Random Variables

CSE 312 24Su

Lecture 8

Outline

Today we will talk about **random variables!**

> Introduce motivation and idea of random variables

> How to describe random variables and their properties
  Support, probability mass function, cumulative distribution function, expectation

Defining Events Can Be *Tedious*

For example, if we're interested in analyzing the sum of 2 random dice…

We might define events for all the possible outcomes:

E_1 ~ the sum is 1

E_2 ~ the sum is 2

E_3 ~ the sum is 3

…

E_{12} ~ the sum is 12

Defining and Using Events Can Be *Tedious*

For example, if we're interested in analyzing the sum of 2 random dice…

Now, how might we express the event that the sum is more than 6?

We could define an event $A \sim \textbf{the sum is more than 6}$ or since that's a bit undescriptive we might look for $P(E_7 \cup E_8 \cup E_9 \cup E_{10} \cup E_{11} \cup E_{12})$

We want a way to easily say something "[the sum of the dice] > 6" or be able to summarize things like "what is [the sum of the dice] on average?"

Random Variables

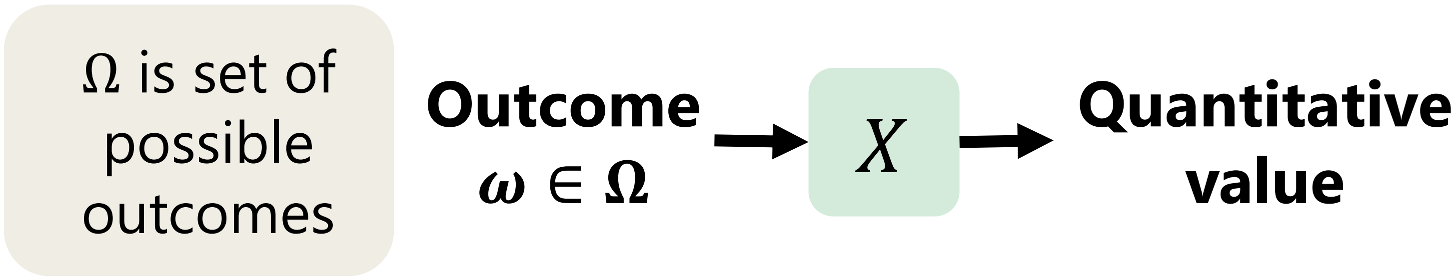
Random Variable

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

Random Variable

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

Formally: *Function* assigning a value to outcomes of a random experiment



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"?

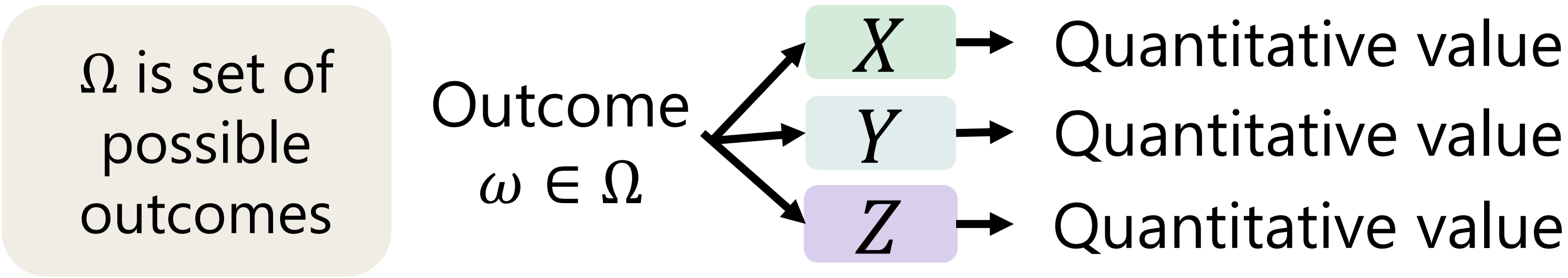
RANDOM VARIABLE

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

X is the sum of the two dice.

More random variables

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



E.g., Roll a fair red die and a fair blue die, Ω is the set of possible outcomes

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 $M(4,2) = 4$

…

Notational Notes

> We will always use capital letters for random variables.

