Random Variables

CSE 312 Spring 21
Lecture 10

Announcements

We clarified problem 5 on HW3 (details on edge cases, like whether q can be 1).

Implicitly defining Ω

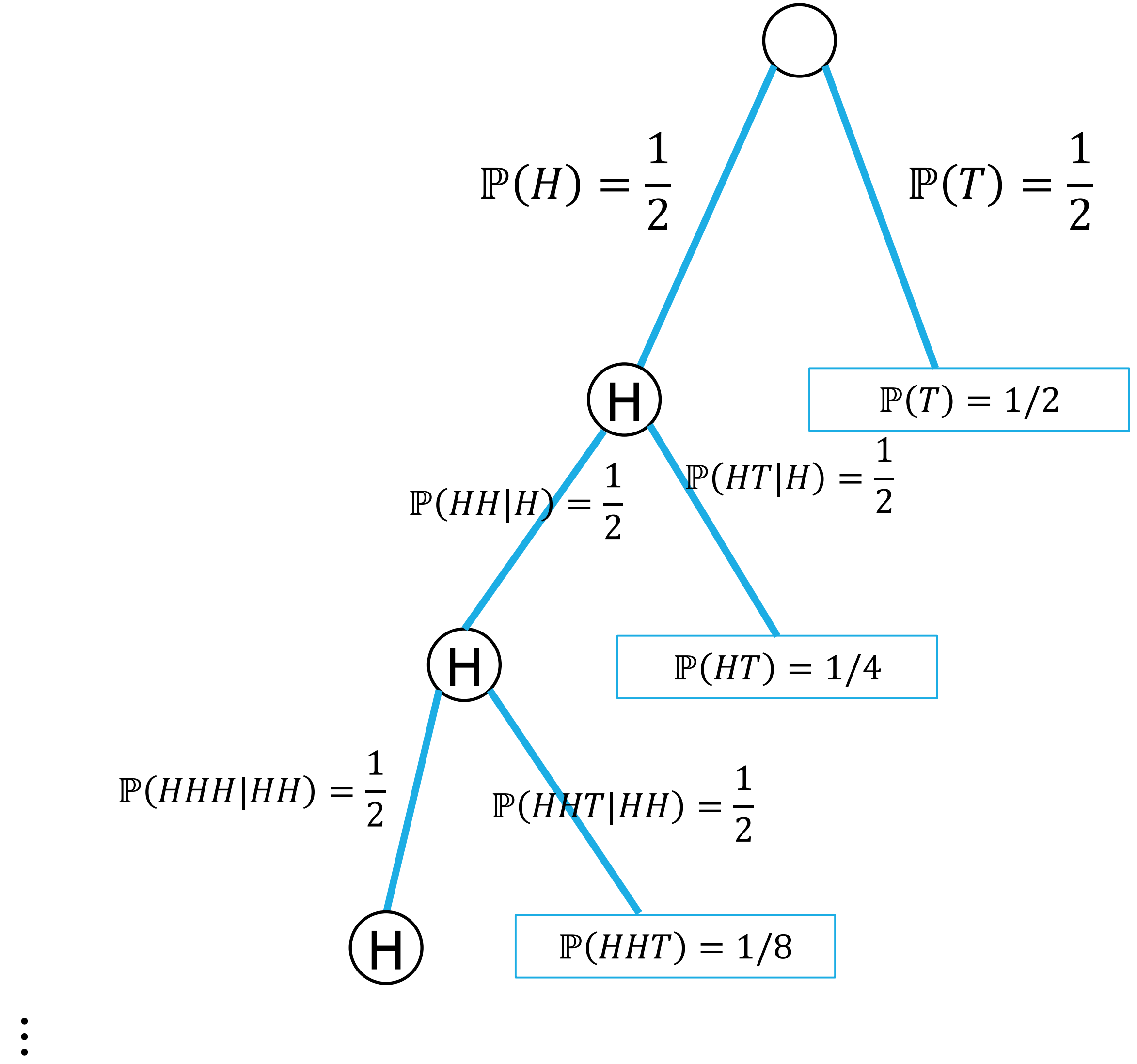
We've often skipped an explicit definition of Ω.

Often $|\Omega|$ is infinite, so we really couldn't write it out (even in principle).

How would that happen?

Flip a fair coin (independently each time) until you see your first tails.
what is the probability that you see at least 3 heads?

An infinite process.



Ω is infinite.

A sequential process is also going to be infinite…

But the tree is "self-similar"

To know what the next step looks like, you only need to look back a finite number of steps.

