CS 331, Fall 2024

Lecture 19 (11/4)

Today:
- Find Pivot
- Linearity of expectation
- Union bound
- Contention resolution

Find Pivot (Part VII, Section 1)

So far: deterministic algos only

This unit: explore power of randomness

One extra power:

```
def Algo():
  • Step 1
  • Step 2
  • Flip Coin
  • Step 3
```

On HW: arbitrary bias is also OK

Reminder: both QuickSort and Selection use... FindPivot(L)

Input: L is n real numbers

Output: $e \in L$ s.t. $\text{rank}(e) \in \left[\frac{n}{4}, \frac{3n}{4}\right)$

2	7	9	1	10	-6	-4	3
✓			✓			✓	✓

Why? Recurse on two halves, both sides $< \frac{3}{4}n$ entries

Our solution in Part II:

FindPivot → Selection → Median of Medians → ...

... very complicated ...

How can H/T help?

Idea: Let e = random element of L

Say e is middling if $\text{rank}(x) \in \left[\frac{n}{4}, \frac{3n}{4}\right]$

$$\Pr[e \text{ is middling}] = \frac{\frac{3n}{4} - \frac{n}{4}}{n} = \frac{1}{2}$$

we can check in $O(n)$ time: compute $\text{rank}(x)$

FindPivot(L), take 2:

$e \leftarrow$ uniform random ele of L

While e is not middling:

  $e \leftarrow$ uniform random ele of L

} $O(n)$ time

Return e

$$E[\text{runtime}] = O(n) \cdot E[\underbrace{\#\text{ loops}}_{:= X}]$$

(expectation, "average")

What is distribution of X?

$$\Pr[X = i] = \left(\frac{1}{2}\right)^{i-1} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^{i}$$

$$E[X] = \sum_{i \geq 1} \Pr[X = i] \cdot i$$

$$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \dots$$

$$= \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots\right)$$

$$+ \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots\right)$$

$$+ \left(\frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \dots\right)$$

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

$\mathbb{E}$[runtime]: $O(n)$!!! so simple :)

Good "tail behavior": w.p. $> 99.9\%$, $X \le 10$

Recurring theme: randomized algos are

- simple & intuitive
- sometimes <u>much faster</u> than deterministic
- tricky to analyze... we care about

1) expected behavior 2) "high prob." behavior

Key issue: no <u>independence</u>!

X independent of Y

if $X \mid Y$ same as X

(review: Part 1, Sec. 6)

...

Step 1:

Step 2:

Step 3: highly dependent on 1/2...

Linearity of Expectation (Part VII, Section 2.1)

Let X, Y be r.v.s in $\mathbb{R}$

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

e.g. algo step 1   e.g. algo step 2

There are no caveats. This is always true!

Extends fine to sums of ≥ 2 r.v.s (recurse)

Intuition:

	H	T	
H	$1/4$	$1/2$	$X=1$
T	$1/4$	0	$X=0$
	$Y=1$	$Y=0$	

$$\mathbb{E}[X+Y] = \tfrac{1}{4}(1+1) + \tfrac{1}{4}(0+1) + \tfrac{1}{2}(1+0) + 0(0+0)$$

group like terms together...

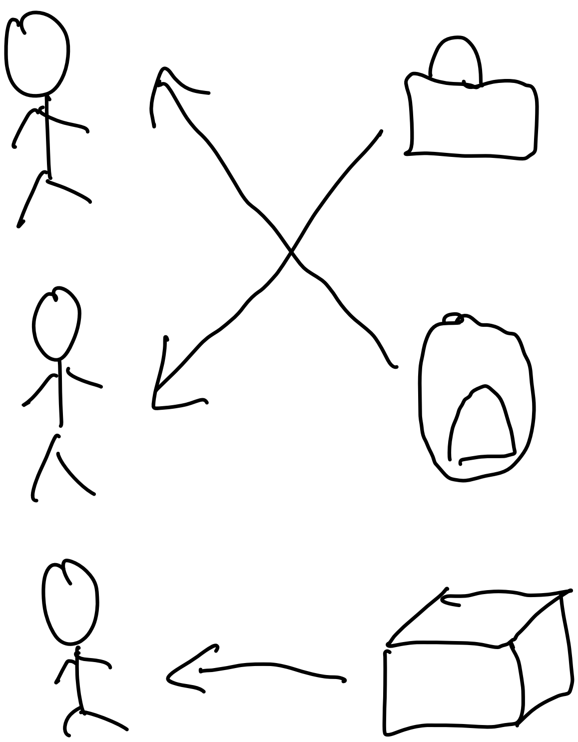
$\boxed{\text{Example}}$ Permutation fixed points

$\swarrow$ n students

Everyone in class gives backpack to doorman

Doorman remembers you, but shift changes...