> It’s common to use lower-case letters for the values they could take on.

> When we say $X = 2$, we are referring to the set of outcomes that the random variable X assigns the value 2

For example, if X is the number of heads in three coin flips, $X = 2$ corresponds to the set of outcomes $\{HHT, HTH, THH\}$

Notational Notes

> We will always use capital letters for random variables.

> It's common to use lower-case letters for the values they could take on.

> When we say $X = 2$, we are referring to the set of outcomes that the random variable X assigns the value 2

For example, if X is the number of heads in three coin flips, $X = 2$ corresponds to the set of outcomes $\{HHT, HTH, THH\}$

Note that $X = 2$ is a set of outcomes, so it is an event!

How do we describe random variables?

Random variable gives a *quantitative property* of an outcome in a random experiment.

e.g., the number of coin flips till the first head, or the sum of two dice rolls

> **Support**

e.g., what are the possible "number of coin flips" it could possibly take?

> **Probability Mass Function**

e.g., what's the probability it takes 3 coin flips till the 1^{st} head? what about 5 flips?

> **Cumulative Distribution Function**

e.g., what's the probability it takes *less than* 3 coin flips till the 1^{st} head?

> **Expectation**

e.g., how many coin flips can we *expect* it to take till the first head on average?

> **Variance**

e.g., on average, how much does the "number of coin flips" deviate from the expectation?

Support

The "support" (aka "the range") is the set of values X can actually take.
We called this the "image" in 311.

E.g., We roll a red and a blue dice and define these 3 random variables:

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

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

M (max of the two dice) has support ______________

Each value in the support corresponds to some outcome(s) from Ω

Support

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

We called this the "image" in 311.

We roll a red and a blue dice.

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

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

M (max of the two dice) has support $\{1,2,3,4,5,6\}$

Probability Mass Function

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

Which is the event…that $X = k$.

e.g., $S = 7$ is the event that two dice sum to 7

⚠️ A random variable X is <u>not</u> an event.

But, $\boldsymbol{X = 2}$ **is an event** – it's the event/set of outcomes where the random variable takes on the value 2

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

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

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

We'll often write $p_X(k) = \mathbb{P}(X = k)$ for the PMF.

Probability Mass Function

Let T be the number of 2's rolling a (fair) red and blue die.

What is the range of T*?* $\Omega_T = \{0,1,2\}$

What is the PMF of T*?*

$p_T(0) = 25/36$, $p_T(1) = 10/36$, $p_T(2) = 1/36$

$$p_T(k) = \begin{cases} 25/36 & k = 0 \\ 10/36 & k = 1 \\ 1/36 & k = 2 \\ 0 & \text{otherwise} \end{cases}$$

> Often use **piecewise function** to give probabilities for different values of k
> This is a function, so include **otherwise case**, so it is defined for all values

Partition

A random variable partitions Ω.

$\sum_{k \in \Omega_x} p_X(k) = 1$

Let T be the number of 2's rolling a (fair) red and blue die.

$p_T(0) = 25/36$

$p_T(1) = 10/36$

$p_T(2) = 1/36$

	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)

Partition

A random variable partitions Ω.

$\sum_{k \in \Omega_x} p_X(k) = 1$

Let T be the number of 2's rolling a (fair) red and blue die.

$p_T(0) = 25/36$

$p_T(1) = 10/36$

$p_T(2) = 1/36$

	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)

Partition

A random variable partitions Ω.

$\sum_{k \in \Omega_x} p_X(k) = 1$

Let T be the number of 2's rolling a (fair) red and blue die.

$p_T(0) = 25/36$

$p_T(1) = 10/36$

$p_T(2) = 1/36$

	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)

Partition

A random variable partitions Ω.

$\sum_{k \in \Omega_x} p_X(k) = 1$

Let T be the number of 2's rolling a (fair) red and blue die.

$p_T(0) = 25/36$

$p_T(1) = 10/36$

$p_T(2) = 1/36$

	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 $\{1, \ldots, 20\}$ $\}$, $\mathbb{P}()$ is uniform measure.