From every node, the children look identical (H with probability ½, continue pattern; T to a leaf with probability ½)

Finding $\mathbb{P}$(at least 3 heads)

Method 1: infinite sum.

Ω includes $H^i T$ for every i. Every such outcome has probability $1/2^{i+1}$

What outcomes are in our event?

$\sum_{i=3}^{\infty} 1/2^{i+1} = \frac{\frac{1}{2^4}}{1-1/2} = \frac{1}{8}$

Infinite geometric series, where common ratio is between -1 and 1 has closed form $\frac{\text{first term}}{1-\text{ratio}}$

Finding $\mathbb{P}$(at least 3 heads)

Method 2:

Calculate the complement

$\mathbb{P}$(at most 2 heads) = $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$

$\mathbb{P}$(at least 3 heads)= $1-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right)=\frac{1}{8}$

Random Variables

Random Variable

What's a random variable?

Formally

Random Variable

$X: \Omega \to \mathbb{R}$ **is a random variable**
$X(\omega)$ **is the summary of the outcome** ω

Informally: A random variable is a way to **summarize** the important (numerical) information from your outcome.

The sum of two dice

EVENTS

We could define

E_2 =“sum is 2”

E_3 =“sum is 3”

…

E_{12} =“sum is 12”

And ask “which event occurs”?

RANDOM VARIABLE

$X: \Omega \to \mathbb{R}$

X is the sum of the two dice.

More random variables

From one sample space, you can define many random variables.

Roll a fair red die and a fair blue die

Let D be the value of the red die minus the blue die $D(4,2) = 2$

Let S be the sum of the values of the dice $S(4,2) = 6$

Let M be the maximum of the values $M(4,2) = 4$

…

Support

The "support" (aka "the range") is the set of values X can actually take.

We called this the "image" in 311.

D (difference of red and blue) has support $\{-5, -4, -3, \dots, 4,5\}$

S (sum) has support $\{2,3, \dots, 12\}$

What is the support of M (max of the two dice)

Probability Mass Function

Often we're interested in the event $\{\omega: X(\omega) = x\}$

Which is the event…that $X = x$.

We'll write $\mathbb{P}(X = x)$ to describe the probability of that event

So $\mathbb{P}(S = 2) = \frac{1}{36}$, $\mathbb{P}(S = 7) = \frac{1}{6}$

The function that tells you $\mathbb{P}(X = x)$ is the "probability mass function"

We'll often write $f_X(x)$ for the pmf.

Partition

A random variable partitions Ω.

Let T be the number of twos in rolling a (fair) red and blue die.

$f_T(0) = 25/36$

$f_T(1) = 10/36$

$f_T(2) = 1/36$

	D2=1	D2=2	D2=3	D2=4	D2=5	D2=6
D1=1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1.6)
D1=2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
D1=3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
D1=4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
D1=5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
D1=6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Try It Yourself

There are 20 balls, numbered 1,2,…,20 in an urn.

You'll draw out a size-three subset. (i.e. without replacement)

$\Omega = \{$size three subsets of $\{1, \dots, 20\}$ $\}$, $\mathbb{P}()$ is uniform measure.

Let X be the largest value among the three balls.

If outcome is $\{4,2,10\}$ then $X = 10$.

Write down the pmf of X

Fill out the poll everywhere so Robbie knows how long to explain
Go to pollev.com/cse312

Try It Yourself

There are 20 balls, numbered 1,2,…,20 in an urn.

You'll draw out a size-three subset. (i.e. without replacement)

Let X be the largest value among the three balls.

$$f_X(x) = \begin{cases} \binom{x-1}{2}/\binom{20}{3} & \text{if } x \in \mathbb{N},\ 3 \leq x \leq 20 \\ 0 & \text{otherwise} \end{cases}$$

Good check: if you sum up $f_X(x)$ do you get 1?

Good check: is $f_X(x) \geq 0$ for all x? Is it defined for all x?

Describing a Random Variable

The most common way to describe a random variable is the PMF.

