Chap 1 Power of randomization

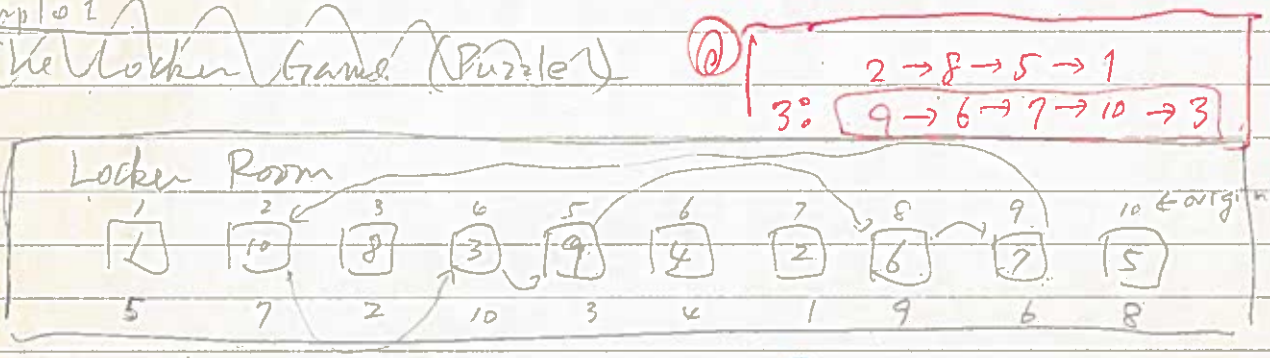
1.1 the Locker Game

Chap 6

Power of Randomization only when used appropriately

P
(P) ⇄ (V)
(A) hiding info.

Example 1
The Locker Game (Puzzle)



name 1 ~ 10

9 Team Game. A team of 10 people + Referee

- Each locker has a name of a player
- Each player enters the room one by one
- He/She can open at most 5 (½ of the whole doors) lockers
- If he/she finds her own name, she wins. ... successful.
- If All (10) players win! ... successful ⇒ the Team wins

using some alg.

② The team can discuss its strategy before the game, but once it started, no communication is allowed

notes, computers, ...

otherwise, player 1 opens the first 5 lockers, ...

our evaluation is for the worst case.

Strategy 1 All players open the first 5 lockers

Strategy 2 Even (odd) number players open even (odd) number lockers. worst-case

All deterministic strategies work.

Strategy 3 Each player opens 5 lockers at random

$(\frac{1}{2})^{10} = 1/1024 > 0$

randomization works

but # of lockers increases, then ---

Strategy 4. "shared randomness"

~~A random permutation, say 5,7,2,10,3,4,1,9,6,8,~~
is shared by all the players. Then --

Analysis of the winning (losing) probability

Shared randomness ⇒

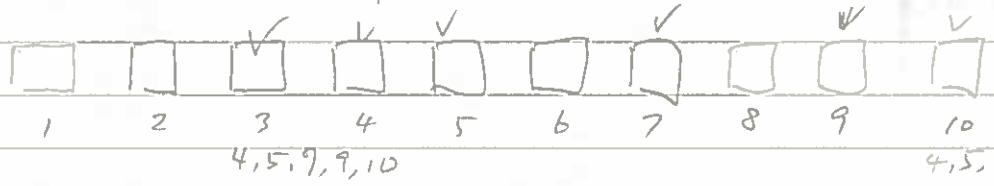
Names inside ≡ a random permutation of 1–10.

The players lose if ∃ a cycle with length 6 or more

Let's calculate its probability or
# of permutations having such a cycle
$L(6), L(7), \cdots, L(10)$

Case 1 : length 6

To fix one such permutation



(1) Fix 6 positions $\binom{10}{6}$

(2) Fix the order of the cycle $5 \times 4 \times \cdots \times 1$

(3) The other 4 numbers $4!$

$$L(6) = \binom{10}{6} \cdot 5! \cdot 4! = \frac{10!\,5!\,4!}{6!\,4!} = 10!/6$$

Case 2 Length 7 $L(7) = 10!/7$ ...

