# RANDOMIZED ALGORITHM

## Motivating Example:

Dont care about average case, we only care about the "pessimistic" case

**Example 1:** Toss a coin, how many times should you toss to get a tail ?

$$\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \dots = 2$$

$\Rightarrow$ 2 times on average

But if flips the coin 20 times, then we can almost certain to get at least 1 tail :

$$\left(\frac{1}{2}\right)^{20} \sim \frac{1}{10^6} \dashrightarrow \text{failure}$$

$$\Rightarrow 1 - \frac{1}{10^6} \dashrightarrow \text{success}$$

$\Rightarrow$ Goal is to give a k tries that makes prob. of success close to 1

**Example 2:** X or smaller

An interactor program repeatedly output either :

- integer X with prob 30%
- smaller integer ( < X) with prob 70%

Decide when to stop the interactor and guess the number correctly.

$$P(<X) = 0.7$$

After large k tries :

$$P(<X) = 0.7^k \approx 0$$

$$\Rightarrow P(X) \approx 1$$

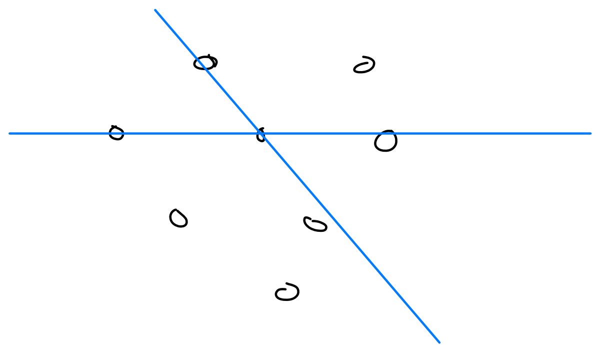
At this point you can almost certain that the output is X

## Example 3: Line through N/4 points

Given N points, find line pass through at many points as possible.

The answer is at least N/4 points.

(When given answer hint that is big, that might mean randomized algorithm)



**Brute force:**

For every pair of points (a line),
count number of points on that line
Return line with maximum points recorded.
$\Rightarrow \quad O(2^n)$

**Randomized:**

Let $\ell^*$ be the line we looking for.

$$P(\text{point} \in \ell^*) \leqslant 1/4$$

$$\Leftrightarrow \quad P(\text{line} = \ell^*) \leqslant 1/4 \cdot 1/4$$

$$= 1/16$$

$$\Leftrightarrow \quad P(\text{line} \neq \ell^*) > 15/16$$

Consider this routine: "Pick random 2 points to create a line, count the points on that line repeatedly for large k times". Then:

$$P(\text{line} \neq \ell^*) > (15/16)^k$$

$$\approx 0$$

At this point:

$$P(\text{line} = \ell^*) \approx 1$$

Return the line with maximum points recorded.

## Example 4: Greatest common divisor (GCD)

Max GCD of at least $N/2$ of given $N$ numbers

**Randomized:**

Let $\begin{cases} X = x_1, x_2, x_3, \ldots, x_n & \text{be the } n \text{ numbers} \\ X^* \text{ be the solution set s.t.} & |X^*| \geq |X|/2 = N/2 \end{cases}$

Consider this subroutine: For each $x_i$:

- Pick a random number $x_i \in X$
- If $x_i \in X^*$, then one of its GCDs is the max GCD
- $P(x_i \in X^*) \geq 1/2 \Rightarrow P(x_i \notin X^*) < 1/2$
- Repeatedly perform this subroutine for large $k$ times:

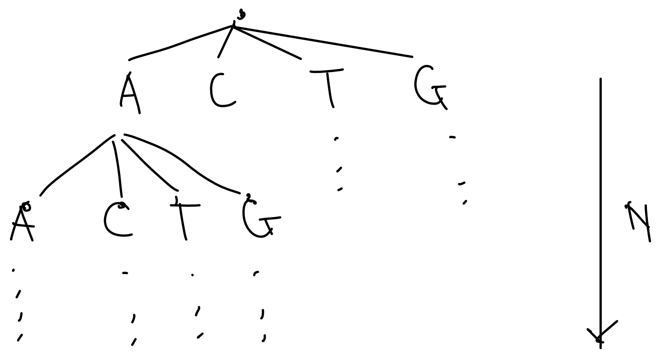
$$P(x_i \notin X^*) < (1/2)^k$$
$$\approx 0$$
$$\Rightarrow P(x_i \in X^*) \approx 1$$

Return the max GCD recorded so far

## Example 5: ACTG prefix

Guess a hidden string $S$ with characters ACTG. You can ask if something is a prefix of $S$. $N$ = length of $S$

- **Brute force:**




$\Rightarrow O(4^N)$

- **Randomized:**

Let $\begin{cases} S = c_1 c_2 c_3 \ldots c_N & \text{be string} \\ S^* = c_1^* c_2^* c_3^* \ldots c_N^* & \text{be the solution string} \end{cases}$

