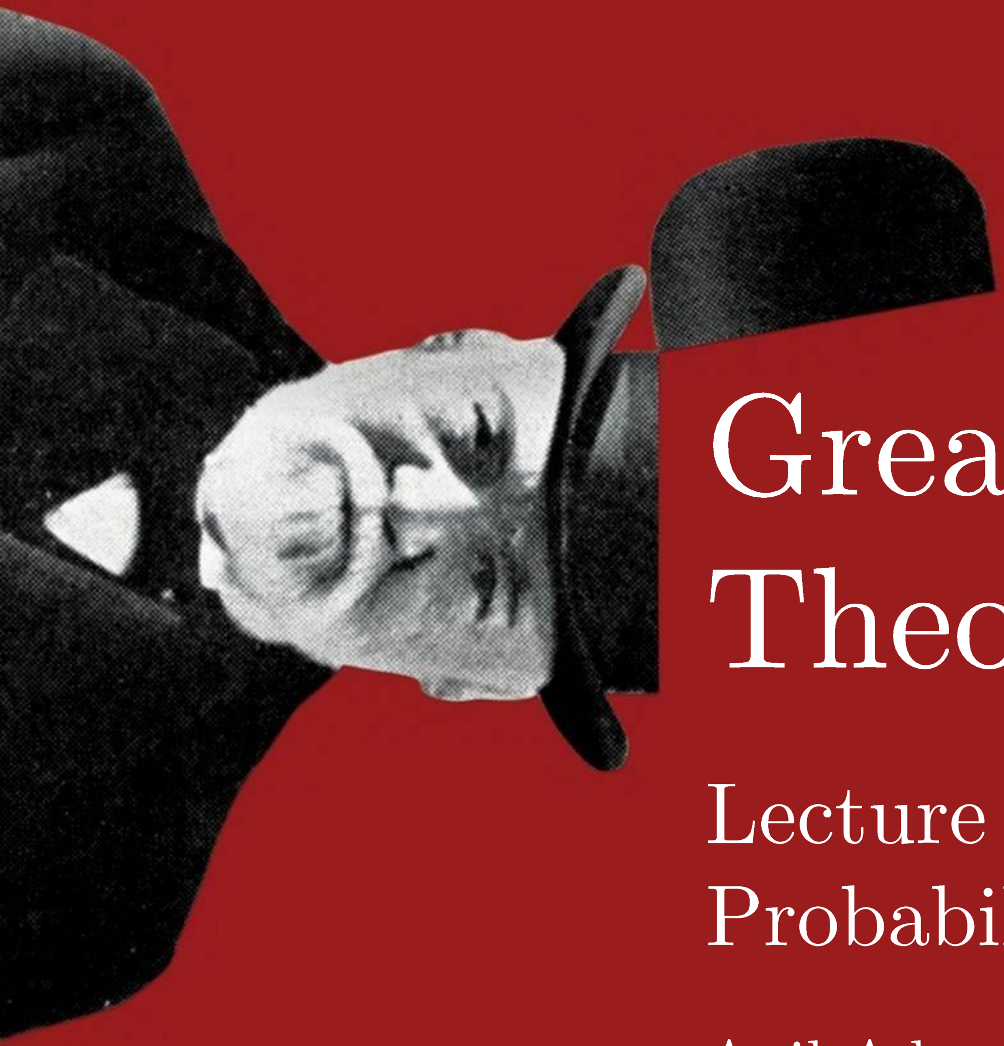


Great Ideas in Theoretical CS

Lecture 22: Probability II

Anil Ada
Ariel Procaccia (this time)

Deep questions

If I randomly put 100 letters into 100 addressed envelopes, on average how many will end up in the correct envelope?

How many times on average do I need to flip a fair coin to get heads?

Great expectations

Random variable

- Let S be a sample space
- A random variable is a function $X: S \to \mathbb{R}$
- Examples:
  - $X =$ value of red die when red and blue are rolled:

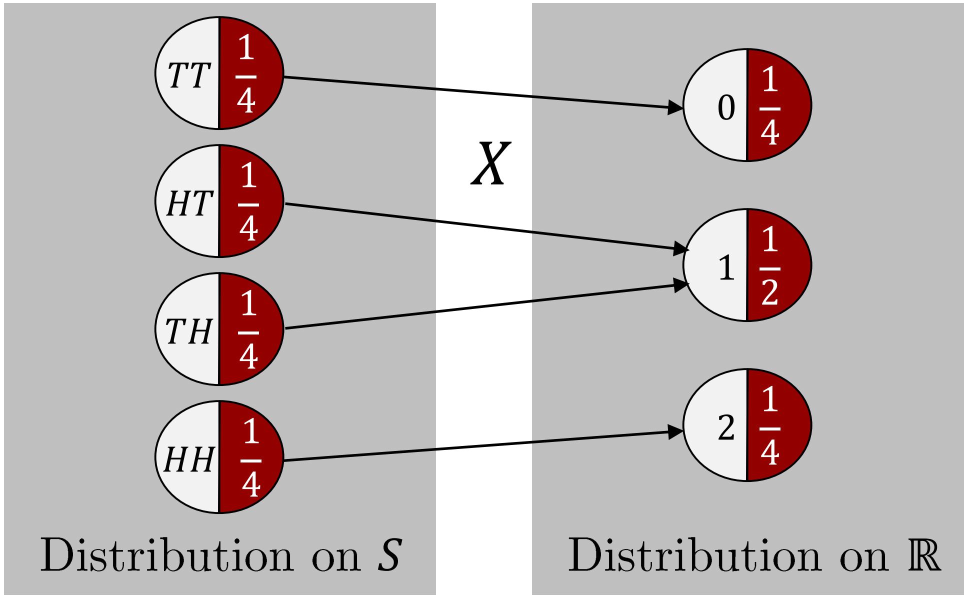
    $$X(3,4) = 3, \qquad X(1,5) = 1$$

  - $X =$ value of the sum of two dice when red and blue are rolled:

    $$X(3,4) = 7, \qquad X(1,5) = 6$$

Two coins tossed

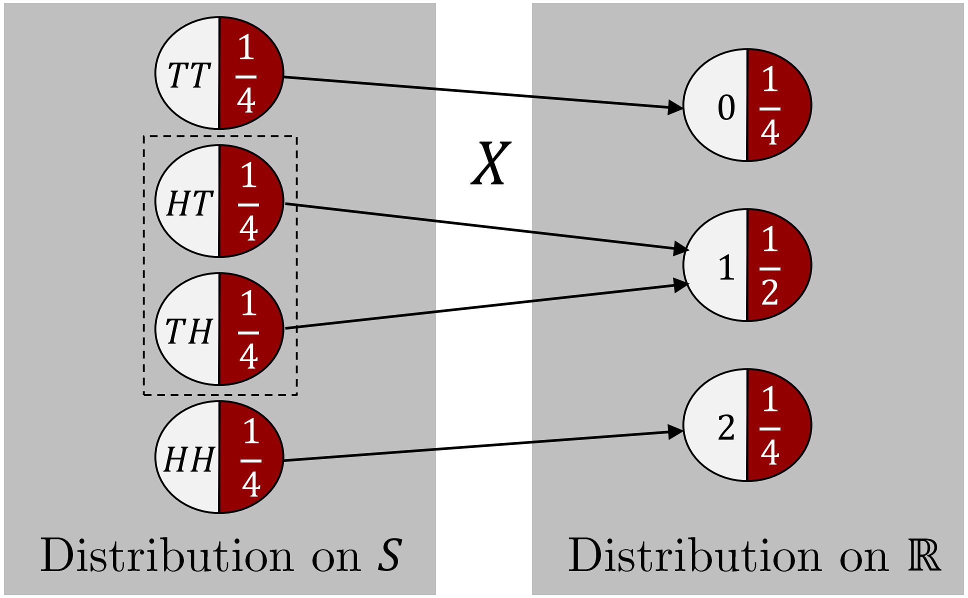
- $X: \{TT, HT, TH, HH\} \to \{0,1,2\}$ counts the number of heads



From RVs to events

- For RV X and $a \in \mathbb{R}$ we can define the event E that $X = a$:

$$\Pr[E] = \Pr[X = a] = \Pr[\{t \in S \mid X(t) = a\}]$$

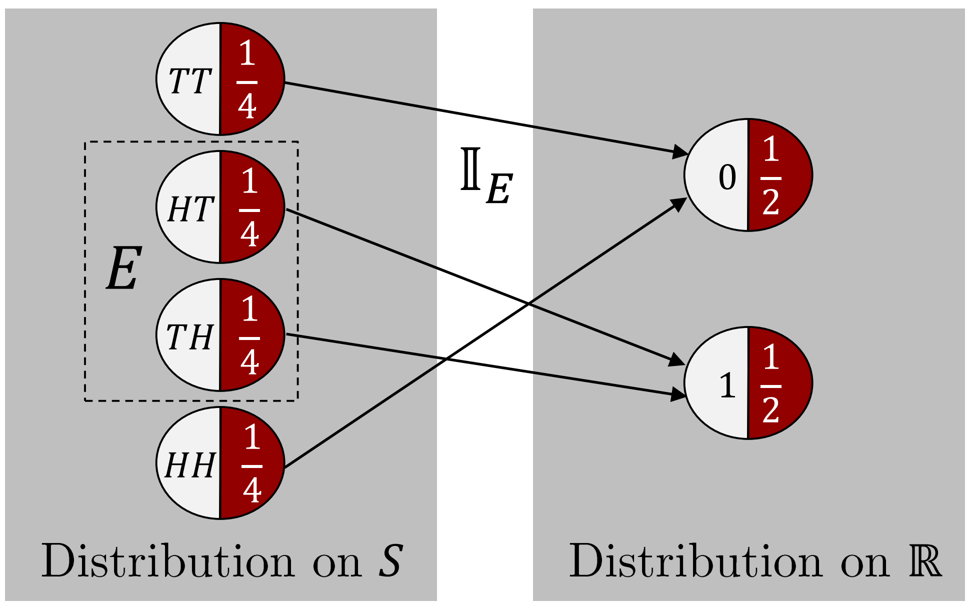


$$\Pr[X = 1] = \Pr[\{HT, TH\}] = 1/2$$

From events to RVs

- For any event E, define the indicator random variable for E:

$$\mathbb{I}_E(t) = \begin{cases} 1 & t \in E \\ 0 & t \notin E \end{cases}$$



E = exactly one head

Independent RVs

- Two random variables are independent if for every a, b, the events $X = a$ and $Y = b$ are independent

Expectation

- The expectation of a random variable X is:

$$\mathbb{E}[X] = \sum_{t \in S} \Pr[t] \times X(t) = \sum_{k} \Pr[X = k] \times k$$

- Poll 1: X is the #heads in 3 coin tosses. $\mathbb{E}[X] = ?$
  1. 1
  2. 4/3
  3. 3/2
  4. 2

Don't always expect the expected!
$\Pr[X = \mathbb{E}[X]]$ could be 0

Expectation

- If $\mathbb{I}_E$ is the indicator RV for the event E,
$$\mathbb{E}[\mathbb{I}_E] = \Pr[\mathbb{I}_E = 1] \times 1 = \Pr[E]$$
- If X and Y are two RVs (on the same sample space S) then $Z = X + Y$ is also an RV, defined by $Z(t) = X(t) + Y(t)$
- Example: X is one die roll, Y is another, and Z is their sum

Linearity of expectation

- If $Z = X + Y$, then

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

Even if X and Y are not independent!

Linearity of expectation

$$\mathbb{E}[Z] = \sum_{t \in S} \Pr[t] Z(t)$$

$$= \sum_{t \in S} \Pr[t] \left(X(t) + Y(t)\right)$$

$$= \sum_{t \in S} \Pr[t]\, X(t) + \sum_{t \in S} \Pr[t]\, Y(t)$$

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

Using linearity of expectation

General approach:
View thing you care about as expected value of some RV; write this RV as sum of simpler RVs (typically indicator RVs); Solve for their expectations and add them up!

Using linearity of expectation

- If I randomly put 100 letters into 100 addressed envelopes, what is the expected number of letters that will end up in their correct envelopes?

$$\sum_k k \times \Pr[k \text{ correct letters}] = \sum_k k \times [\text{..aargh!!..}]$$

Using linearity of expectation

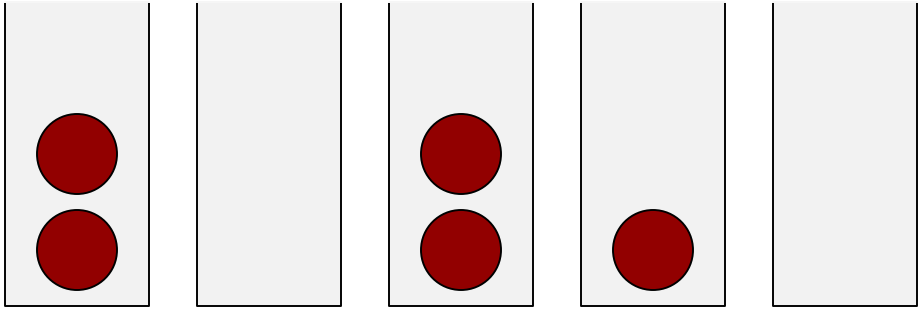
- Let A_i be the event that the ith letter is in the correct envelope
- Let $\mathbb{I}_{A_i}$ be the indicator variable for A_i
  - Not independent!
- $\mathbb{E}[\mathbb{I}_{A_i}] = \Pr[A_i = 1] = 1/100$
- We are interested in $Z = \sum_{i=1}^{100} \mathbb{I}_{A_i}$
- $\mathbb{E}[Z] = 100 \times \frac{1}{100} = 1$