The losing prob $= \frac{1}{6} + \frac{1}{7} + \cdots + \frac{1}{10} \doteq 0.645$

The winning prob. ≥ 0.35 (random choice)

(1/1024 for a simple)

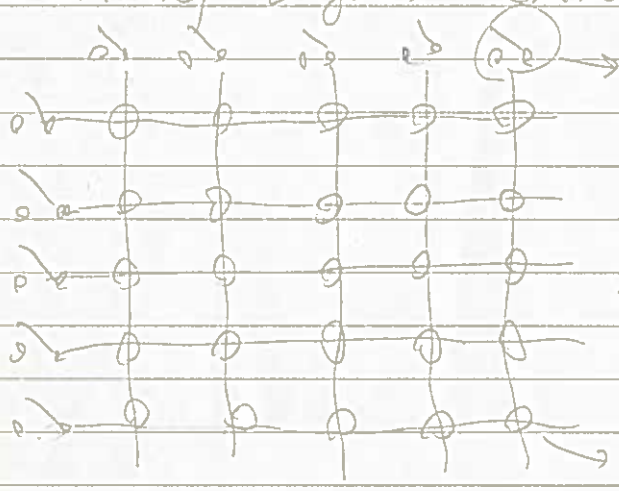
For a general $2n$ lockers $\frac{1}{n+1} + \cdots + \frac{1}{2n} \approx \int_n^{2n} \frac{1}{x}\,dx = \ln 2 < 0.7$

ⓒ proper use of randomness

No.

( )

1.2 Unbalancing Lights Game



switch
push: flips on/off of the lights on the column

n × n mesh

light

Game: 100 × 100 mesh

prize money (# of on lights − # of off lights) × 10 czk

Game change: 1,000 czk

you can manipulate the switches for 10 minutes

@ Initial settings are important.

Thm 1.1 For any $a_{ij} = \pm 1$ $(1 \le i, j \le n)$, we can select the values of x_i, y_j $(= \pm 1)$ such that

$$\sum_{ij} a_{ij} x_i y_j \ge \sqrt{\frac{1}{2\pi}}\, n^{\frac{3}{2}} \quad (\approx 4{,}000 \text{ if } n = 100)$$

$\therefore \approx 0.4$

a_{ij} : initial on/off of the lights (+1: on, −1: off)

x_i, y_j : −1 : push the switch

row column +1 : does not

No.

(proof) Consider the following randomized alg.

Select $y_1, \cdots, y_n$ $(=\pm 1)$ at random

$y_1\ y_2\ \cdots\ y_n$

x_i | 0 0 ⋯ 0 |

±1 " "

Let $R_i = \sum_{j=1}^{n} a_{ij} y_i$ $\quad R = \sum_{i=1}^{n} |R_i|$

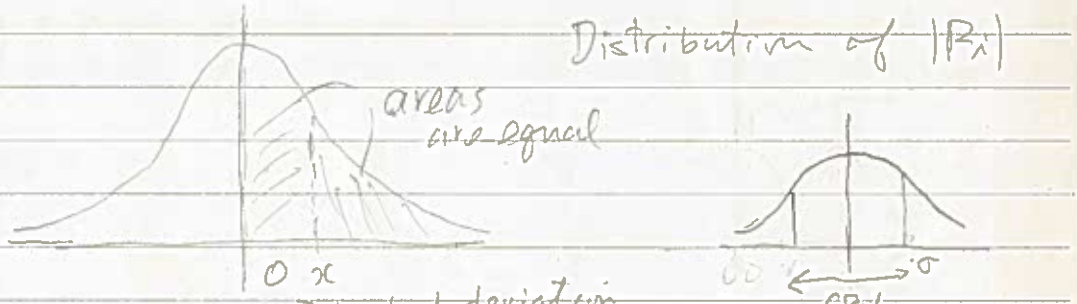
↳ sum of the values (± 1 at random) in row i

$= n$ values of fair coin

(head = +1, Tail = −1)

