Old notes from 2018. To be updated after class.

CS5330 Week 2

Jan 22 2018

Puzzle of the Day

- → 100 Prisoners $p_1, p_2, \ldots, p_{100}$
- → 100 Boxes $b_1, b_2, \ldots, b_{100}$
- → each prisoner's name is put in a random box

random permutation

Game:

- → With no communication, each prisoner looks in 50 boxes
- → Boxes left unchanged
- → Win if all prisoners find their own name

Can you win w.p. $\geq 1/4$?

Today: Algorithms — Quicksort, Treaps, Stable Matching

Tools — Linearity of Expectation, Conditional Expectation, Markov's Inequality, Coupon Collector, Balls-in-Bins

---

```
QuickSort(A[1,..,n], begin, end)
    if begin = end then return
    p = random(begin, end)
    j = Partition(A, p)
    QuickSort(A, begin, j-1)
    QuickSort(A, j+1, end)
```

Ex. [3, 7, 22, 1, 5, 8, 2, 9]

$p \xleftarrow{R} 6$, pivot $= 8$

Partition:

$$\underbrace{[3, 7, 2, 1, 5,}_{\text{recurse}} \overset{=j}{8}, \underbrace{22, 9]}_{\text{recurse}}$$

Standard Analysis: $E[T] = O(n \log n)$

Input: $A =$ permutation of $[a_1, a_2, \dots, a_n]$

$a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n$

Define: $X_{ij} = 1$ if compare a_i and a_j (at any point in sort)

$= 0$ otherwise

$T \leq O(\sum X_{ij})$ ← comparisons dominate
for each comparison, O(1) swaps
compare each pair at most once

$E[T] \leq O(E[\sum X_{ij}])$

$\leq O(\sum E[X_{ij}])$ ← linearity expectation
$E[A+B] = E[A] + E[B]$

$E[X_{ij}] = 1 \cdot \Pr[X_{ij} = 1] + 0 \cdot \Pr[X_{ij} = 0]$

$= \Pr[X_{ij} = 1]$

What is probability we compare a_i and a_j?

(assume wlog $i < j$)

$\cdots \; a_i \quad a_{i+1} \quad a_{i+2} \cdots a_{j-1} \quad a_{j-1} \quad a_j \; \cdots$

Only compare a_i, a_j if:

→ a_i or a_j is pivot

→ a_i and a_j are in same recursive sub-problem

In execution, which becomes pivot first?

1. $a_i \rightarrow X_{ij} = 1$
2. $a_j \rightarrow X_{ij} = 1$
3. $a_k : i < k < j \rightarrow X_{ij} = 0$

} some item in $[i, j]$ is first to become pivot

↳ after partition on a_k, a_i and a_j are in different subproblems

↳ never compared

Key Claim: Each item in $[a_i, .., a_j]$ is equally likely to become pivot first

Handwave: symmetry?

More careful: Until first pivot, all in same subproblem ← induction

Pivot chosen uniformly from subproblem

$$\Rightarrow \Pr(a_j \text{ is first pivot}) \leq \frac{1}{j-i+1}$$

← why ≤ ??

$$\Pr(a_i \text{ is first pivot}) \leq \frac{1}{j-i+1}$$

Conclusion: $\Pr(X_{ij}) \le \frac{1}{j-i+1}$

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = 2\sum_{i=1}^{n} \sum_{k=2}^{n-i+1} \frac{1}{k} \quad [\text{define } k = j-i+1]$$