But there's a second representation:

The cumulative distribution function (CDF) gives the probability $X \leq x$

More formally, $\mathbb{P}(\{\omega: X(\omega) \leq x\})$

Often written $F_X(x) = \mathbb{P}(X \leq x)$

$F_X(x) = \sum_{i:i \leq x} f_X(i)$

Try it yourself

What is the CDF of X where

X be the largest value among the three balls. (Drawing 3 of the 20 without replacement)

Try it yourself

What is the CDF of X where

X be the largest value among the three balls. (Drawing 3 of the 20 without replacement)

$$F_X(x) = \begin{cases} 0 & \text{if } x < 3 \\ \binom{\lfloor x \rfloor}{3} / \binom{20}{3} & \text{if } 3 \leq x \leq 20 \\ 1 & \text{otherwise} \end{cases}$$

Try it yourself

What is the CDF of X where

X be the largest value among the three balls. (Drawing 3 of the 20 without replacement)

$$F_X(x) = \begin{cases} 0 & \text{if } x < 3 \\ \binom{\lfloor x \rfloor}{3} / \binom{20}{3} & \text{if } 3 \leq x \leq 20 \\ 1 & \text{otherwise} \end{cases}$$

Good checks: Is $F_X(\infty) = 1$? If not, something is wrong.

Is $F_X(x)$ increasing? If not something is wrong.

Is $F_X(x)$ defined for all real number inputs? If not something is wrong.

Two descriptions

PROBABILITY MASS FUNCTION

Defined for all $\mathbb{R}$ inputs.

Usually has “0 otherwise” as an extra case.

$\sum_x f_X(x) = 1$

$0 \le f_X(x) \le 1$

$\sum_{z:z\le x} f_X(z) = F_X(x)$

CUMULATIVE DISTRIBUTION FUNCTION

Defined for all $\mathbb{R}$ inputs.

Usually has “0 otherwise” and 1 otherwise” extra cases

Non-decreasing function

$0 \le F_X(x) \le 1$

$\lim_{x\to-\infty} F_X(x) = 0$

$\lim_{x\to\infty} F_X(x) = 1$

More Random Variable Practice

Roll a fair die n times. Let X be the number of rolls that are $5s$ or $6s$.

What is the pmf?

Don't try to write the CDF…it's a mess…

Or try for a few minutes to realize it isn't nice.

More Random Variable Practice

Roll a fair die n times. Let Z be the number of rolls that are $5s$ or $6s$.

What's the probability of getting exactly k 5's/6's? Well we need to know which k of the n rolls are 5's/6's. And then multiply by the probability of getting exactly that outcome

$$f_Z(z) = \begin{cases} \binom{n}{z} \cdot \left(\frac{1}{3}\right)^z \left(\frac{2}{3}\right)^{n-z} & \text{if } z \in Z, 0 \leq z \leq n \\ 0 & \text{otherwise} \end{cases}$$

More Practice: Infinite sequential processes

Infinite sequential process

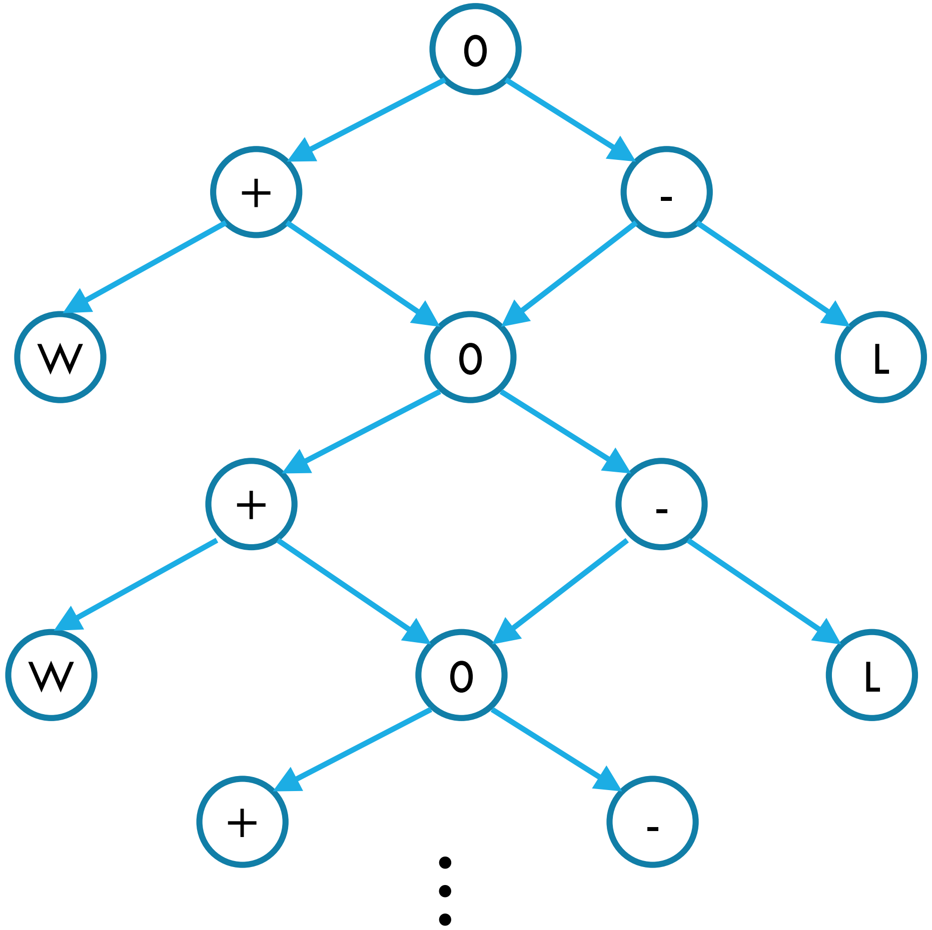
In volleyball, sets are played first team to