⇓

It follows the binomial distribution



mean deviation

$E[|R_i|] \doteq \sqrt{\frac{2}{\pi}}\,\sigma = \sqrt{\frac{2}{\pi}}\sqrt{npq} = \sqrt{\frac{1}{2\pi}}\sqrt{n}$ ← standard deviation

$E[R] = \sum_{i} E[|R_i|] = \sqrt{\frac{1}{2\pi}}\sqrt{n}\cdot n$

$\Rightarrow \exists$ selection of $y_1, \cdots, y_n$ that achieves $|R_i| \geq \sqrt{\frac{1}{2\pi}}\sqrt{n}$

Now set $x_i = -1$ iff $R_i < 0$. #

Note that we can use this "algorithm" whose success probability is very high

→ play it !

NO.

DATE

1.3 Probabilistic Methods

Note that Th. 1.1. has nothing to do with probability

The probability that an event E happens > 0

$\Rightarrow$ E can actually happen.

The " " " does not < 1.0

$\Rightarrow$ E can actually happen

Usually we design a probabilistic algorithm A that makes E happen (if lucky)

If the success prob of A > 0, ---

If the failure " " A < 1.0, ---

Namely, we have more than 0.35 as the winning prob.

If we have more lockers,

$$\frac{1}{n}+\frac{1}{n+1}+\cdots+\frac{1}{2n} \approx \int_n^{2n} \frac{1}{x}\,dx = \ln 2 \leq 0.7$$

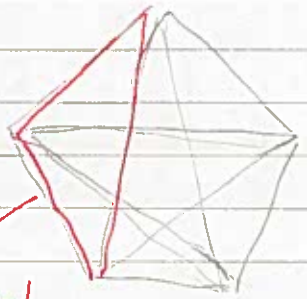
We still have more than 0.3.

Example

① Ramsey Number

② Edge Coloring

K_5 : a complete graph with 5 vertices

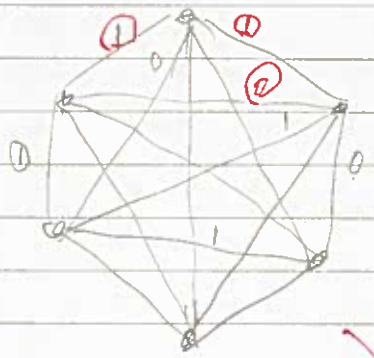


not allowed

Color each edge with "blue" or "red", so that no triangle of the same color appears.

Possible ? Yes !

最初に現れる3



K_6 : Do the same thing.

Possible ? NO !

Prove this

↳ Then obviously NO for $K_7, \ldots$

$R(k,h)$: The minimum n such that for any coloring of K_n, there must exist a single ← colored K_k

↳ no successful coloring exists

$R(3,3) = 6$

Triangle = K_3

↓ 6 people ∋ 3 people all know each other or mutually unknown (no one knows each other)

↳ $n=5$ ∋ successful coloring

$n=6$ ∌ " "

$R(6,6) > 8$: K_8 can be colored without monochromatic K_6

No.

( )

↓ nothing to do with probability, but ...

"Another power of randomization"

Thm 1.2 $R(k,k) > 2^{k/2}$ for all $k \geq 3$

We prove: $n = 2^{k/2} \Rightarrow \exists$ successful coloring (Goal: avoid a single-colored K_k)

Proof Consider the following simple algorithm and its analysis

- Just color each edge of K_n with red or blue uniformly at random

$\Rightarrow$ Compute the probability P that this algorithm fails, i.e., that a single-colored K_k appears

$\Rightarrow$ It turns out that $P < 1.0$ if $n = 2^{k/2}$ if $n = 2^{k/2}$

$\Rightarrow$ E = "A single-colored K_k appears by random coloring" ith

$\Rightarrow$ The probability that E happens is < 1.0

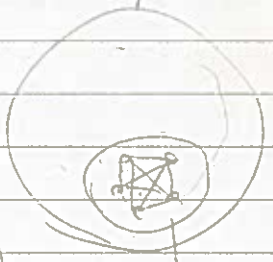
$\exists$ coloring E does not happen

① Computing probability P.