New doorman returns random permutation



Let X =

correct assignments

What is $\mathbb{E}[X]$?

$$X = X_1 + X_2 + \dots + X_n$$

$X_1 = \begin{cases} 0 & \text{if student 1 wrong} \\ 1 & \text{if student 1 right} \end{cases}$

Observation: $\mathbb{E}[X_i] = \frac{1}{n}$

We're immediately done by lin. ex.:

$$\mathbb{E}[X] = \sum_{i \in [n]} \mathbb{E}[X_i] = n \cdot \frac{1}{n} = 1 \quad \blacksquare$$

X_i and X_j <u>not independent</u> but that's ok!

$$\boxed{\text{Indicators \& Events}}$$

Let $\mathcal{E}$ be <u>event</u> (random, either occurs or not)

We define

$$\mathbb{1}(\mathcal{E}) = \begin{cases} 1 & \mathcal{E} \text{ occurs} \\ 0 & \mathcal{E} \text{ doesn't occur} \end{cases}$$

indicator random variable

ex. $X_i = \mathbb{1}\left(\begin{array}{c}\text{student } i \\ \text{gets backpack}\end{array}\right)$

We have

$$\mathbb{E}[\mathbb{1}(\mathcal{E})] = 1 \cdot \Pr[\mathcal{E}] + 0 \cdot (1 - \Pr[\mathcal{E}])$$

$$= \Pr[\mathcal{E}] \quad \text{(also always true)}$$

$\boxed{\text{Example}}$ Coupon collector

Every week you get uniformly random coupon

If you collect all n, you win a prize!

How many weeks in $\mathbb{E}$?

Let $X = X_1 + X_2 + \ldots + X_n$

total weeks waited

X_i = # weeks after $(i-1)^{th}$ coupon before i^{th} coupon received

$\boxed{\text{Aside}}$ Geometric r.v.s

Let Z = # tosses of coin before first H

H w.p. p

T w.p. $1-p$

e.g. X_i above has $p = \frac{n-(i-1)}{n}$ $\frac{\text{(new coupons)}}{\text{(total coupons)}}$

$$\Pr[Z=i] = (1-p)^{i-1} p$$

$$\mathbb{E}[Z] = \sum_{i \geq 1} i \cdot (1-p)^{i-1} p$$

$$= 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{p}$$

$$\Rightarrow \mathbb{E}[X_i] = \frac{n}{n-(i-1)} \quad \forall i \in [n]$$

At this point: lin ex gives

$$\mathbb{E}[X] = \sum_{i \in [n]} \mathbb{E}[X_i]$$

total weeks

$$= \frac{n}{n} + \frac{n}{n-1} + \dots + \frac{n}{1}$$

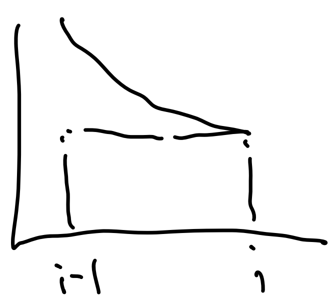
$$= n\left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right)$$

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

"Harmonic number"

$$1 + \frac{1}{2} + \dots + \frac{1}{n} \leq 1 + \int_1^n \frac{1}{t}\, dt \leq \log(n) + 1$$

$$\int_{i-1}^{i} \frac{1}{t}\, dt \geq \frac{1}{i}$$



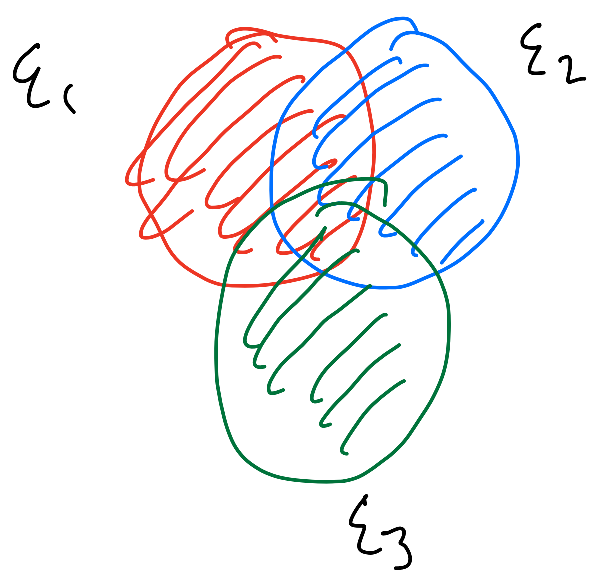
Union Bound (Part VII, Section 2.1)

2nd most useful tool for dependent r.v.s

Let $\mathcal{E}_1, \ldots \mathcal{E}_k$ be <u>any</u> events

$$\underbrace{\Pr\left[\bigcup_{i\in[k]} \mathcal{E}_i\right]}_{\text{"any event happens"}} \le \sum_{i\in[k]} \underbrace{\Pr[\mathcal{E}_i]}_{\text{"event } i \text{ happens"}}$$

Proof 1: picture



total area

$\le$ area( )

$+$ area( )

$+$ area( )

Proof 2: lin. ex.