$$= 2\sum_{k=2}^{n} \sum_{i=1}^{n-k+1} \frac{1}{k} \quad [\text{Reorder}]$$

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

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

$$= 2(n+1)(H_n - 1) - 2(n-1)$$

$$H_n = \sum_{j=1}^{n} \frac{1}{j} \le \log(n) + 1$$

$$\le 2(n+1)\log n - 2(n-1) \le 2n\log n$$

#comparisons in QuickSort

$$\int_1^{n+1} \frac{1}{j}\,dj \le H_n \le \int_1^{n} \frac{1}{j}\,dj + 1$$

$$\log(n+1) \le H_n \le \log n + 1$$

$$E[T(n)] = O\left(\sum_{i=1}^{n}\sum_{j=2}^{n} \Pr(X_{ij}=1)\right)$$

$$\leq O(n H_n) = O(n \log n)$$

Markov's Inequality

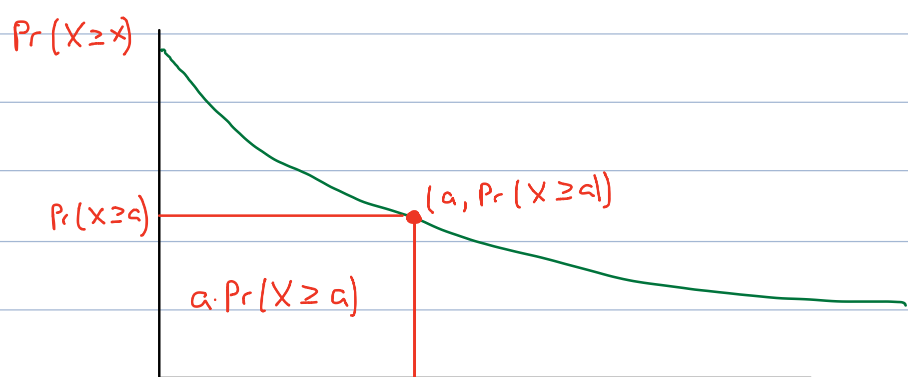
$$\text{For all } \underbrace{X}_{\text{non-negative random variable}} : \Pr(X \geq \underbrace{a}_{a>0}) \leq \frac{E[X]}{a}$$

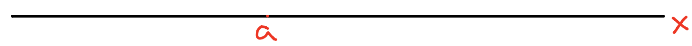
Ex: $\Pr(X \geq 2E[X]) \leq 1/2$ "At most 1/2 of the time you have $\geq$ twice the average"

I = event $X \geq a \Rightarrow aI \leq X$

$$E[aI] \leq E[X]$$

$$a \Pr(X \geq a) \leq E[X]$$





$$a \Pr(X \geq a) \leq \sum_{j=1}^{\infty} \Pr(X \geq j) = E[X]$$

$\Pr($QuickSort has $\geq 4n\log n$ comparisons$)$

$$\leq \frac{2n\log n}{4n\log n} \leq 1/2$$

Markov

Let Y = #comparisons QuickSort, $E[Y] \leq 2n\log n$

$$\Pr(Y \geq 2tn\log n) \leq 1/t$$

$$= \Pr(Y^2 \geq 4t^2n^2\log^2 n)$$

$$\leq \frac{E[Y^2]}{4t^2n^2\log^2 n}$$

See next page

$$\leq \frac{4n^2\log^2 n}{4t^2n^2\log^2 n}$$

$$\leq 1/t^2$$

"2nd moment method" $\longrightarrow$ Chebychev's Inequality

$$Y = \sum_{i=1}^{n} \sum_{j=i}^{n-1} X_{ij}$$

$$E[Y^2] = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=1}^{n} \sum_{l=k+1}^{n} \frac{2}{j-k+1} \cdot \frac{2}{k-l+1}$$

$$= \sum_{i=1}^{n} \sum_{j=2}^{n-i+1} \sum_{k=1}^{n} \sum_{l=2}^{n-k+1} \frac{2}{j} \cdot \frac{2}{l}$$

$$= \sum_{j=2}^{n} \sum_{l=2}^{n} \sum_{i=1}^{n-j+1} \sum_{k=1}^{n-l+1} \frac{4}{jl}$$

$$= 4 \sum_{j=2}^{n} \sum_{l=2}^{n} \frac{(n-j+1)(n-l+1)}{jl}$$

$$\leq 4n^2 \sum_{j=2}^{n} \sum_{l=2}^{n} \frac{1}{jl}$$

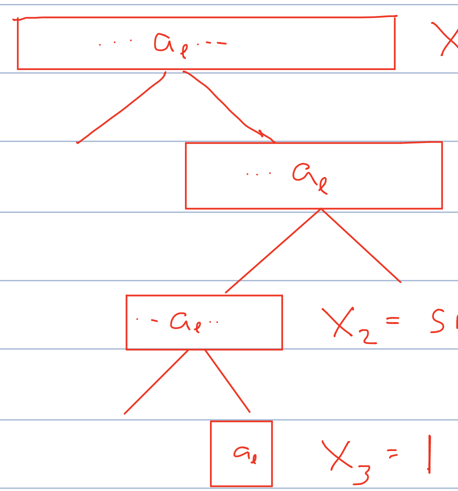
$$\leq 4n^2 \sum_{j=2}^{n} \frac{1}{j}(H_n - 1)$$

$$\leq 4n^2 (H_n - 1)^2 \leq 4n^2 \log^2 n$$

Another Approach

Fix an element a_ℓ.

