Great Theoretical Ideas In Computer Science

Anupam Gupta

CS 15-251 Spring 2005

Lecture 23 April 5, 2005 Carnegie Mellon University

Probability III:

The probabilistic method & infinite probability spaces

Recap

Random Variables

- An event is a subset of S.
- A Random Variable (RV) is a (real-valued) function on S.

Example:

- Event A: the first die came up 1.
- Random Variable X: the value of the first die.

E.g., $X(\langle 3,5\rangle)=3$, $X(\langle 1,6\rangle)=1$.

x	$P(x)$
$\langle 1,1\rangle$	$1/36$
$\langle 1,2\rangle$	$1/36$
$\langle 1,3\rangle$	$1/36$
$\langle 1,4\rangle$	$1/36$
$\langle 1,5\rangle$	$1/36$
$\langle 1,6\rangle$	$1/36$
$\langle 2,1\rangle$	$1/36$
$\vdots$	
$\langle 6,5\rangle$	$1/36$
$\langle 6,6\rangle$	$1/36$

It's a floor wax *and* a dessert topping

It's a function on the sample space S.

It's a variable with a probability distribution on its values.

You should be comfortable with both views.

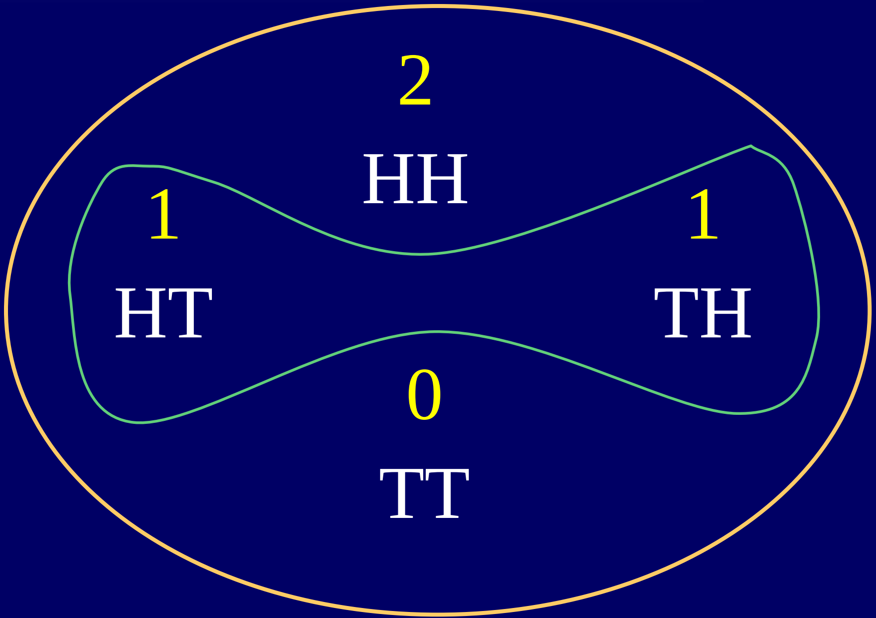
Definition: expectation

The <u>expectation</u>, or <u>expected value</u> of a random variable X is

$$\sum_{x \in S} \Pr(x) X(x) = \sum_k k \Pr(X = k)$$

E.g, 2 coin flips, X = # heads.

What is E[X]?



Thinking about expectation

D	S	X	Distrib on X
$\frac{1}{4}$ ---	TT	$\longrightarrow$	0 --- $\frac{1}{4}$
$\frac{1}{4}$ ---	TH	$\longrightarrow$	1 --- $\frac{1}{2}$
$\frac{1}{4}$ ---	HT	$\longrightarrow$	2 --- $\frac{1}{4}$
$\frac{1}{4}$ ---	HH	$\longrightarrow$	

$E[X] = \frac{1}{4}*0 + \frac{1}{4}*1 + \frac{1}{4}*1 + \frac{1}{4}*2 = 1.$

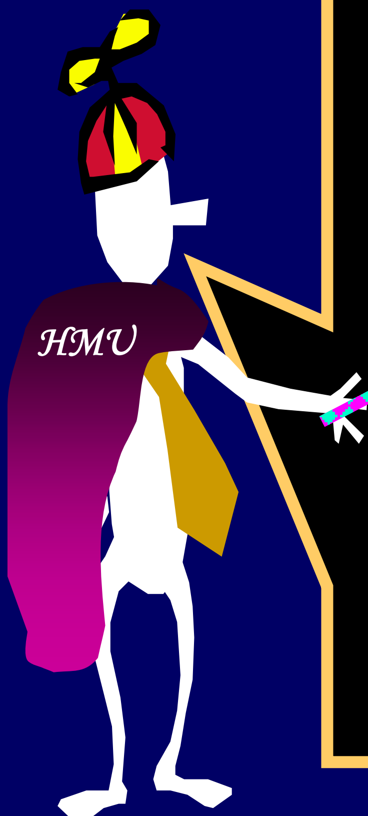
$E[X] = \frac{1}{4}*0 + \frac{1}{2}*1 + \frac{1}{4}*2 = 1.$

Linearity of Expectation

If $Z = X+Y$, then

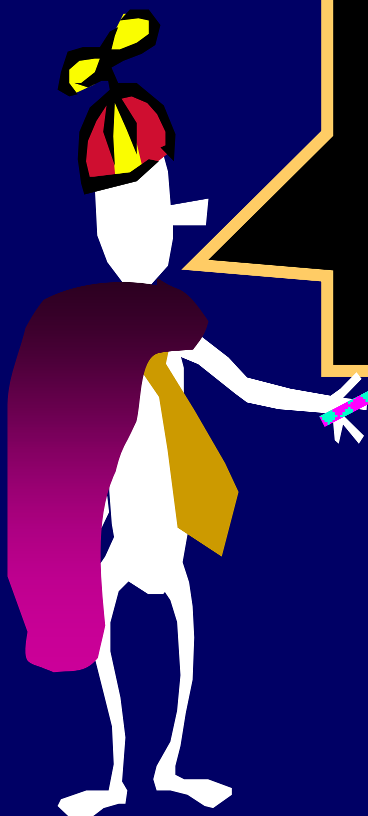
$$E[Z] = E[X] + E[Y]$$

Even if X and Y are not independent.