Recall $\Pr[\mathcal{E}_i] = \mathbb{E}[\mathbb{1}(\mathcal{E}_i)]$

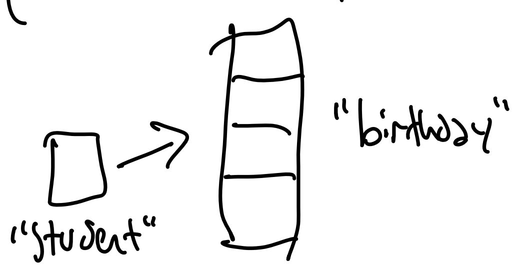
$$\mathbb{1}\Big(\bigcup_{i\in[k]} \mathcal{E}_i\Big) \le \mathbb{1}(\mathcal{E}_1) + \dots + \mathbb{1}(\mathcal{E}_k)$$

Take $\mathbb{E}$ of both sides.

$\boxed{\text{Example}}$ Birthday paradox

n students, m days of year

Q: if every student has uniformly random birthday, at what n do we expect <u>collisions</u> (shared bday)?

Real-life app: hashing 

For every pair of students $1 \le i < j \le n$

$$\mathcal{E}_{i,j} = \text{students } i,j \text{ same bday}$$

$$\Pr\left[\begin{array}{c}\text{some two} \\ \text{students collide}\end{array}\right] = \Pr\left[\bigcup_{1 \le i < j \le n} \mathcal{E}_{i,j}\right]$$

$$\le \sum_{1 \le i < j \le n} \Pr[\mathcal{E}_{i,j}] \approx \frac{n^2}{2m}$$

Small if $n^2 \ll 2m$:

$$\text{e.g. } \sqrt{2 \cdot 365} \approx \underset{\text{students}}{27}$$

(surprisingly low: "paradox")

Surprisingly accurate threshold $\sqrt{2m}$

(next time: <u>concentration</u>)

Contention resolution (Part VII, Section 2.2)

n processes want to run jobs (1 unit of time)

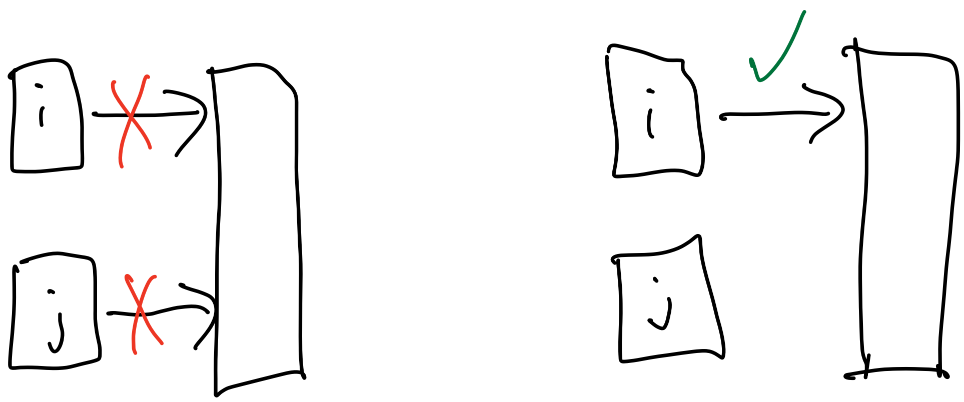
need to access shared central server

no communication! each time step:

- make an attempt

or

- make no attempt



If you are unique attempt, you get served.

$\gtrsim n$ steps necessary. Achievable?

Nearly: with randomness, $O(n \log(n))$ steps!

Idea: randomly attempt w.p. $\frac{1}{n}$ for T steps

$\mathcal{E}_i$: i^{th} process never uniquely served

In one step,

$$\Pr[i \text{ served}] = \frac{1}{n}\underbrace{\left(1-\frac{1}{n}\right)^{n-1}}_{\approx \frac{1}{e}} \geq \frac{1}{4n} \quad \forall n \geq 2$$

Hence $\Pr[\mathcal{E}_i] \leq \left(1-\frac{1}{4n}\right)^{T}$

If $T = 4Cn\log(n)$

$$\underbrace{\Pr\left[\bigcup_{i\in[n]} \mathcal{E}_i\right]}_{\text{some process not done}} \leq n \cdot \left(1-\frac{1}{4n}\right)^{4Cn\log(n)}$$

$$\leq n \cdot \left(\frac{1}{e}\right)^{C\log(n)} \leq \underbrace{\frac{1}{n^{C-1}}}_{\text{very low!}}$$