Let X_j = Size of j^{th} subproblem containing a_ℓ.



$X_0 = n$

X_1 = Size of

X_2 = Size of ...

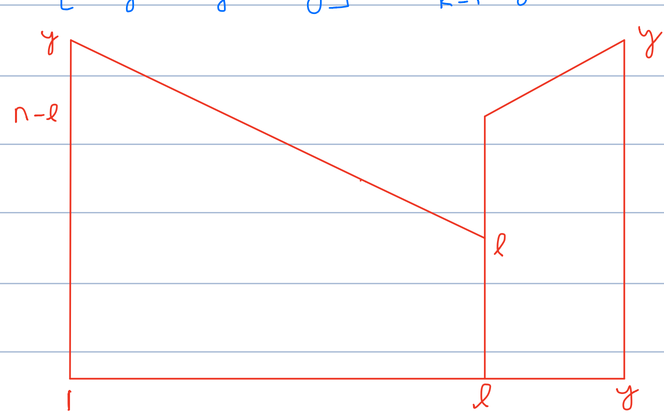
$X_3 = 1$

$X_4 = 0$

$E[X_j] = ??$

$$E[X_j \mid X_{j-1} = y] \le \sum_{k=1}^{y} \frac{1}{y} \cdot \max(k, y-k) \le \frac{1}{4} \cdot \left[y \cdot \frac{y}{2} + 2 \cdot \frac{1}{2} \cdot \frac{y}{2} \cdot \frac{y}{2} \right]$$

$$\le 3/4\ y$$



$$\boxed{E[X_j] = E_{X_{j-1}}\left[E_{X_j}[X_j \mid X_{j-1}]\right]}$$

$$= \sum_y \Pr(X_{j-1} = y)\, E[X_j \mid X_{j-1} = y]$$

$$\leq \sum_y \Pr(X_{j-1} = y)\left(\tfrac{3}{4} y\right)$$

$$\leq \tfrac{3}{4} E[X_{j-1}]$$

$$E[X_j] \leq \tfrac{3}{4} E[X_{j-1}] \leq \cdots \leq \left(\tfrac{3}{4}\right)^j E[X_0]$$

$$\leq \left(\tfrac{3}{4}\right)^j n$$

$$\Pr[X_j \geq 1] \overset{\text{Markov}}{\leq} \frac{E[X_j]}{1} \leq \left(\tfrac{3}{4}\right)^j n$$

$\Rightarrow$ Fix $j^* = (c+2)\log_{4/3} n \leq 3(c+2)\log n$

$$\Pr[X_{j^*} \geq 1] \leq \frac{1}{n^{c+2}} \cdot n \leq 1/n^{c+1}$$

Depth of recursion for a_ℓ is $\leq j^*$ wp $\geq 1 - \frac{1}{n^{c+1}}$

Union bound over all a_k:

$$\Pr[\text{any } a_k \text{ has depth} \geq j^*] \leq \sum_{k=1}^{n} \frac{1}{n^{c+1}}$$

$$\leq 1/n^c$$

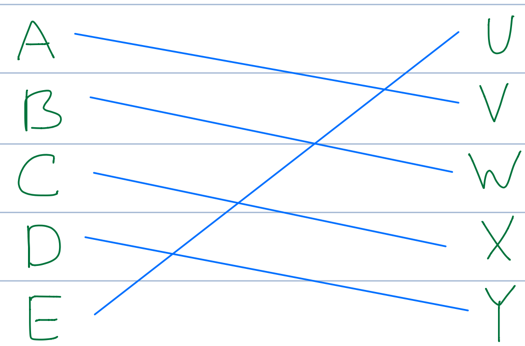
If QuickSort recursion is depth j^*, then total # comparisons $\leq n j^*$.

Theorem: # comparisons in QuickSort is $\leq 3(c+2)\, n \log n$ w.p. $\geq 1 - 1/n^c$

Corollary: $E[\#\text{ comparisons}] \leq (1-\frac{1}{n})\, 9 n \log n + \frac{1}{n} \cdot n^2$
$\leq 10 n \log n$

Stable Matching

n women   n men



Goal: find a matching

Lists of preferences:

A: V U X W Z Y

B: U W V Z X Y

⋮

U: C A B D E

V: A B D C E

⋮

Each list is a permutation of the available choices

Matching is <u>stable</u> if: for all matched pairs (a, x), (b, z)
either a prefers x to z
or z prefers b to a.