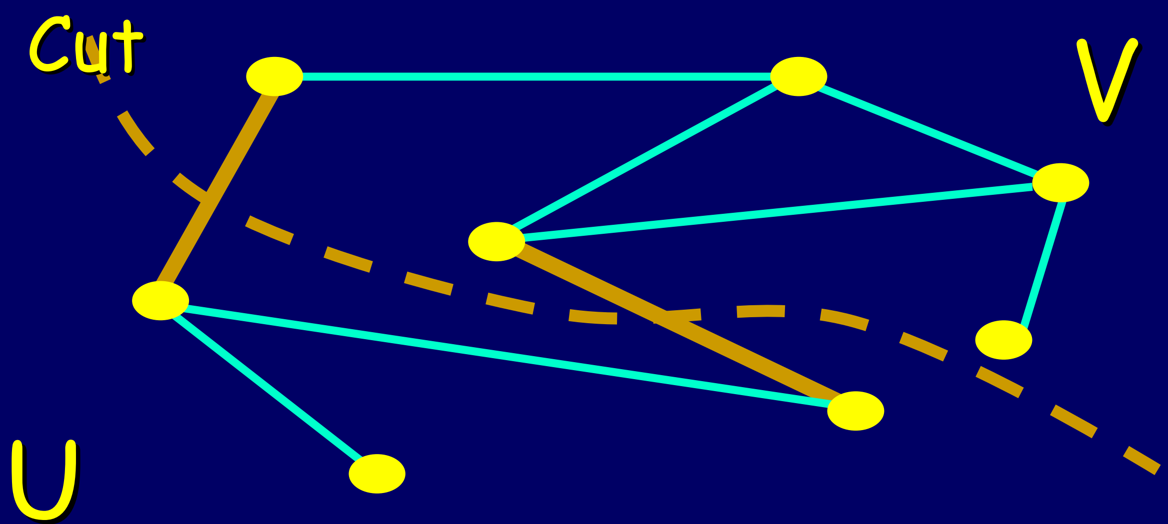
New topic: The probabilistic method

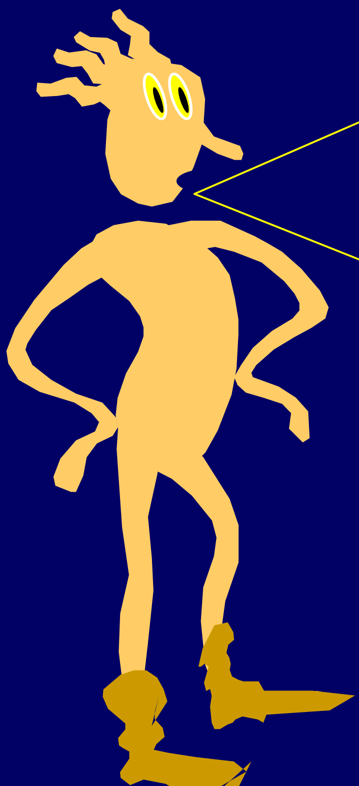
Use a probabilistic argument to prove a non-probabilistic mathematical theorem.



Definition: A cut in a graph.

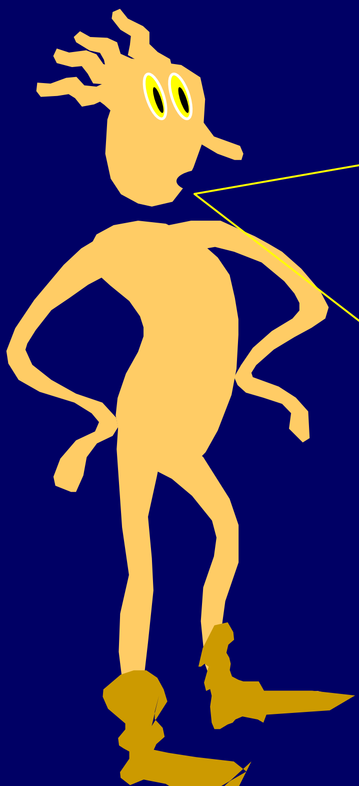
A *cut* is a partition of the nodes of a graph into two sets: U and V. We say that an edge crosses the cut if it goes from a node is U to a node in V.





Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.



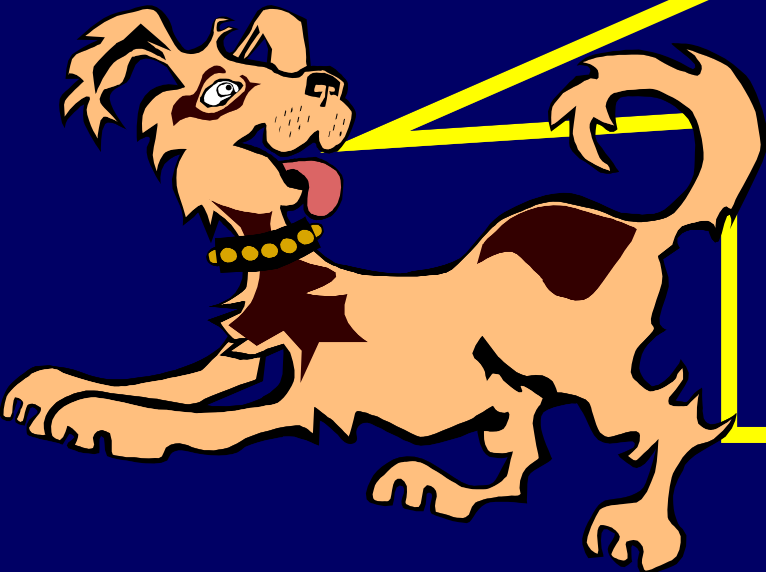
Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.

How are we going to prove this?

Will show that if we pick a cut <u>at random</u>, the <u>expected</u> number of edges crossing is $\frac{1}{2}$(# edges).

How does this prove the theorem?

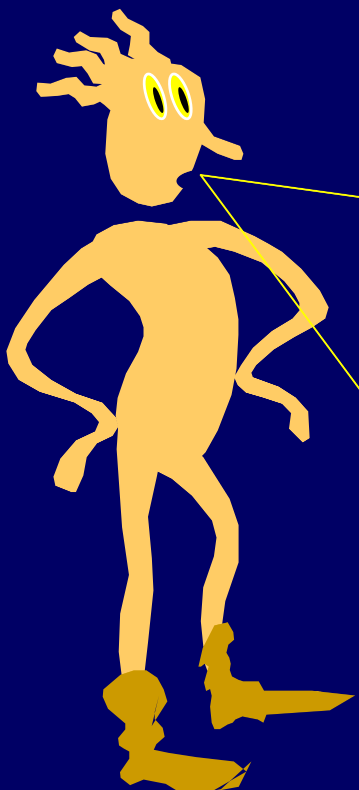


What might be is surely possible!

Goal: show exists object of value at least v.

Proof strategy:

- Define distribution D over objects.
- Define RV: X(object) = value of object.
- Show $E[X] \geq v$. Conclude it must be possible to have $X \geq v$.



Theorem:

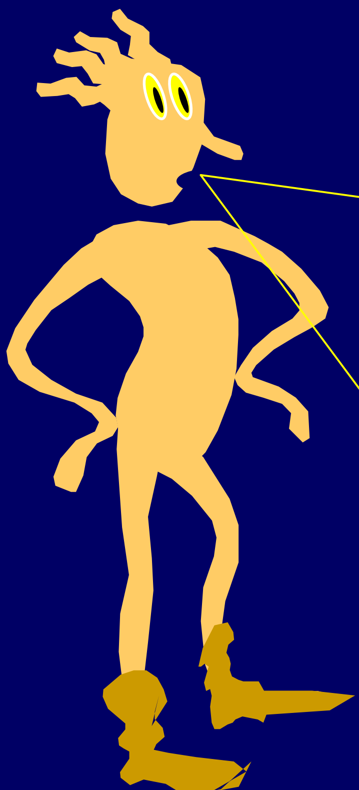
In any graph, there exists a cut such that at least half the edges cross the cut.