Let X be the largest value among the three balls.
e.g., if the outcome is $\{4,2,10\}$ then $X = 10$.

> What is the **support of $\boldsymbol{X}$**?
> Write down the **PMF of $\boldsymbol{X}$**
i.e., what is $p_X(k) = \mathbb{P}(X = k)$?

Fill out the poll everywhere so Claris 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)

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

Let X be the largest value among the three balls.

$\Omega_X =$

$p_X(k) = \Big\{$

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.

$$\Omega_X = \{3,4,5,\ldots,19,20\}$$

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

Good checks:
if you sum up $p_X(k)$ do you get 1?
is $p_X(k) \geq 0$ for all k? Is it defined for all k?

Cumulative Distribution Function (CDF)

The **PMF gives the probability $\boldsymbol{X = k}$**
(and is the most common way to describe a random variable)

There's a second representation:

The **cumulative distribution function (CDF) gives the probability $\boldsymbol{X \leq k}$**

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

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

$F_X(k) = \sum_{i:i\leq k} p_X(i)$
"sum up the probabilities of X taking all possible numbers less than k"

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.

$\Omega_X = \{3,4,5, \dots, 19, 20\}$

$$F_X(k) = \mathbb{P}(X \leq k) = \begin{cases} \\ \\ \end{cases}$$

Think "what is the probability the largest value is less than or equal to 10?"

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.

$\Omega_X = \{3,4,5, \dots, 19, 20\}$

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

Think "what is the probability the largest value is less than or equal to k?"

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.

$$\Omega_X = \{3,4,5, \dots, 19, 20\}$$

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

Good checks:

Is $F_X(\infty) = \mathbb{P}(X \leq \infty) = 1$? If not, something is wrong.

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

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

Summary (PMF and CDF)

PROBABILITY MASS FUNCTION

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

Usually has "0 otherwise" as an extra case.

$\sum_x p_X(x) = 1$

$0 \leq p_X(x) \leq 1$

$\sum_{i:i\leq k} p_X(k) = F_X(k)$

CUMULATIVE DISTRIBUTION FUNCTION

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

Often has "0 otherwise" and 1 otherwise" extra cases

Non-decreasing function

$0 \leq F_X(k) \leq 1$

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

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

Expectation

Expectation

Expectation

The "expectation" (or "expected value") of a random variable X is:

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot \mathbb{P}(X = k)$$

Intuition: The **weighted average of values X could take on**.

Weighted by the probability you actually see them.

e.g., if $Y \sim$ num. flips to get a head,

$E[Y]$ is num. of flips we expect it to take *on average*

Example 1

Flip a fair coin twice (independently).

What is the *expected number of heads* we see?

Let X be the number of heads.

Example 1

Flip a fair coin twice (independently).
What is the *expected number of heads* we see?

Let X be the number of heads.

$\Omega = \{TT, TH, HT, HH\}$, $\mathbb{P}()$ is uniform measure.

$\Omega_X = \{0,1,2\}$

$$\begin{aligned}\mathbb{E}[X] &= p_X(0) \cdot 0 + p_X(1) \cdot 1 + p_X(2) \cdot 2 \\ &= \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 = 0 + \frac{1}{2} + \frac{1}{2} = 1.\end{aligned}$$

Example 2

You roll a biased die.

It shows a 6 with probability $\frac{1}{3}$, and 1,…,5 with probability $2/15$ each.

Let X be the value of the die. What is $\mathbb{E}[X]$?

1. Write the PMF for X: $p_X(k) = \begin{cases} 2/15 & k \in \{1,2,3,4,5\} \\ 1/3 & k = 6 \\ 0 & \text{otherwise} \end{cases}$

2. Plug in formula: $\mathbb{E}[X] = \frac{2}{15} \cdot 1 + \frac{2}{15} \cdot 2 + \frac{2}{15} \cdot 3 + \frac{2}{15} \cdot 4 + \frac{2}{15} \cdot 5 + \frac{1}{3} \cdot 6 = 4$

$\mathbb{E}[X]$ is not just the most likely outcome!