After the random coloring, the probability that a fixed K_k is single-colored is

$$\left(\frac{1}{2}\right)^{\binom{k}{2}} \cdot 2 = 2^{1-\binom{k}{2}}$$

(red or blue)



The # of ways of selecting K_k is $\binom{n}{k}$, an arbitrary K_k

so the prob. that at least one single-colored K_k appears is at most

$$\binom{n}{k} 2^{1-\binom{k}{2}}$$

$\Rightarrow$ $P \leq \binom{n}{k} 2^{1-\binom{k}{2}}$

② This vale < 1 if $n = 2^{k/2}$

Union bound event $A_1, \cdots, A_m$

$P(A_1 \cup \cdots \cup A_m) \leq P(A_1) + \cdots + P(A_m)$

No.

$$1-\binom{k}{2} = 1-\frac{k(k-1)}{2} = -\frac{k^2}{2}+\frac{k}{2}+1$$

$$\binom{n}{k}2^{1-\binom{k}{2}} = \frac{n!}{k!\,(n-k)!}\cdot 2^{-\frac{k^2}{2}+\frac{1}{2}(k+2)}$$

$$\leq \frac{n^k}{k!}\,2^{-\frac{k^2}{2}+\frac{1}{2}(k+2)}$$

$$= \frac{2^{\frac{1}{2}(k+1)}}{k!}\cdot\frac{n^k}{2^{\frac{k^2}{2}}}$$

$$\nwarrow \qquad \| \quad \text{if } n = 2^{\frac{k}{2}}$$

$$1 \qquad 1$$

$$< 1.0 \qquad //$$

In fact $P \ll 1.0$ (if k is large), namely The algorithm succeeds with high probability, ie, it is a good algorithm.

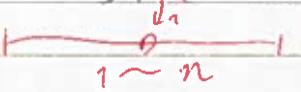
Define new random variables $X_i(\sigma)$

$$X_i(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = i \\ 0 & \text{otherwise} \end{cases}$$

(2, 1, 3, 4, 5) $X_1(\sigma) = X_2(\sigma) = 0$, $X_3(\sigma) = X_4(\sigma) = X_5(\sigma) = 1$

Obviously $X(\sigma) = X_1(\sigma) + X_2(\sigma) + \cdots + X_n(\sigma)$.

Also, $X_i(\sigma) = 1$ with prob $1/n$.

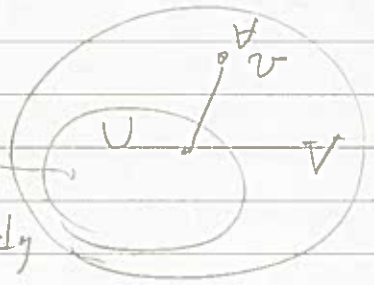


So, $E[X(\sigma)] = E[X_1(\sigma)] + E[X_2(\sigma)] + \cdots + E[X_n(\sigma)]$

$= 1/n + \cdots + 1/n = 1$.

② Dominating set of a graph $G = (V, E)$

$U \subseteq V$ is a dominating set iff for any $v \in V - U$, there is a vertex $u \in U$ that is adjacent to v.



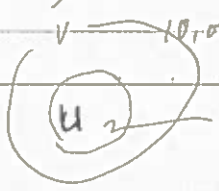
Set of representatives s.t. everybody in V is known by somebody in U.

Thm 1.3 Suppose that $G = (V, E)$ has a minimum degree of δ (≥ 1). Then there exists a dominating set whose size is at most $n[1 + \ln(\delta+1)]/(\delta+1)$

nothing to do with probability

$n = 10{,}000$, $\delta = 100$ ⇒ this value ≈ 455

Proof Trick: not evaluating the value directly.



Consider a random vertex set of $n\frac{\ln(\delta+1)}{\delta+1}$

Let $p = \ln(\delta+1)/(\delta+1)$.

For each $v \in V$, put it into X with prob. p.