"No pair wants to cheat"

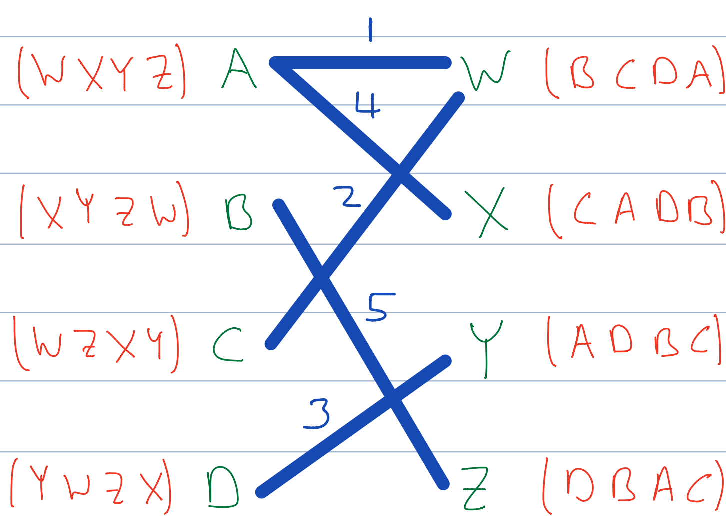
Is it always possible to find a <u>stable</u> matching?

YES! Gale-Shipley (1962) → Nobel Prize in Econ (2012)

Proof by Algorithm:

Repeat until all matched:

1. Fix A unmatched woman
2. A asks first man X on her list not already asked
3. If X unmatched, then match (A, X)
4. If (B, X) matched and X prefers A to B, then swap: Match (A, X)
5. Otherwise, continue.



Termination: $\leq n^2$ possible proposals in total

Everyone matched: if X rejects A, then X is matched to <u>someone</u> at end

if A is unmatched, then A asked n men.

So all n men are matched at end

$\Rightarrow$ contradicts A is unmatched

Stable: Assume not: $(\cdots Y \cdots X \cdots)$ A —— X

B —— Y $(\cdots A \cdots B \cdots)$

$\Rightarrow$ A asked Y before X

Y switched to $\cdots\cdots$ to B

So Y prefers B to A.

Average-case: assume each preference list is a random permutation

1) How fast?

2) How many get #1 choice?
How many get #K choice?

Step 1: Deferred Decisions

A: Choose random permutations for all players
Run algorithm

B: Run algorithm
Whenever query list, choose next random entry

Process A = Process B
Doesn't matter when we flip the random coins

Easier to analyze B...

Step 2: Independent Process

→ Choices not independent: next choice depends on
whether prior choices were
rejected
Also, accept/reject depends on
history.

New Algorithm: Ask man uniformly at random
→ if already asked ⇒ rejected
→ if not already asked ⇒ as before

Same outcome as original algorithm.
(may take longer)

Stochastic Domination

2 R.V.s X, Y

"$X \geq Y$" if $\forall k$: $Pr(X \geq k) \geq Pr(Y \geq k)$

If $X \geq Y$, then $E[X] \geq E[Y]$

Claim: if T is # proposals in new process
T' is # proposals in old process
then $T \geq T'$

[Typical proof: coupling argument... Postponed.]

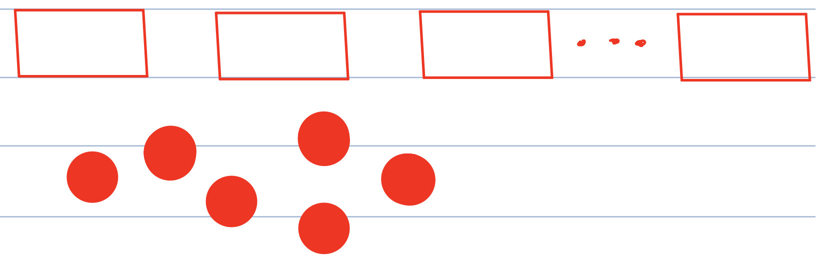
Observation: terminates when each man receives ≥ 1 proposal

To bound T: How many proposals until each man has received ≥ 1 ??