Try it yourself

> Let X be the result shown on a fair die. What is $\mathbb{E}[X]$?

> Let Y be the sum of two (independent) fair die rolls. What is $\mathbb{E}[Y]$?

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

Try it yourself

Let X be the result shown on a fair die. What is $\mathbb{E}[X]$

1. Write the PMF: $p_X(k) = \begin{cases} 1/6 & k \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$

2. Plug into formula: $\mathbb{E}[X] = 6 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6}$

$$= \frac{21}{6} = 3.5$$

$\mathbb{E}[X]$ is not necessarily a possible outcome!

That's ok, it's an average!

Try it yourself

Let Y be the sum of two (independent) fair die rolls. What is $\mathbb{E}[Y]$?

$\mathbb{E}[Y] =$

$\frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \frac{3}{36} \cdot 4 + \frac{4}{36} \cdot 5 + \frac{5}{36} \cdot 6 + \frac{6}{36} \cdot 7 + \frac{5}{36} \cdot 8 + \frac{4}{36} \cdot 9 + \frac{3}{36} \cdot 10 + \frac{2}{36} \cdot 11 + \frac{1}{36} \cdot 12$

$= 7$

$\mathbb{E}[Y] = 2\mathbb{E}[X]$. That's not a coincidence…we'll talk about why next time!

Subtle but Important

X is random. You don't know what it is (at least until you run the experiment).

$\mathbb{E}[X]$ is not random. It's a number.

You don't need to run the experiment to know what it is.

Summary

Today we have talked about **random variables** which assign quantitative values to outcomes of a random experiment.

How to describe a random variable?

> **Range/Support:** $\mathbf{\Omega_X}$ is set of possible values X can be

> **Probability Mass Function:** $p_X(k) = \mathbb{P}(X = k)$

> **Cumulative Distribution Function:** $F_X(k) = \mathbb{P}(X \leq k)$

> **Expectation**: A single value that is the weighted average of values on the support, weighted on the probabilities X takes on each

> **Variance:** Coming soon!

More Practice: Infinite sequential processes

Infinite sequential process

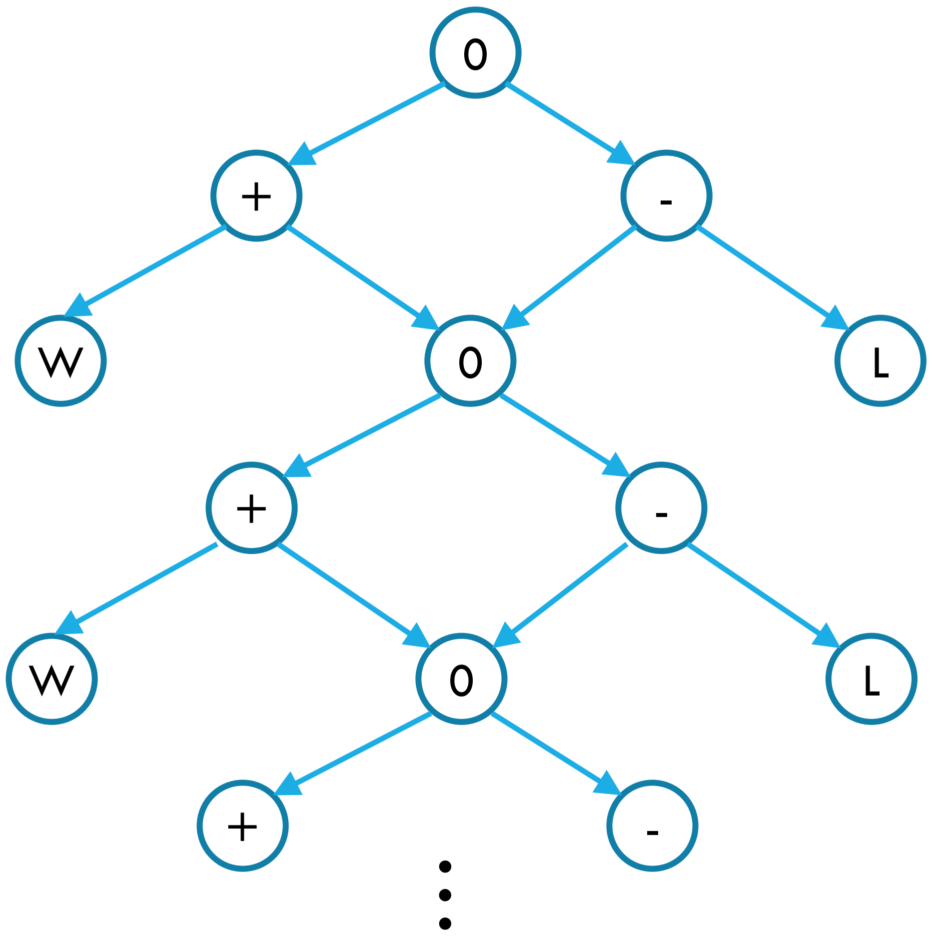
In volleyball, sets are played first team to