Let Y be a set of vertices $v \in V-X$ such that v is not adjacent to any vertex u in X. (※)

$E[|X|] = np$

The prob. that a vertex $v \in V$ is in Y

$= \Pr[$neither v nor its neighbors comes into $X]$

$\leq (1-p)^{\delta+1}$

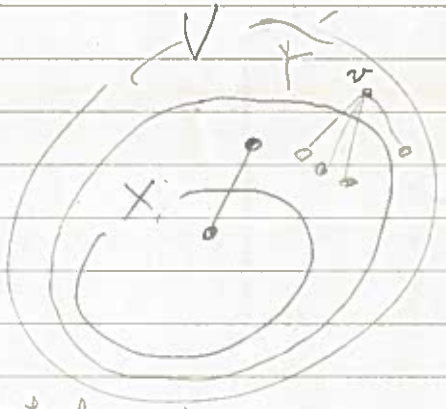
$\Downarrow$

$E[|Y|] \leq n(1-p)^{\delta+1}$

$E[|X|+|Y|] \leq np + n(1-p)^{\delta+1}$

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

$\leq n\cdot\frac{1+\ln(\delta+1)}{\delta+1}$



$\rightarrow \left(1-\frac{\ln k}{k}\right)^k \leq \frac{1}{k}$

using $\ln(1-x) \leq -x$

This means there exist X and Y such that $|X|+|Y| \leq n\cdot\frac{1+\ln(\delta+1)}{\delta+1}$ they satisfy (※)

and it is easy to obtain such ones.

Note that $X \cup Y$ is a dominating set. //

BPP : bounded-error probabilistic polytime

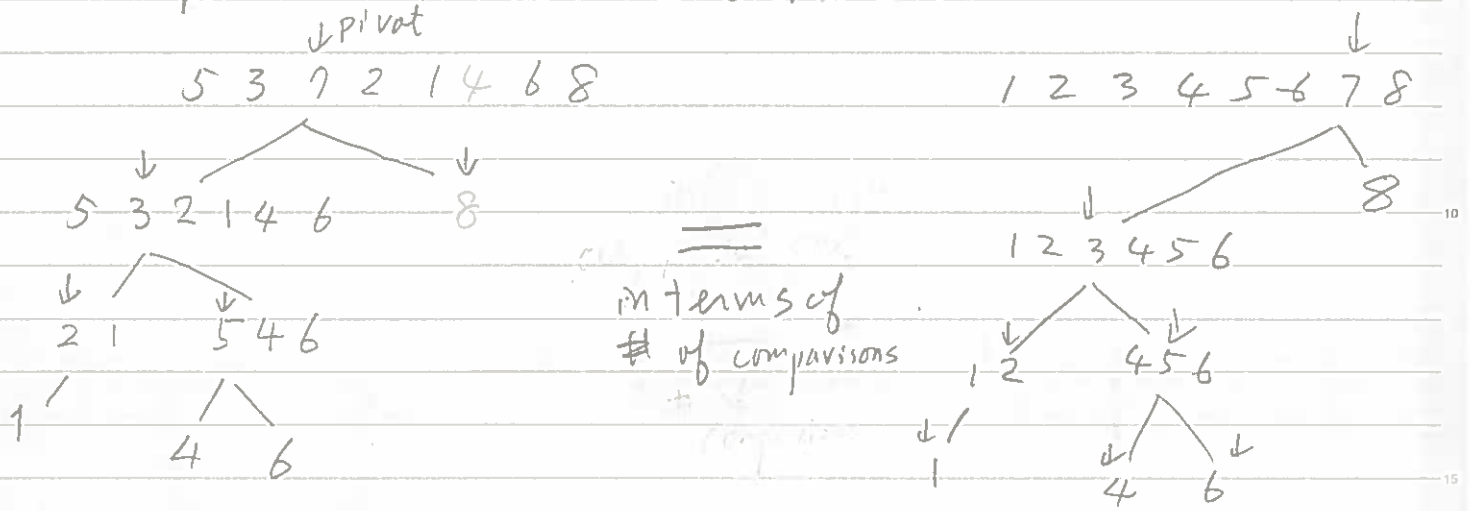
NO.