Proof: Pick a cut uniformly at random. I.e., for each node flip a fair coin to determine if it is in U or V.

Let X_e be the indicator RV for the event that edge e crosses the cut.

What is $E[X_e]$?

Ans: $\frac{1}{2}$.



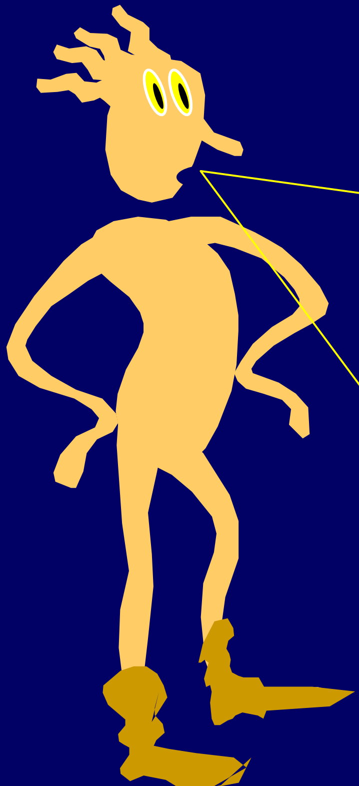
Theorem:

In any graph, there exists a cut such that at least half the edges cross the cut.

Proof:

- Pick random cut.
- Let $X_e=1$ if e crosses, else $X_e=0$.
- Let X = total #edges crossing.
- So, $X = \sum_e X_e$.
- Also, $E[X_e] = \frac{1}{2}$.
- By linearity of expectation,

$$E[X] = \tfrac{1}{2}(\text{total \#edges}).$$



Pick a cut uniformly at random. I.e., for each node flip a fair coin to see if it should be in U.

E[#of edges crossing cut]

= # of edges/2

The sample space of all possible cuts must contain at least one cut that at least half the edges cross: if not, the average number of edges would be less than half!

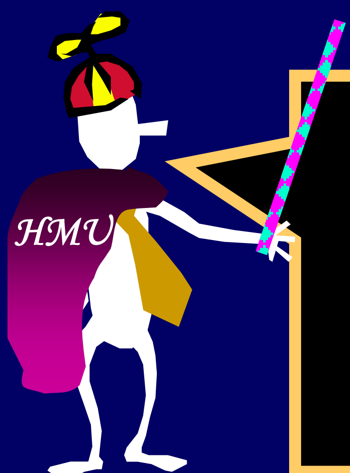
Another example of prob. method

What you did on hwk #8.

- If you color nodes at random, Pr(every v has a neighbor of a different color) > 0.
- So, must exist coloring where every v has a neighbor of a different color.
- This then implied existence of even-length cycle.

Back to cuts: Can you use this argument to also <u>find</u> such a cut?

In this case you can, through a neat strategy called the conditional expectation method

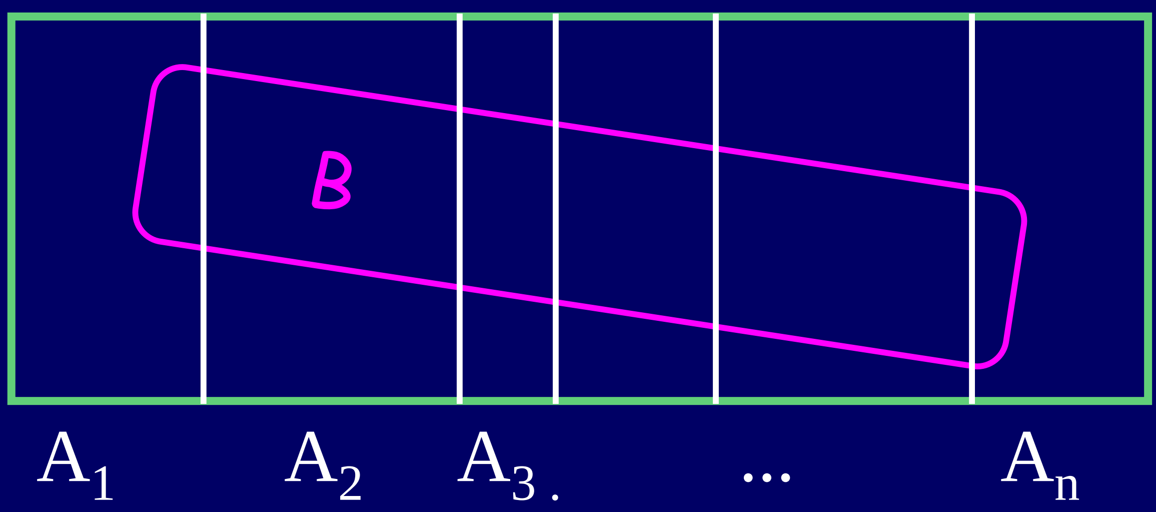


Idea: make decisions in greedy manner to maximize expectation-to-go.

First, a few more facts…

For any partition of the sample space S into disjoint events $A_1, A_2, \ldots, A_n$, and any event B,

$\Pr(B) = \sum_i \Pr(B \setminus A_i) = \sum_i \Pr(B|A_i)\Pr(A_i)$.



Def: Conditional Expectation

For a random variable X and event A, the conditional expectation of X given A is defined as:

$$E[X|A] = \sum_k k \Pr(X = k|A)$$

E.g., roll two dice. X = sum of dice, E[X] = 7.

Let A be the event that the first die is 5.

E[X|A] = 8.5

Def: Conditional Expectation

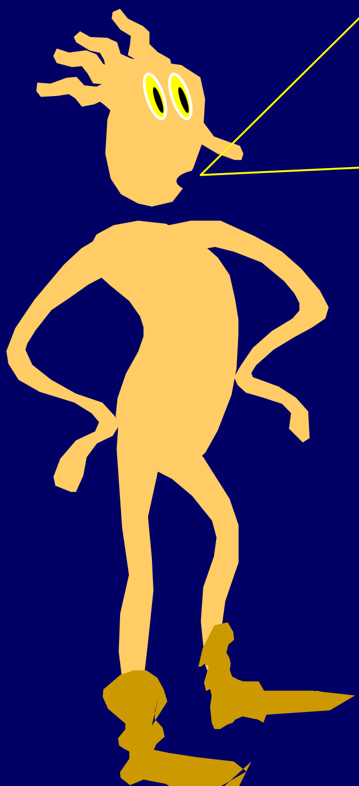
For a random variable X and event A, the conditional expectation of X given A is defined as:

$$E[X|A] = \sum_k k \Pr(X = k|A)$$

Useful formula: for any partition of S into $A_1, A_2, \ldots$ we have: $E[X] = \sum_i E[X|A_i]\Pr(A_i)$.

Proof: just plug in $\Pr(X=k) = \sum_i \Pr(X=k|A_i)\Pr(A_i)$.

Recap of cut argument



Pick random cut.

- Let X_e=1 if e crosses, else X_e=0.
- Let X = total #edges crossing.
- So, $X = \sum_e X_e$.
- Also, $E[X_e] = \frac{1}{2}$.
- By linearity of expectation,

$$E[X] = \tfrac{1}{2}(\text{total \#edges}).$$

Conditional expectation method

Say we have already decided fate of nodes 1,2,...,i-1. Let X = number of edges crossing cut if we place rest of nodes into U or V at random.