- Score 25 points
- Lead by at least 2

At the same time wins a set.

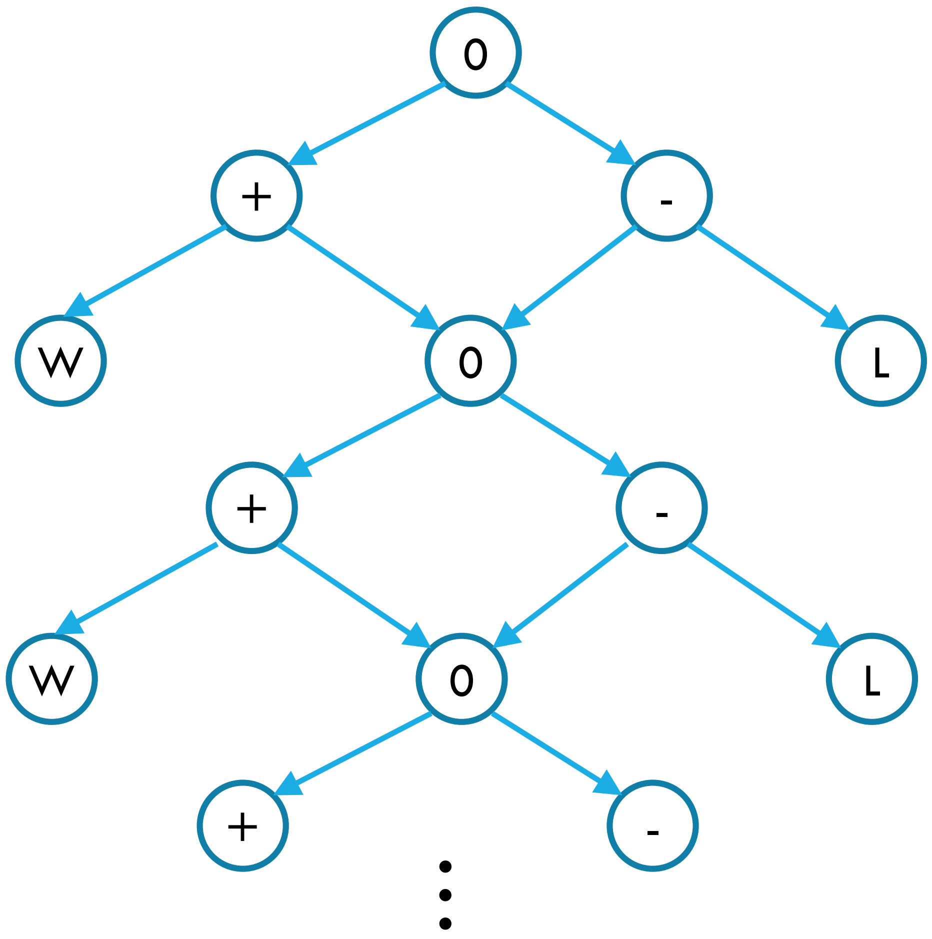
Suppose a set is 23-23. Your team wins each point independently with probability p.
What is the probability your team wins the set?

Sequential Process



$$\mathbb{P}(win\ from\ even) = p^2 + 2p(1-p)\mathbb{P}(win\ from\ even)$$

Sequential Process



$$\mathbb{P}(win\ from\ even) = p^2 + 2p(1-p)\mathbb{P}(win\ from\ even)$$

$$x - x[2p - p^2] = p^2$$
$$x[1 - 2p + p^2] = p^2$$

$$x = \frac{p^2}{p^2 - 2p + 1}$$