DATE

1.4 How to prove the success prob. of randomized Algorithm

@ Randomized algs are usually VERY simple
but not easy to prove their performances

Example 1 : Randomized Quick Sort



@ Complexity = # of comparisons

pivot sequence
= 7 3 8 2 5 1 4 6
looks like a random permutation

Th. 1.4 # of comparisons is $O(n \log n)$ whp. → next page

Consequently:

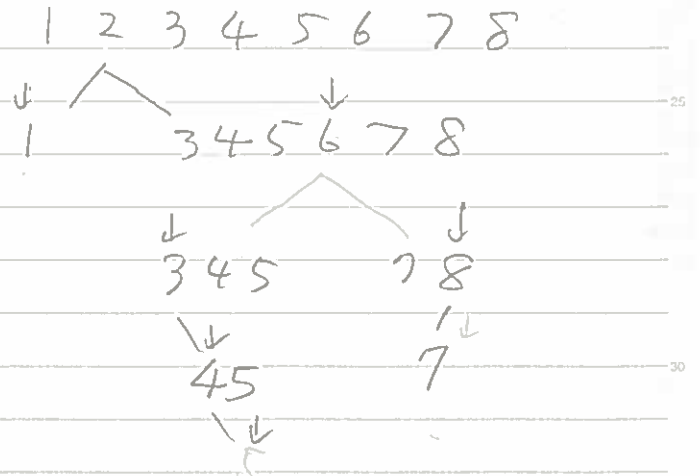
A (random) permutation
$\sigma = 2\ 6\ 1\ 8\ 3\ 4\ 5\ 7$

⇓

determines the execution of Q.S.

⇓

determins # of comparisons $7 + 5 + 3 + \cdots$



NO.

DATE

Simple argument such as ---

if a random pivot falls in this interval, then the length of unsorted sequence becomes at most 2/3

The prob. that happens is 1/3

⇓

Do this $9 \log n$ times ⇒ Above event happens at least $2 \log n$ times w.h.p.

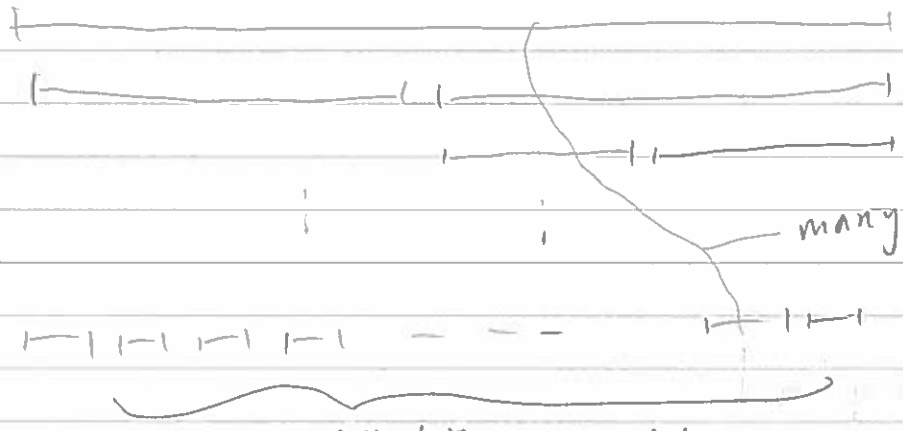
⇓

$\left(\frac{2}{3}\right)^{2 \log n} \approx n^{-1.1}$

⇓

The above length becomes 1 w.h.p.

However, there are many "intervals"



All of them must become single elements.

NO.

DATE

$T(\sigma)$: # of comparisms determined by σ

Random Variable $X_{ij}(\sigma) = \begin{cases} 1 : i \text{ and } j \text{ compared under } \sigma \\ 0 : \text{ — . not — —} \end{cases}$

i j

1 2 3 4 5 6 7 8

$\sigma = 2\ 6\ 1\ 8\ 3\ 4\ 7\ 5$

1 | 3 4 5 6 7 8

3 4 5 | 7 8

Not compared

$\sigma = 2\ 7\ 1\ 8\ 3\ 4\ 6\ 5$

1 | 3 4 5 6 7 8

3 4 5 6 | 8

Compared

~~k : the first number in σ s.t. $i \le k \le j$~~

$k = i$ or j $\Rightarrow$ i and are compared

$i < k < j$ $\Rightarrow$ " NOT "