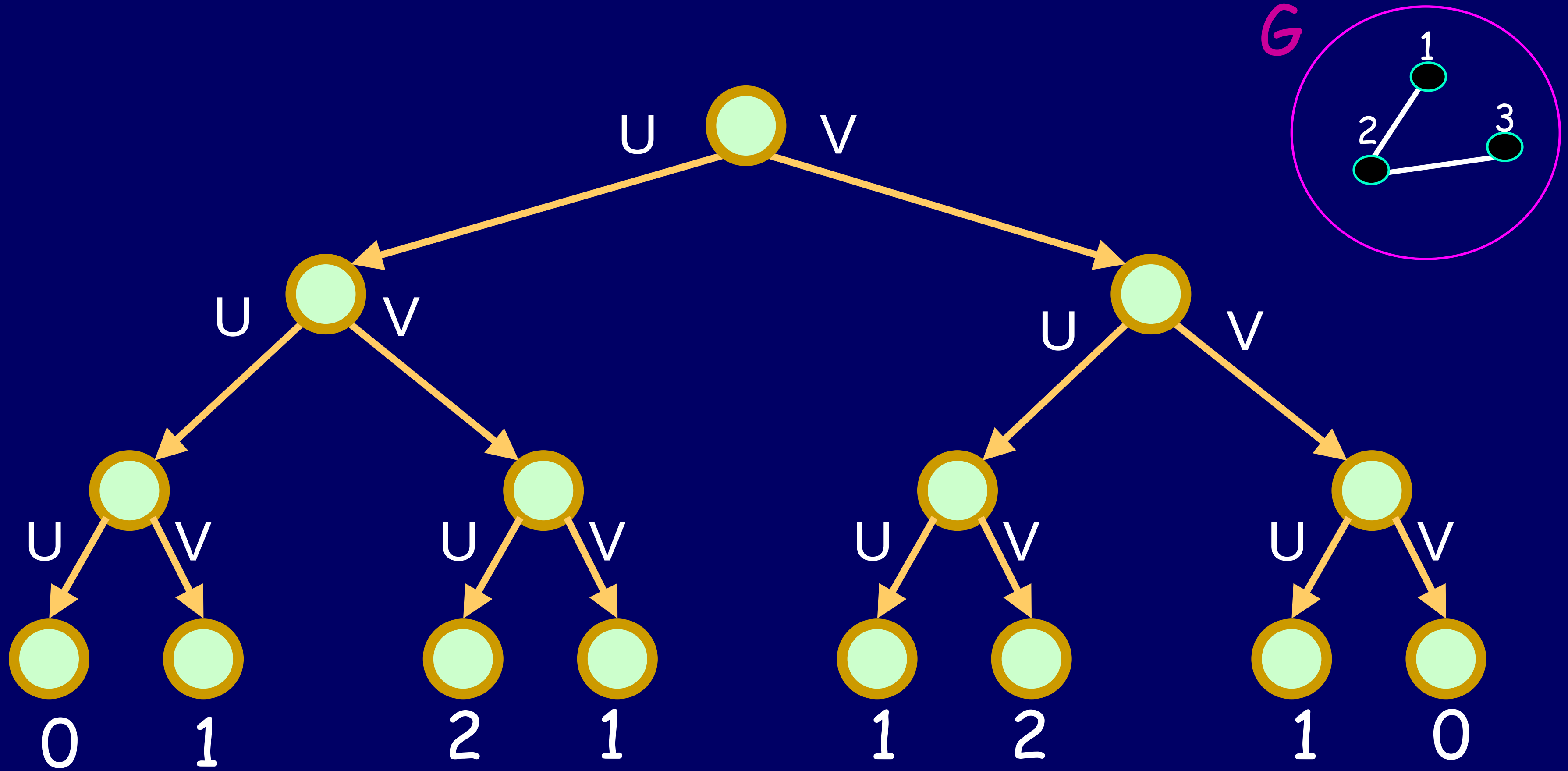
Let A = event that node i is put into U.

So, $E[X] = \frac{1}{2}E[X|A] + \frac{1}{2}E[X|\neg A]$

It can't be the case that both terms on the RHS are smaller than the LHS. So just put node i into side whose C.E. is larger.

Pictorial view (important!)

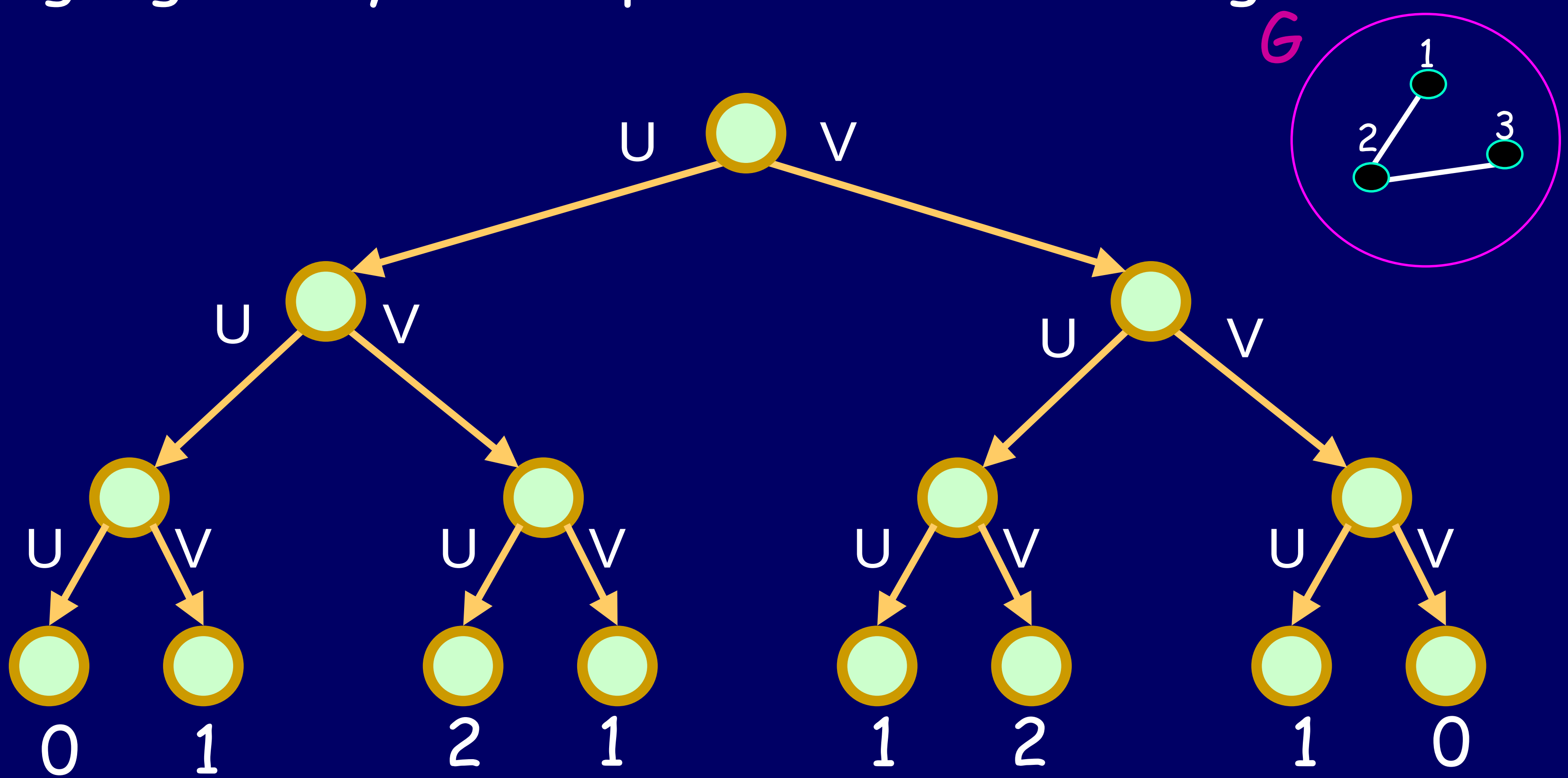
View S as leaves of choice tree. i^{th} choice is where to put node i. Label leaf by value of X. $E[X]$ = avg leaf value.