USING LINEARITY OF EXPECTATION

- Poll 2: We flip n coins of bias p; what is the expected number of heads?
  1. 1
  2. p
  3. n
  4. np

Balls and bins

- n jobs are assigned to n servers uniformly at random
- What is the expected number of jobs per server?

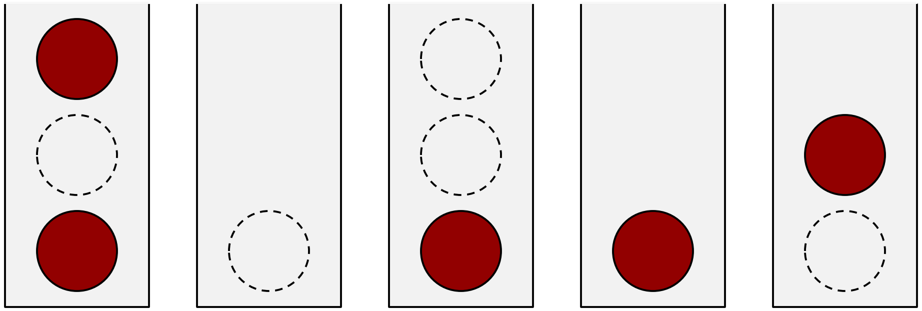


Balls and bins

- Let X_i be the number of jobs on server i
- $n = \mathbb{E}[X_1 + \cdots X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]$
- By symmetry, $\mathbb{E}[X_i] = 1$ for all i
- **Fact:** The expected number of jobs assigned to the busiest server is roughly $\log n \,/\, \log\log n$

Balls and bins

- n jobs are assigned to n servers, but for every job we choose two servers uniformly at random, and use the less busy server
- Poll 3: Expected number of jobs per server?
  1. $1/n$
  2. $1/4$
  3. $1/2$
  4. 1



- Fact: Busiest server has $\sim \log \log n$ jobs

Conditional expectation

- $\mathbb{E}[X \mid E] = \sum_k \Pr[X = k \mid E] \times k$
- Similarly to conditional probability:
  $\mathbb{E}[X] = \mathbb{E}[X \mid E] \times \Pr[E] + \mathbb{E}[X \mid \bar{E}] \times \Pr[\bar{E}]$

GEOMETRIC RVs

- Flip a coin with probability p of heads
- $X = \#\text{flips until first heads}$
- $$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X \mid H]\Pr[H] + \mathbb{E}[X \mid T]\Pr[T] \\ &= 1 \cdot p + (\mathbb{E}[X] + 1)(1-p) \\ &= 1 + \mathbb{E}[X](1-p)\end{aligned}$$
- It follows that $\mathbb{E}[X] = 1/p$

251 Land

- All 15251 students fly off to space and colonize Mars
- Faced with the problem that there are 57% men, the authorities impose a new rule: When having kids, stop after you have a girl
- Poll 4 yes/no: Will the number of new boys be larger than the number of new girls?
- Poll 5 yes/no: What if the rule is: stop after having two girls?

Capstone project *

computational
social choice +
approximation algs +
linearity of expectation
(+ randomized algs)

* Just for fun

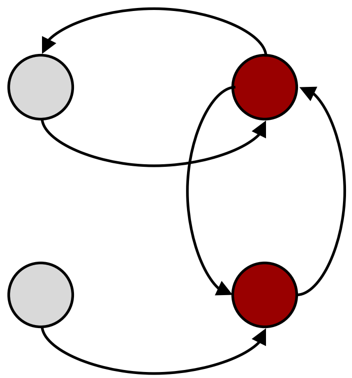
Selecting a subset *

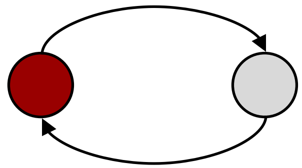
- A *k-selection system* receives a directed graph as input and outputs $V' \subseteq V$ such that $|V'| = k$
- Edges are interpreted as approval votes, trust, or support
- Think of graph as directed social network
- A k-selection system f is *impartial* if $i \in f(G)$ does not depend on the votes of i