$$E(T(\sigma)) = E\left(\sum_{i<j} X_{ij}(\sigma)\right)$$

$$= \sum_{ij} E(X_{ij}(\sigma)) = \sum_{ij} \text{prob}(X_{ij} = 1)$$

$$= \sum_{ij} 2/j-i+1 \quad (k \text{ can be this many})$$

the first k is equal prob. for $i, i+1, \cdots, j$

$$= 2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) + 2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1}\right) + \cdots$$

$i=1,\ j=2,3,\cdots$ $\quad$ $i=2,\ j=3,4,\cdots$

$$\le 2n \ln n = \sout{O(n \log n)} \qquad 1 + \frac{1}{2} + \cdots + \frac{1}{n} \times \ln n$$

Q Markov Bound $\rightarrow$ $O(n \log n)$ whp

NO.

DATE

Example 2 : 3SAT

Input : 3CNF formulan

$f = (x_1 \vee \bar{x}_3 \vee x_4)(\bar{x}_2 \vee x_4 \vee x_5)(\ )\cdots(\ )$

Question : Is f satisfiable ?

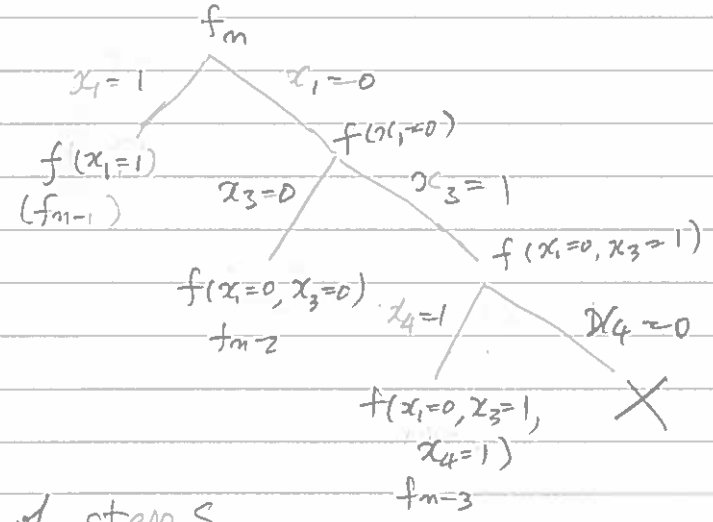
$\exists$ assignment making f true

check $(x_1, \cdots, x_n)$ from $(0, 0, \cdots, 0)$

to $(1, 1, \cdots, 1)$

$\Rightarrow 2^n$ steps

But many "good" algorithms running faster



$T(n)$: # of steps

$T(n) = T(n-1) + T(n-2) + T(n-3)$

$\Downarrow$

$T(n) = 1.84^n \ll 2^n$

$T(n) = 2T(n-1) \Rightarrow T(n) = 2^n$

$C^n = C^{n-1} + C^{n-2} + C^{n-3}$

$C^3 = C^2 + C + 1$

1.5^n

No.

$(1.334^n$

1.321^n

Local Search for 3SAT. [Schoening 99]

Input $f = (x_1 \vee x_2 \vee x_3)(\ )(\ ) \cdots (\ )$

$\vec{x} = (a_1, a_2, \cdots, a_n) \leftarrow$ random assignment

Repeat $3n$ times

if $f(a_1, a_2, \cdots, a_n) = 1$ Then output "yes"

otherwise

select an arbitrary false clause.

$(x_6 \vee \bar{x}_7 \vee x_9)$

$0 \quad 1 \quad 0$

$(1 \to 0$
$0 \to 1$

Select one of the three variables and flip it at random

Analysis of the success prob.

$\vec{x} = (1,0,0,1,1,0,1)$ satisfying assignment

$x = (0,0,0,0,1,0,0)$

Hamming distance

$k-1 \quad k \quad k+1$

↑ We are here

[Assume there is only one sat. assignment

by a single flip:

← with at least prob $1/3$

→ with at most prob. $2/3$ (why?)