Pictorial view (important!)

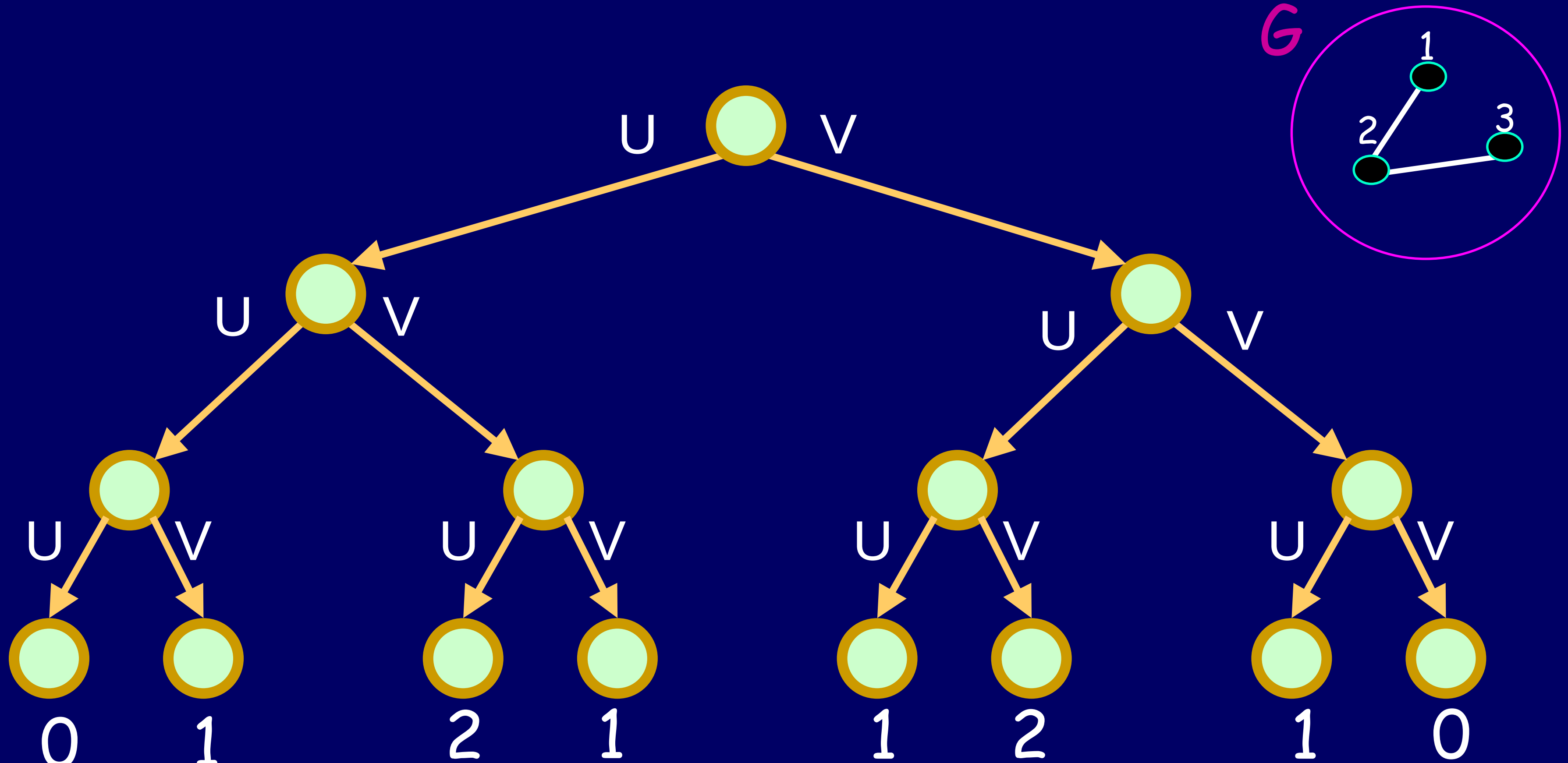
If A is some node (the event that we reach that node), then $E[X|A]$ = avg value of leaves below A.

Alg = greedily follow path to maximize avg.