- Subroutine: For each $c_i$ $\forall\, 1 \leq i \leq N$
  - Assign A, C, T, G to $c_i$ randomly

    $P(c_i = c_i^*) = 1/4 \Leftrightarrow P(c_i \neq c_i^*) = 3/4$
  - Repeat above step for at most 4 times

$$P(c_i \neq c_i^*) = (3/4)^4 \approx 0$$
$$\Rightarrow P(c_i = c_i) \approx 1$$

$\Rightarrow O(N \cdot 4) \Rightarrow$ Average $\Theta(N \cdot 2.5)$

# ESSENTIAL TOOLS FOR RANDOMIZED ALGORITHM

## Linear of Expectation

For any r.v.s $X$, $Y$ and constant $c$:

- $E(X + Y) = E(X) \, . \, E(Y)$
- $E(cX) = c\,E(X)$

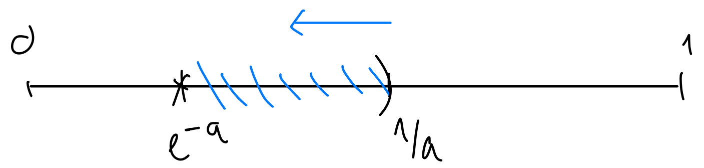
## Markov Inequality

If $X \geq 0$ and $a > 0$ $\Rightarrow$ $P(X \geq a) \leq \frac{E(X)}{a}$

Example: $X \sim Exp(\lambda = 1)$ : $P(X \geq a) \leq$ ?

Markov inequality : $P(X \geq a) \leq \frac{E(X)}{a}$

$$= \frac{1}{a}$$

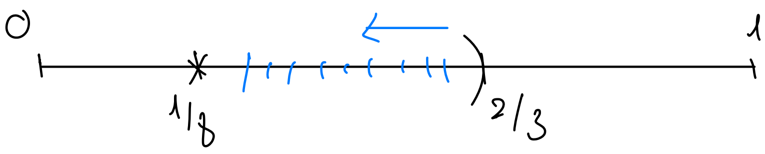



Example: $X \sim$ Uniform$([-4, 4])$ : $P(X \geq 3) \leq$ ?

Since $X$ can be negative, we use $|X|$ instead, Markov inequality:

$\Rightarrow P(|X| \geq 3) \leq \frac{E(|X|)}{3}$

$$= \frac{2}{3}$$




## Chebyshev Inequality

"If the variance is small, then $X$ is unlikely to be too far from the mean"

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

Proof: $P(|X - \mu| \geq c) = P((X-\mu)^2 \geq c^2)$

$\leq \frac{1}{c^2} E[(X-\mu)^2]$ (Markov inequality)

$= \frac{1}{c^2} \sigma^2$

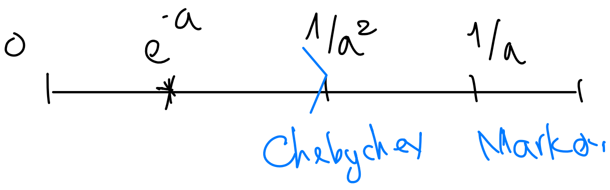
Example (tighter bound than Markov)

Consider $X \sim Exp(\lambda = 1)$. We know that using Markov inequality we can obtain this bound:

$$P(X \geq a) \leq \frac{1}{a}$$

Now using Chebyshev inequality, assume $a > \mu = 1$

$$\begin{aligned} P(X \geq a) &= P(X - \mu \leq a - \mu) \\ &= P(X - 1 \leq a - 1) \\ &\leq P(|X-1| \geq a-1) \\ &\leq \frac{1}{(a-1)^2} \approx \frac{1}{a^2} \end{aligned}$$




$\Rightarrow$ Chebyshev gives better bound

## Chernoff Bound:

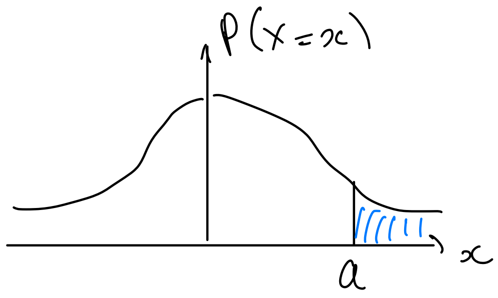
Given $X \sim Bin(\mu, \sigma^2)$, for any $t > 0$:

$$P(X \geq a) \leq e^{-at} \cdot M_x(t)$$ $\rightarrow$ momentum func

$$= e^{-at} \cdot e^{\mu t + \frac{\sigma^2}{2} t^2}$$

$\Rightarrow$ The lower $t$, the tighter the bound

Example: $X \sim Bin(0,1)$, $P(X \geq a) \leq ?$




Chernoff bound:

$$P(X \geq a) \leq e^{-at} \cdot e^{t + \frac{1}{2}t^2}$$

Pick $t = a$, then:

$$P(X \geq a) \leq -\frac{a^2}{2}$$

tightest bound