- Score 25 points
- Lead by at least 2

At the same time wins a set.

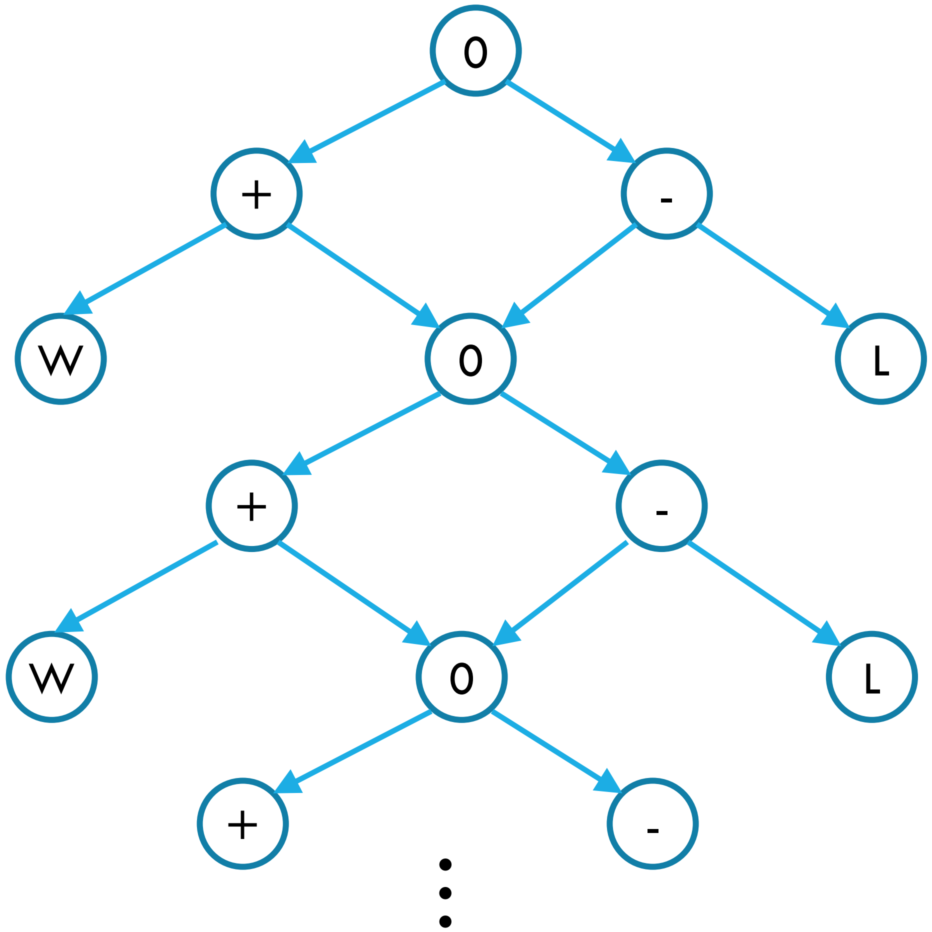
Suppose a set is 23-23. Your team wins each point independently with probability p.
What is the probability your team wins the set?

Sequential Process



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

Sequential Process



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

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

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

More Random Variable Practice

Roll a fair die n times. Let Z 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

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