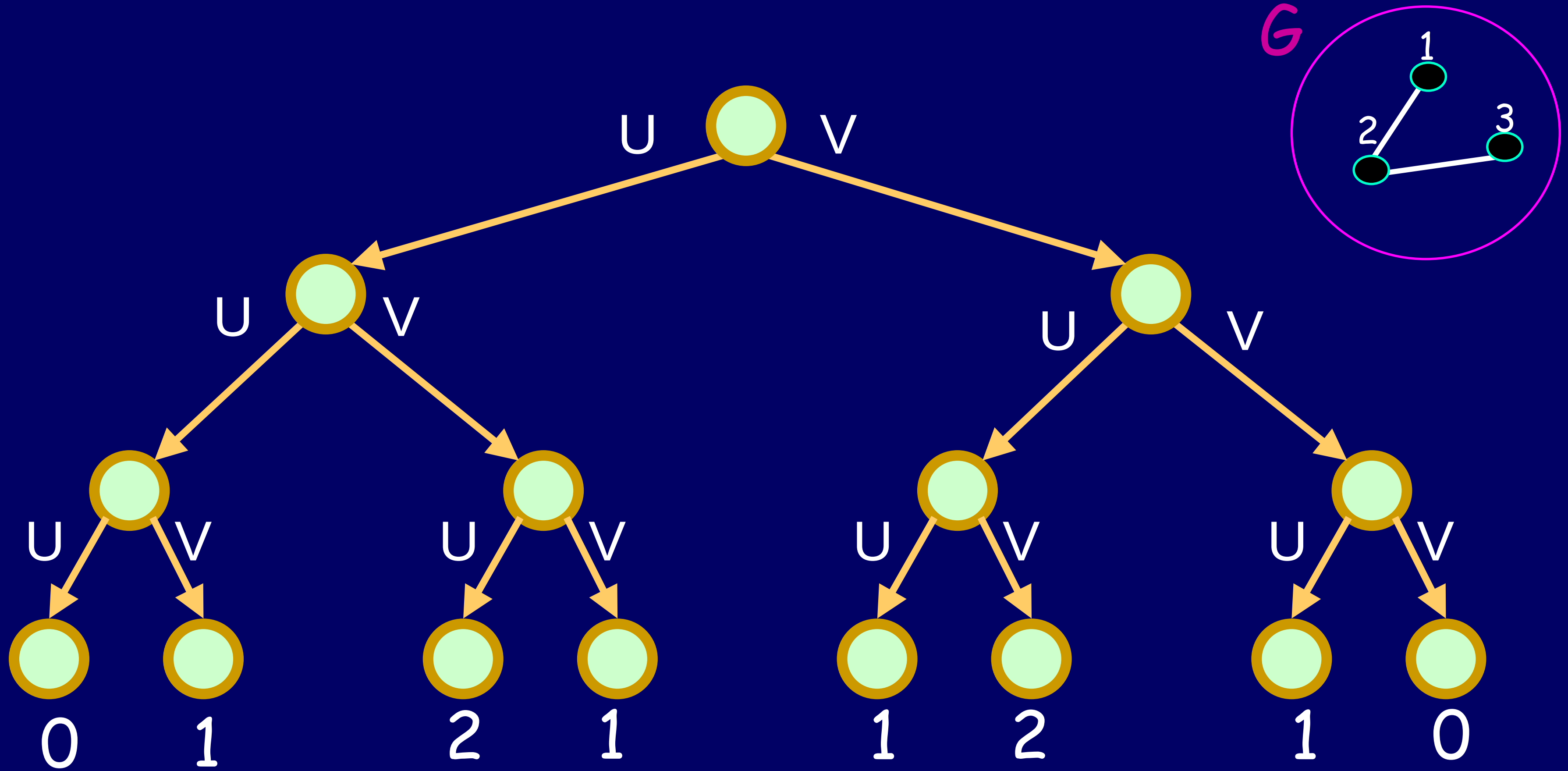
Pictorial view (important!)

Linearity of expectation gives us a way of magically computing E[X|A] for any node A.
(Even though the tree has 2^n leaves)



Pictorial view (important!)

In particular, $E[X|A]$ = (# edges crossing so far) + ½(# edges not yet determined)

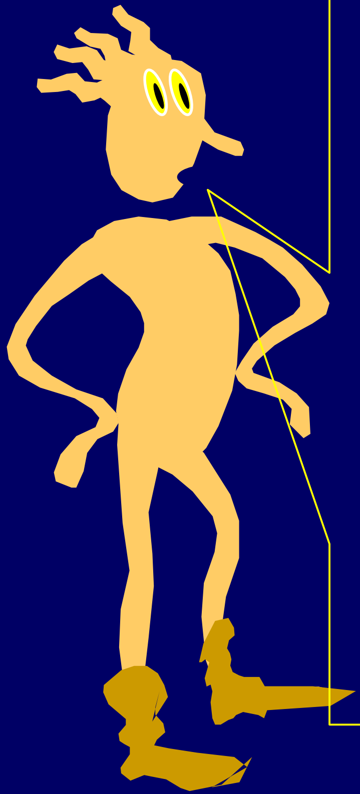


Conditional expectation method

In fact, our algorithm is just: put node i into the side that has the fewest of its neighbors so far.

(The side that causes the most of the edges determined so far to cross the cut).

But the probabilistic view was useful for proving that this works!

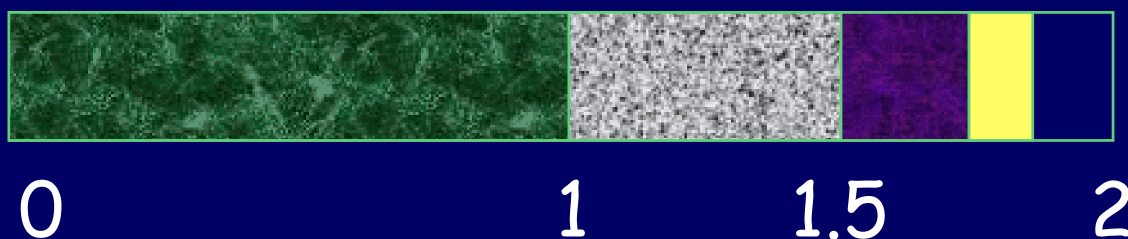


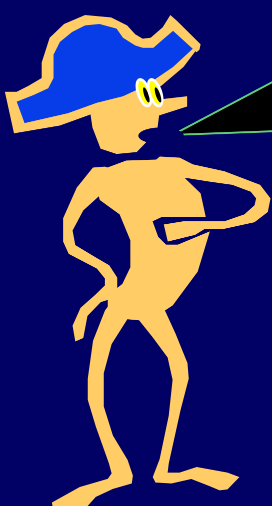
In many cases, though, we can't get an exact handle on these expectation. Probabilistic method can often give us proof of existence without an algorithm for finding the thing.

In many cases, no efficient algorithms for finding the desired objects are known!

An easy question

What is $\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i$? A: 2.





But it never actually gets to 2. Is that a problem?

But it never actually gets to 2. Is that a problem?

No, by $\sum_{i=0}^{1} f(i)$, we really mean $\lim_{n! \, 1} \sum_{i=0}^{n} f(i)$.

[if this is undefined, so is the sum]

In this case, the partial sum is $2-(\frac{1}{2})^n$ which goes to 2.

A related question

Suppose I flip a coin of bias p, stopping when I first get heads.

What's the chance that I:

- Flip exactly once?

  Ans: p
- Flip exactly two times?

  Ans: $(1-p)p$
- Flip exactly k times?

  Ans: $(1-p)^{k-1}p$
- Eventually stop?

  Ans: 1. (assuming $p>0$)

A related question

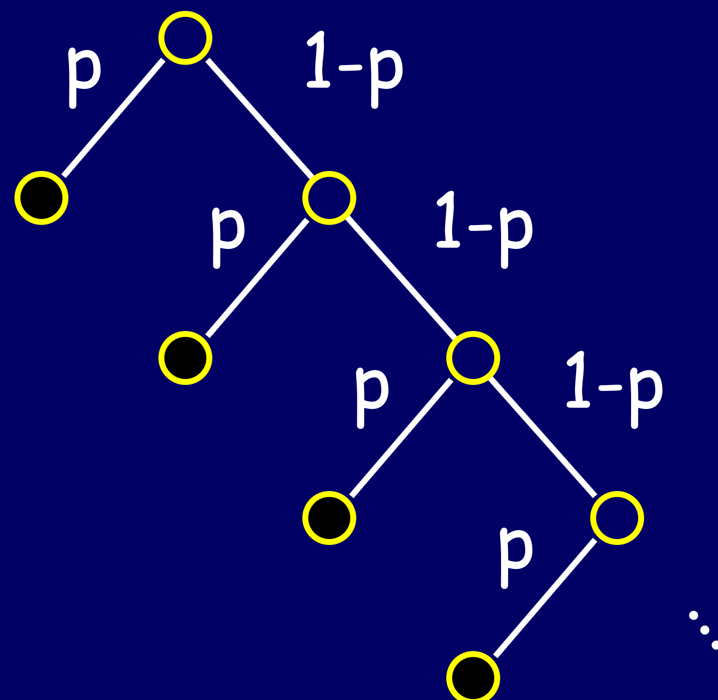
Pr(flip once) + Pr(flip 2 times) + Pr(flip 3 times) + … = 1.