$x_6 \vee \bar{x}_7 \vee x_9$

$0 \quad 1 \quad 0$ current

$1 \quad 1 \quad 0$ sat

$1 \quad 0 \quad 0$ or

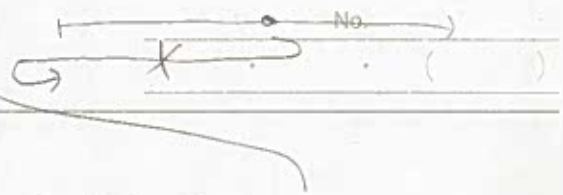
For example ($k=5$)

0 1 2 3 4 5 6 7

→ good direction 7 times &

bad " twice

$\sum_{i=0}^{n} q_k(i) = c^{-n}$ ⇒ repeat c^n times gives us a good success prob.



((we are successful))

This can happen with prob. $\binom{9}{2}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right)^2$ approximately

We are also successful
good: 9 times
bad: 4 times

In general: bad direction i times
good " " $k+i$ "

with prob. $q_k(i) = \binom{k+2i}{i}\left(\frac{1}{3}\right)^{k+i}\left(\frac{2}{3}\right)^i$

Summing up from $i=0$ to n ($\to \xi_{2n}$)

$\sum_{i=0}^{n} q_k(i)$ total success prob.

However, the value of $q_k(i)$ ~~changes~~ is exponentially small (~~due to the binomial coefficient~~), so it is

$\sum_{i=1}^{n}\binom{n}{i} \approx \binom{n}{n/2}$ approximately OK to consider a maximum ~~one~~ single term, i.e., $q_k(k)$ (ie, when $i=k$)

$$\sum_{i=0}^{n} q_k(i) \geq q_k(k) = \binom{3k}{k}\left(\frac{1}{3}\right)^{2k}\left(\frac{2}{3}\right)^k \approx \left(\frac{1}{2}\right)^k$$

(1) A useful formula

$$\binom{n}{\alpha n} \approx \left(\frac{1}{\alpha}\right)^{\alpha n}\left(\frac{1}{1-\alpha}\right)^{(1-\alpha)n}$$

Recall that we use a random assignment at the beginning of [illegible]

At position (Hamming dist.) k with prob

correct (O, O, O, O, O) $\binom{n}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \binom{n}{k}\left(\frac{1}{2}\right)^n$

initial (O x x O O)

From there we can get to the solution with prob. $\left(\frac{1}{2}\right)^k$

$n=3k$, $\alpha=1/3$

$\binom{3k}{k} = 3^k\left(\frac{3}{2}\right)^{2k}$

No. ( )

So the total success prob. =

$$\sum_k \binom{n}{k}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^k$$

$$\geq \binom{n}{n/3}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^{n/3} \quad (\max \text{ at } k=n/3)$$

$$=\left(\frac{3}{4}\right)^n$$

$1.334 \ll 2$

Namely, if we repeat the procedure $\left(\frac{4}{3}\right)^n$ times, we can get an answer with high prob.

$$\binom{n}{n/3} = 3^{\frac{n}{3}}\left(\frac{3}{2}\right)^{\frac{2}{3}n} \qquad 2^{0.416n}$$

$$\left(3^{\frac{1}{3}}\right)^n\left(\left(\frac{3}{2}\right)^{\frac{2}{3}}\right)^n\left(\frac{1}{2}\right)^n\left(\left(\frac{1}{2}\right)^{1/3}\right)^n$$

$$3^{\frac{1}{3}+\frac{2}{3}}\cdot\left(\frac{1}{2}\right)^{\frac{2}{3}} \qquad \left(\frac{1}{2}\right)^{4/3}$$

$$3 \qquad \left(\frac{1}{2}\right)^{-2}$$

$$3/4$$

Th 1.5 (Algorithm) Schoening solves 3SAT in time $O\left(\left(\frac{3}{4}\right)^n\right)$ whp.