* Just for fun

SELECTING A SUBSET *

- Optimization target: sum of indegrees of selected vertices
- Optimal solution: not impartial
- $k = n$: no problem
- $k = 1$: no positive impartial approximation
- $k = n - 1$: no positive impartial approximation, even if each vertex has at most one outgoing edge!





* Just for fun

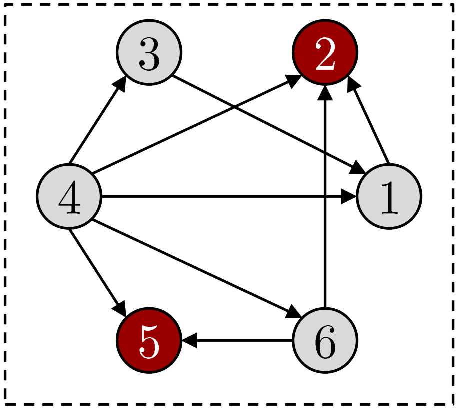
Selecting a subset *

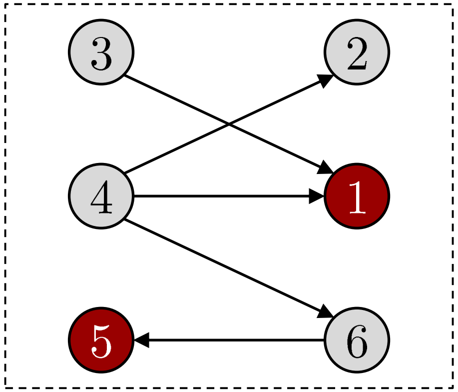
- Each tribe member votes for at most one member
- One member must be eliminated
- Impartial rule cannot have property: if unique member received votes he is not eliminated

* Just for fun

Selecting a subset *

- The random partition algorithm:
  - Assign vertices uniformly at random to 2 subsets
  - For each subset, select $\sim \frac{k}{2}$ vertices with highest indegrees based on edges from the other subset
- This mechanism is clearly impartial

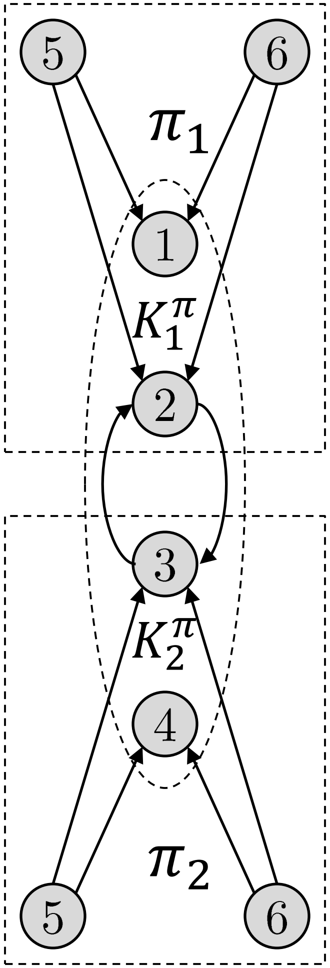




* Just for fun

Selecting a subset *

- Theorem [Alon et al. 2011]: Random Partition is a $\frac{1}{4}$-approximation algorithm
- Proof:
  - Assume for ease of exposition: k is even
  - Let K be the optimal set
  - A partition $\pi = (\pi_1, \pi_2)$ divides K into two subsets $K_1^\pi = K \cap \pi_1$ and $K_2^\pi = K \cap \pi_2$
  - $d_1^\pi = \{(u, v) \in E | u \in \pi_2, v \in K_1^\pi\}$, d_2^π defined analogously
  - $\mathbb{E}[d_1^\pi + d_2^\pi] = \frac{OPT}{2}$ by linearity of expectation
  - We get at least $\frac{d_1^\pi + d_2^\pi}{2}$ ■



* Just for fun

Summary

- Terminology:
  - Random variables
  - Expectation
  - Conditional expectation
  - Geometric RVs
- Principles:
  - Using the linearity of expectation by writing RVs as sums of simple RVs