So, $p + (1-p)p + (1-p)^2p + (1-p)^3p + \ldots = 1$.

Or, using $q = 1-p$,

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1-q}.$$

Pictorial view

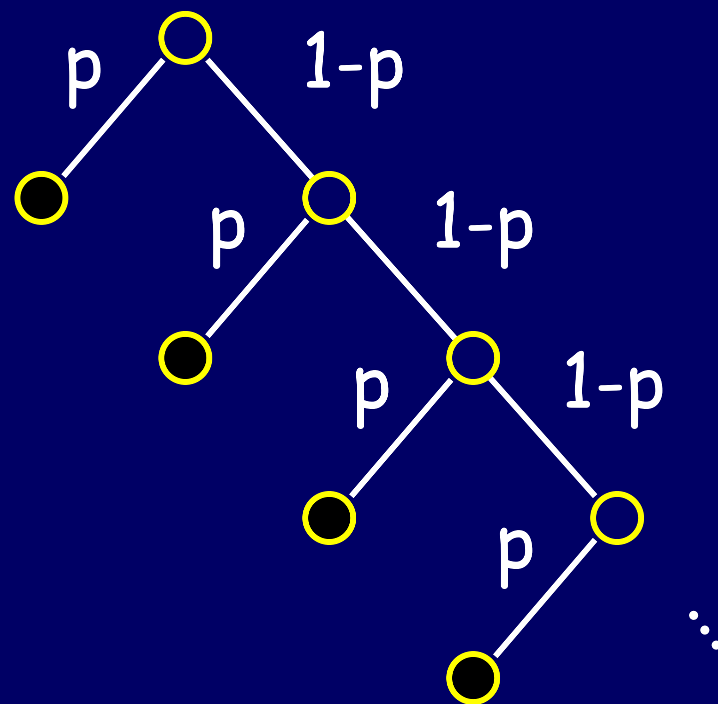


Sample space S = leaves in this tree.

Pr(x) = product of edges on path to x.

If p>0, prob of not halting by time n goes to 0 as n!1.

Use to reason about expectations too

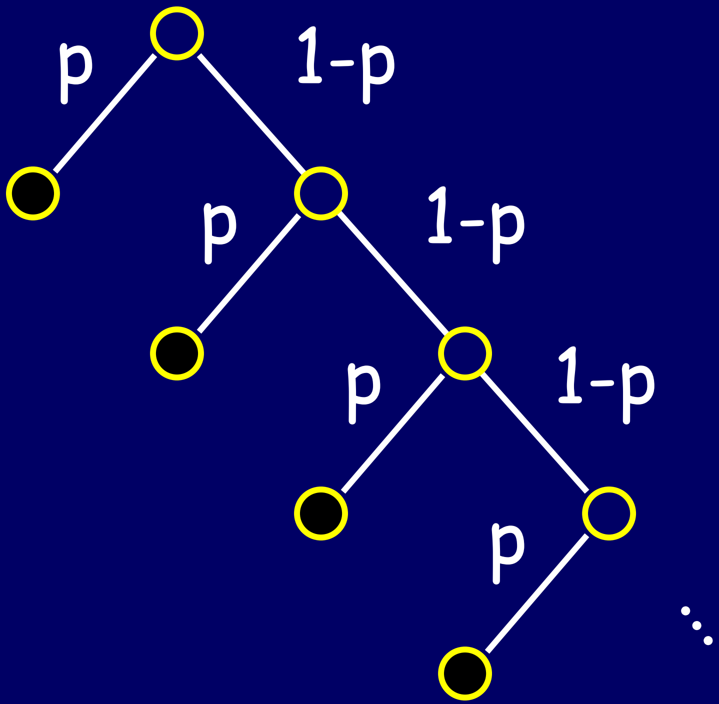


Pr(x|A)=product of edges on path from A to x.

$E[X] = \sum_x Pr(x)X(x)$.

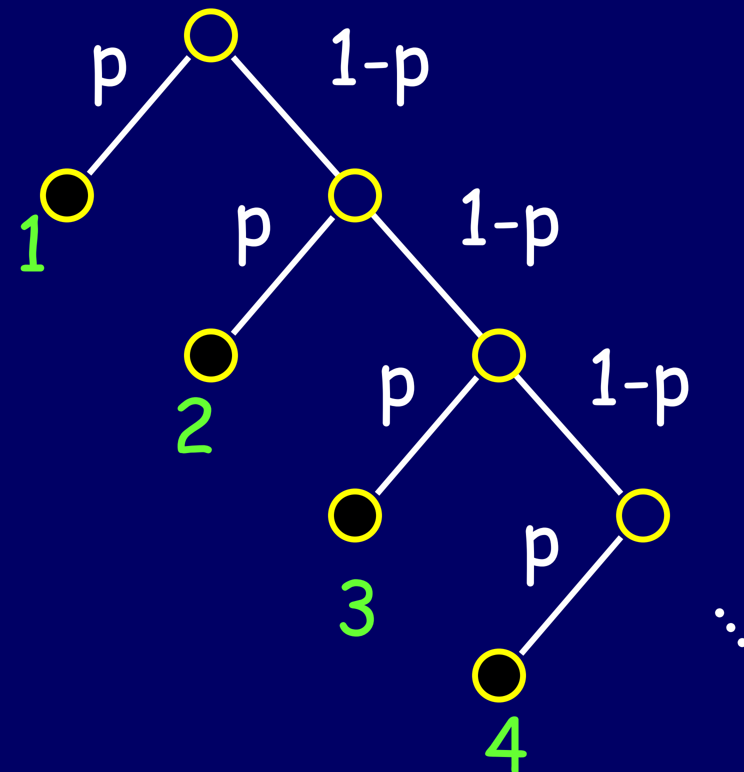
$E[X|A] = \sum_{x2\ A} Pr(x|A)X(x)$. I.e., it is as if we started the game at A.

Use to reason about expectations too



Flip bias-p coin until heads. What is expected number of flips?

Use to reason about expectations too



Let X = # flips.

Let A = event that 1st flip is heads.

$E[X] = E[X|A]Pr(A) + E[X|{:}A]Pr({:}A)$

$= 1*p + (1 + E[X])*(1-p).$

Solves to $pE[X] = p + (1-p)$, so $E[X] = 1/p$.