Balls-in-Bins

n bins

m balls



Game:

for each ball:

put it in a random bin

$\Pr(\text{ball } i \text{ in bin } j) = 1/n$

$X_{ij} = 1$ if ball i in bin j

$= 0$ otherwise

$E[\text{balls in bin } j] = \sum_{i=1}^{m} X_{ij} = \sum_{i=1}^{m} \frac{1}{n} = \frac{m}{n}$

Questions:

1) How many balls so each bin has ≥ 1

Stable matching

2) Expected max balls per bin?

Hashing

3) How many bins have exactly 1 ball?

Backoff protocols

And more...

Variants: Weighted balls

Weighted bins

rethrow balls

etc.

Cover all bins:

Let X = # balls when all bins have ≥ 1 ball

Define sequence of R.V.:

X_1 = # balls until 1 bin covered

X_2 = # additional balls until 2 bins covered

$\vdots$

X_j = # additional balls after $j-1$ bins covered until j bins covered.

$$X = \sum_{j=1}^{n} X_n$$

Ex

$$\overbrace{\underset{\text{1 covered}}{b_1}}^{X_1} \quad \overbrace{b_2 \quad b_3 \quad \underset{\text{2 covered}}{b_4}}^{X_2} \quad \overbrace{b_5 \quad b_6 \quad b_7 \quad \underset{\text{3 covered}}{b_8}}^{X_3} \quad b_9$$

$X_1 = 1$

$X_2 = 3$

$X_3 = 4$

$E[X_j] = ?$

→ Assume $j-1$ bins full

→ Throw balls until hit empty

→ Each throw succeeds with Prob: $p_j = \frac{n-j+1}{n}$

$$\Pr(X_j = k) = (1-p_j)^{k-1} p_j$$

Geometric Random Variable

$$E[X_j] = \sum_{k=1}^{\infty} k \cdot (1-p_j)^{k-1} p_j$$

memoryless

$\Pr(X > m+n \mid X > m)$

$= \Pr(X > n)$

$$= p_j \cdot 1 + (1-p_j)(1 + E[X_j])$$

$$\Rightarrow E[X_j] = 1/p_j = \frac{n}{n-j+1}$$

$$E[X] = E[\sum X_j]$$

$$= \sum E[X_j]$$

$$= \sum_{j=1}^{n} \frac{n}{n-j+1}$$

$$= n \sum_{k=1}^{n} \frac{1}{k}$$

$$= n H_n = n \log n + O(1)$$

Conclusion: Expected # proposals = $n \log n + O(1)$

men = bins

proposals = balls

Alternate analysis:

Fix bin B.

$$\Pr(m \text{ balls miss } B) = \left(1-\frac{1}{n}\right)^m \leq e^{-m/n}$$

$$\Rightarrow \text{ if } m \geq (c+1)\, n \ln(n), \text{ then: } \leq 1/n^{c+1}$$

Union Bound: $\Pr(\text{miss } A \text{ or miss } B \text{ or } \ldots) \leq \sum_{j=1}^{n} \frac{1}{n^{c+1}}$

$$\leq 1/n^c$$

W.h.p. $\leq (c+1)\, n \ln n$ proposals

Corollary: $E[T] \leq \left(1-\frac{1}{n}\right)(c+1)\, n \ln n + \left(\frac{1}{n}\right) \cdot ??$
??

How happy are the women?

P_j = # proposals by j.

symmetry: all same

$$E[T] = E[\sum P_j] = \sum E[P_j] = n\,E[P_1] = n \log n + O(1)$$

$$\Rightarrow E[P_j] = \log n + o(1)$$

How happy are the men?

Fix man. Assume m unique proposals

$Y_1, Y_2, \dots Y_m$ = rank of m proposals

$Y = \min(Y_j)$

$$E[Y] = E[\min Y_j]$$
$$= \sum_{k=1}^{n} \Pr(Y \geq k)$$
$$= \sum_{k=1}^{n} \prod_{j=1}^{m} \Pr(Y_j \geq k)$$
$$= \sum_{k=1}^{n} \left(\frac{n-k+1}{n}\right)^m$$
$$= \sum_{i=1}^{n} \left(\frac{i}{n}\right)^m$$

$$= \frac{1}{n^m} \sum_{i=1}^{n} i^m$$

$$= \frac{1}{n^m} \left[\frac{n^{m+1}}{m+1} + O(n^m) \right]$$

$$\simeq O\left(\frac{n}{m+1} \right)$$