ of these sub-arrays shrunk by a 10%-90% or better split

hence # elements of subarray containing α at depth $t \le \left(\frac{9}{10}\right)^{(1-\frac{\gamma}{5})T} \cdot n = \left(\frac{9}{10}\right)^{(1-\frac{\gamma}{5}) \cdot c \cdot \log n} \cdot n$

$= n^{(1-\frac{\gamma}{5}) \cdot c \cdot \log 0.9} \cdot n < 1$

as long as $(1-\frac{\gamma}{5}) \cdot c \cdot \log 0.9 < -1$

e.g. $c = 30$, $\gamma = 2$ works

[< 1 means that there is no subarray as it can't contain any elements]

$\Rightarrow \forall \alpha$: no subarray containing α at depth T (i.e. α became lonely at an earlier depth)

$\Rightarrow$ i.e. quicksort finished earlier than depth $T = c \cdot \log n$

$\Rightarrow$ total runtime $O(n) \times T = O(n \cdot \log n)$

choice of constants c and γ: for the above to go through the constants need to satisfy:

$$\left.\begin{array}{l} c \cdot (\gamma - 1) \geq 30 \\ (1 - \frac{\gamma}{5}) \cdot c \cdot \log_e 0.9 < -1 \end{array}\right\} \text{ setting } c = 30, \gamma = 2 \text{ satisfies both}$$

with this choice of constants:

i. the good event happes with probability $\geq 1 - \frac{1}{n}$
ii. under good event quicksort takes $O(n \cdot \log n)$

$\boxtimes$

Final Remark: Is there anything special with the choice of a 10%-90% split?

A: No using 1%-99% or 49%-51% or more generally θ%-$(1-\theta)$% for any constant θ would also work with exact same proof as long as we choose c to be a large enough constant

Back To Chernoff Bound

Chernoff: Suppose $Y_1, Y_2, \ldots, Y_n \in [0,1]$ are independent random variables. Then $\forall \delta \in (0,1]$:

$$\Pr\left[\sum Y_i > (1+\delta)\mathbb{E}\left[\sum Y_i\right]\right] \leq e^{-\delta^2 \mathbb{E}\left[\sum Y_i\right]/3}$$

and $$\Pr\left[\sum Y_i < (1-\delta)\mathbb{E}\left[\sum Y_i\right]\right] \leq e^{-\delta^2 \mathbb{E}\left[\sum Y_i\right]/2}$$

Proof Of Chernoff:

- for convenience set: $p_i = \mathbb{E}Y_i$, $i = 1, \ldots, n$

$$\mu = \sum_{i=1}^{n} p_i$$

$$p = \frac{\mu}{n}$$

$$Y = \sum Y_i$$

- we'll only bound $\Pr[Y \geq (1+\delta)\mu]$ (upper tail)
  the lower tail is treated similarly

◇ proof compises basic idea + some calculus

$$\Pr[Y \geq (1+\delta)\mu] = \Pr[t \cdot Y \geq t(1+\delta)\mu] \quad \text{for all } t>0$$

$$= \Pr[e^{t \cdot Y} \geq e^{t(1+\delta)\mu}]$$

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

Markov's Inequality:
If X is positive random variable then for all $a>0$:
$\Pr[X > a] \leq \frac{\mathbb{E}[X]}{a}$

$$= e^{-t(1+\delta)\mu} \cdot \mathbb{E}[e^{t \cdot \sum Y_i}]$$

$$= e^{-t(1+\delta)\mu} \cdot \mathbb{E}\left[\prod_i e^{t Y_i}\right]$$

$$= e^{-t(1+\delta)\mu} \cdot \prod_i \mathbb{E}[e^{t Y_i}]$$

independence

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

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

arithmetic mean-geometric mean inequality
$\frac{\alpha_1 + \alpha_2 + \cdots + \alpha_n}{n} \geq \sqrt[n]{\alpha_1 \cdot \alpha_2 \cdots \alpha_n}$

$$= e^{-t(1+\delta)\mu} \cdot \left(e^t \cdot p + 1 - p\right)^n \qquad (\star\star)$$

Bound (**) true for all $t<0$. Minimized by setting

$$t = \log \frac{(1+\delta)(1-p)}{(1-p-\delta p)}$$

and the bound becomes:

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

(a.k.a. the entropic form of Chernoff bound)

where $H_p(x) = x \cdot \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$

is the "relative entropy of x with respect to p"

$$\leq e^{-\delta^2 \mu/3}$$

with a little calculus

use: $\ln(1+x) < x, \ \forall x > 0$

$\boxtimes$

Example:

- Suppose $Y_1, Y_2, \ldots, Y_n$ are independent outcomes of fair coin, i.e. $\mathbb{E}Y_1 = \mathbb{E}Y_2 = \ldots = \mathbb{E}Y_n = 1/2$
- $Y = \sum Y_i$: total number of heads
  $\mathbb{E}Y = n/2$, $\text{Var}\, Y = \frac{n}{4}$, $\text{stdev} = \sqrt{n}/2$
- Q1: Probability Y is a constant factor away from its mean?

A1: Directly from Chernoff

$$\Pr\left[|Y - \mathbb{E}Y| > \delta \cdot \mathbb{E}Y\right] \leq 2 \cdot e^{-\frac{1}{3}\delta^2 \frac{n}{2}}$$

exponentially small probability

$$\Rightarrow \text{with probability} \geq 1 - 2e^{-\frac{1}{6}\delta^2 \cdot n}, \quad Y = \mathbb{E}Y \cdot (1 \pm \delta)$$

often say w.v.h.p (with very high probability) for events that hold with probability $1 - \text{exponentially small}$ such as this one

- Q2: Probability Y is a few standard deviations away from its mean?

$$\forall \delta: \Pr\left[|Y - \mathbb{E}Y| > \delta \cdot \frac{\sqrt{n}}{2}\right] =$$

$$\Pr\left[|Y - \mathbb{E}Y| > \frac{\delta}{\sqrt{n}} \cdot \frac{n}{2}\right] \leq 2 \cdot e^{-\frac{1}{3}\left(\frac{\delta}{\sqrt{n}}\right)^2 \frac{n}{2}} = 2 \cdot e^{-\frac{1}{6}\delta^2}$$

by Chernoff

setting $\delta = 1$ gives $2 \cdot e^{-1/6}$

2 gives $2 \cdot e^{-4/6}$

3 gives $2 \cdot e^{-9/6}$

$\vdots$

$\delta = \sqrt{c \cdot \log n}$ gives $2 \cdot e^{-\frac{1}{6} c \cdot \log n} = 2 \cdot \frac{1}{n^{c/6}}$

polynomially small prob.

$\Rightarrow$ with prob. $\geq 1 - \frac{2}{n^{c/6}}$, $Y = \mathbb{E}Y \pm \frac{\sqrt{n}}{2} \cdot \sqrt{c \cdot \log n}$

often say w.h.p (with high probability) for events that hold with probability $1 - (\text{polynomially small})$

Conclusion: If n coins are tossed, the #heads is:

- $\frac{n}{2} \pm O(\sqrt{n \cdot \log n})$, with high probability
- $\frac{n}{2} \pm 0.1 \cdot n$, with very high probability

can replace this with any constant and will still have w.v.h.p.