Infinite Probability spaces

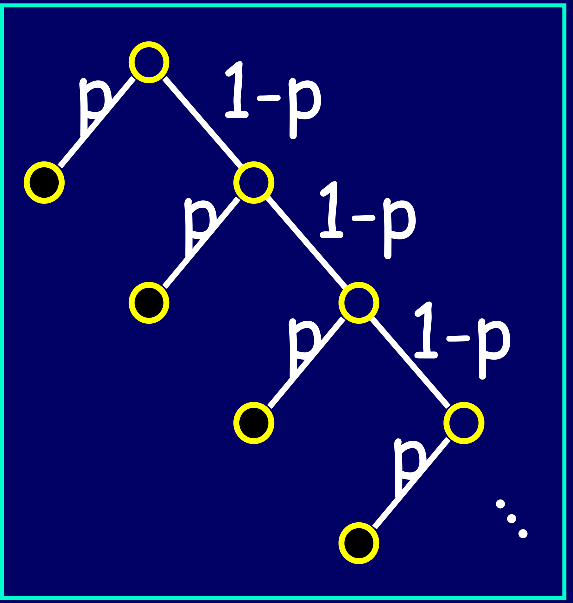
Notice we are using infinite probability spaces here, but we really only defined things for finite spaces so far.

Infinite probability spaces can sometimes be weird. Luckily, in CS we will almost always be looking at spaces that can be viewed as choice trees where

Pr(haven't halted by time t) ! 0 as t!1.

General picture

Let S be a sample space we can view as leaves of a choice tree.



Let S_n = {leaves at depth · n}.

For event A, let $A_n = A\backslash S_n$.

If $\lim_{n!1} \Pr(S_n)=1$, can define:

$$\Pr(A)=\lim_{n!1}\Pr(A_n).$$

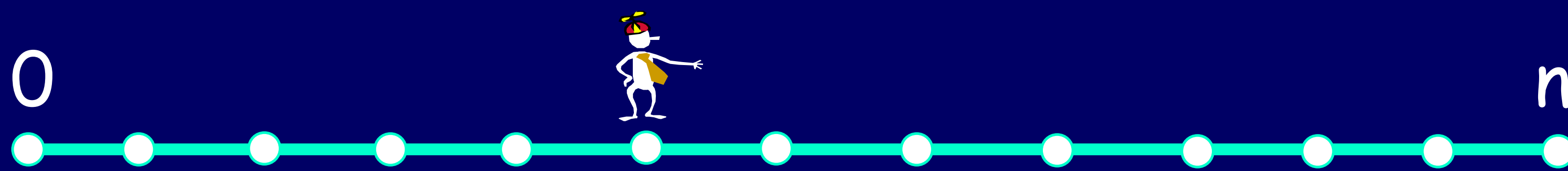
Setting that doesn't fit our model

Flip coin until #heads > 2*#tails.

There's a reasonable chance this will <u>never</u> stop...

Random walk on a line

You go into a casino with \$k, and at each time step you bet \$1 on a fair game. Leave when you are broke or have \$n.



Question 1: what is your expected amount of money at time t?

Let X_t be a R.V. for the amount of money at time t.

Random walk on a line

You go into a casino with \$k, and at each time step you bet \$1 on a fair game. Leave when you are broke or have \$n.

Question 1: what is your expected amount of money at time t?

$X_t = k + \delta_1 + \delta_2 + \ldots + \delta_t$, where δ_i is a RV for the change in your money at time i.

$E[\delta_i] = 0$, since $E[\delta_i | A] = 0$ for all situations A at time i.

So, $E[X_t] = k$.

Random walk on a line

You go into a casino with \$k, and at each time step you bet \$1 on a fair game. Leave when you are broke or have \$n.

Question 2: what is the probability you leave with \$n?

Random walk on a line

You go into a casino with \$k, and at each time step you bet \$1 on a fair game. Leave when you are broke or have \$n.

Question 2: what is the probability you leave with \$n?

One way to analyze:

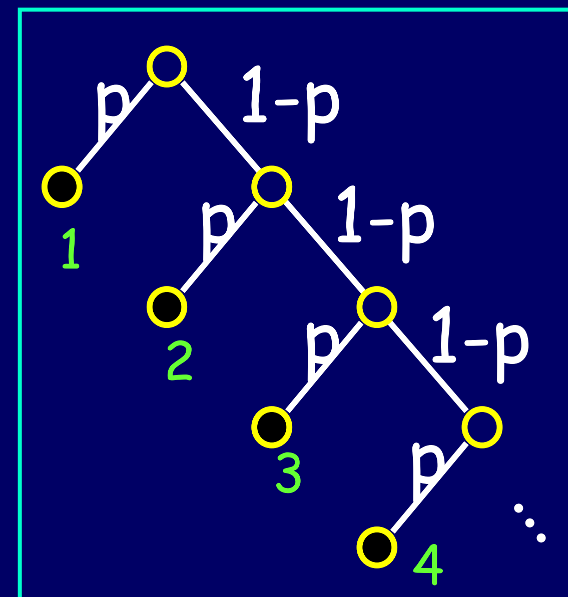
- $E[X_t] = k$.
- $E[X_t] = E[X_t|X_t=0]*Pr(X_t=0) + E[X_t|X_t=n]*Pr(X_t=n) + E[X_t|\text{neither}]*Pr(\text{neither})$.
- So, $E[X_t] = 0 + n*Pr(X_t=n) + \text{something}*Pr(\text{neither})$.
- As t! 1, Pr(neither)! 0. Also $0 < \text{something} < n$.

So, $\lim_{t!1} Pr(X_t=n) = k/n$.

So, Pr(leave with \$n) = k/n.

Expectations in infinite spaces

Let S be a sample space we can view as leaves of a choice tree.



Let S_n = {leaves at depth · n}.

Assume $\lim_{n!1} \Pr(S_n)=1$.

$E[X] = \lim_{n!1} \sum_{x2\ S} \Pr_{n}(x)X(x).$

If this limit is undefined, then the expectation is undefined. E.g., I pay you $(-2)^i$ dollars if fair coin gets i heads before a tail. Can get weird even if infinite. To be safe, should have all E[X|A] be finite.

A slightly different question

If X is a RV in dollars, do we want to maximize E[X]?

Bernoulli's St. Petersburg Paradox (1713)

Consider the following "St. Petersburg lottery" game:

- An official flips a fair coin until it turns up heads.
- If i flips needed, you win 2^i dollars.

What is E[winnings]?

How much would you pay to play?

Similar question

Which would you prefer:

(a) $1,000,000. *Or*,

(b) *A* 1/1000 chance at $1,000,000,000.

Why?

Utility Theory (Bernoulli/Cramer, 1728-1738)

Each person has his/her own utility function.
$U_i(\$1000)$ = value of \$1000 to person i.

Instead of maximizing $E[X]$ (where X is in dollars), person i wants to maximize $E[U_i(X)]$. $U_i(X)$ is a random variable.

Utility Theory

Common utility functions economists consider:

- $U(X) = \log(X)$. E.g., the amount of work you would be willing to do to double your wealth is independent of the amount of money you have.

- $U(X)$ has some asymptote: "no amount of money is worth [fill in blank]"

Utility Theory

Letters between Nicolas Bernoulli, Cramer, Daniel Bernoulli and others: 1713-1732:

see http://cerebro.xu.edu/math/Sources/Montmort/stpetersburg.pdf