Random Variables

CSE 312 Spring 24
Lecture 10

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.

Support Ω_X
the set of values X can take.

Probability Mass Function (pmf $p_X(x)$) on input x, tells you $\mathbb{P}(X = x)$.

$p_X(7) = \frac{1}{6}$

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

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

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)

$$\frac{\binom{7}{3}}{\binom{20}{3}}$$

$$P(X \le x)$$

$$\Updownarrow$$

$$P(B_1 \le x \cap B_2 \le x \cap B_3 \le x)$$

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

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)

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

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)

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

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.

$\sum_x p_X(x) = 1$

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

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

CUMULATIVE DISTRIBUTION FUNCTION

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

Often 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 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.

$$p_Z(z) = \begin{cases} \binom{n}{z}\left(\frac{1}{3}\right)^z\left(\frac{2}{3}\right)^{n-z} \\ \end{cases}$$

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

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

Expectation

Expectation

$\mathbb{E}[X]$

Expectation

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

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

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

Weighted by the probability you actually see them.

Example 1

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

Flip a fair coin twice (independently)

Let X be the number of heads.

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

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

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 $\mathbb{E}[X]$?

$$\frac{1}{3}\cdot 6+\frac{2}{15}\cdot 5+\frac{2}{15}\cdot 4+\frac{2}{15}\cdot 3+\frac{2}{15}\cdot 2+\frac{2}{15}\cdot 1$$

$$=2+\frac{2(5+4+3+2+1)}{15}=2+\frac{30}{15}=4$$

$\mathbb{E}[X]$ 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 Robbie knows how long to explain~~
~~Go to pollev.com/robbie~~

Try it yourself

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

$$6 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6}$$

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

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

That's ok, it's an average!

Try it yourself

$\mathbb{E}[Y] =$

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

$= 7$

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

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.

More Independence

Independence of events

Recall the definition of independence of **events**:

Independence

Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Independence for 3 or more events

For three or more events, we need two kinds of independence

Pairwise Independence

Events $A_1, A_2, \dots, A_n$ are pairwise independent if
$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$ for all i, j

Mutual Independence

Events $A_1, A_2, \dots, A_n$ are mutually independent if
$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k})$
for every subset $\{i_1, i_2, \dots, i_k\}$ of $\{1, 2, \dots, n\}$.

Pairwise Independence vs. Mutual Independence

Roll two fair dice (one red one blue) independently

$R =$ "red die is 3"

$B =$ "blue die is 5"

$S =$ "sum is 7"

How should we describe these events?

Pairwise Independence

R, B, S are pairwise independent

$\mathbb{P}(R \cap B) \mathrel{?=} \mathbb{P}(R)\mathbb{P}(B)$

$\frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$ Yes! (These are also independent by the problem statement)

$\mathbb{P}(R \cap S) \mathrel{?=} \mathbb{P}(R)\mathbb{P}(S)$

$\frac{1}{36} \mathrel{?=} \frac{1}{6} \cdot \frac{1}{6}$ Yes!

$\mathbb{P}(B \cap S) \mathrel{?=} \mathbb{P}(B)\mathbb{P}(S)$

$\frac{1}{36} \mathrel{?=} \frac{1}{6} \cdot \frac{1}{6}$ Yes!

Since all three pairs are independent, we say the random variables are pairwise independent.

Mutual Independence

R, B, S are not mutually independent.

$\mathbb{P}(R \cap B \cap S) = 0$; if the red die is 3, and blue die is 5 then the sum is 8 (so it can't be 7)

$$\mathbb{P}(R)\mathbb{P}(B)\mathbb{P}(S) = \left(\frac{1}{6}\right)^3 = \frac{1}{216} \neq 0$$

Checking Mutual Independence

It's not enough to check just $\mathbb{P}(A \cap B \cap C)$ either.

Roll a fair 8-sided die.

Let A be $\{1,2,3,4\}$

B be $\{2,4,6,8\}$

C be $\{2,3,5,7\}$

$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{2\}) = \frac{1}{8}$

$\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

Checking Mutual Independence

It's not enough to check just $\mathbb{P}(A \cap B \cap C)$ either.

Roll a fair 8-sided die.

Let A be $\{1,2,3,4\}$

B be $\{2,4,6,8\}$

C be $\{2,3,5,7\}$

$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{2\}) = \frac{1}{8}$

$\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

But B and C aren't independent. Because there's a subset that's not independent, A, B, C are not mutually independent.

Checking Mutual Independence

To check mutual independence of events:

Check **every** subset.

To check pairwise independence of events:

Check **every** subset of size two.

Independence of Random Variables

That's for events...what about random variables?

Independence (of random variables)

X **and** Y **are independent if for all** k, ℓ

$$\mathbb{P}(X = k, Y = \ell) = \mathbb{P}(X = k)\mathbb{P}(Y = \ell)$$

We'll often use commas instead of ∩ symbol.

Independence of Random Variables

The “for all values” is important.

We say that the event “the sum is 7” is independent of “the red die is 5”

What about $S =$“the sum of two dice” and $R =$“the value of the red die”

Independence of Random Variables

The “for all values” is important.

We say that the event “the sum is 7” is independent of “the red die is 5”

What about $S =$“the sum of two dice” and $R =$“the value of the red die”

NOT independent.

$\mathbb{P}(S = 2, R = 5) \neq \mathbb{P}(S = 2)\mathbb{P}(R = 5)$ (for example)

Independence of Random Variables

Flip a coin independently $2n$ times.

Let X be “the number of heads in the first n flips.”

Let Y be “the number of heads in the last n flips.”

X and Y are independent.

Mutual Independence for RVs

A little simpler to write down than for events

Mutual Independence (of random variables)

$X_1, X_2, \ldots, X_n$ **are mutually independent if for all** $x_1, x_2, \ldots, x_n$

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_n = x_n)$$

DON'T need to check all subsets for random variables…

But you do need to check all values (all possible x_i) still.

What does Independence give you?

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

More Practice: Infinite sequential processes

Infinite sequential process

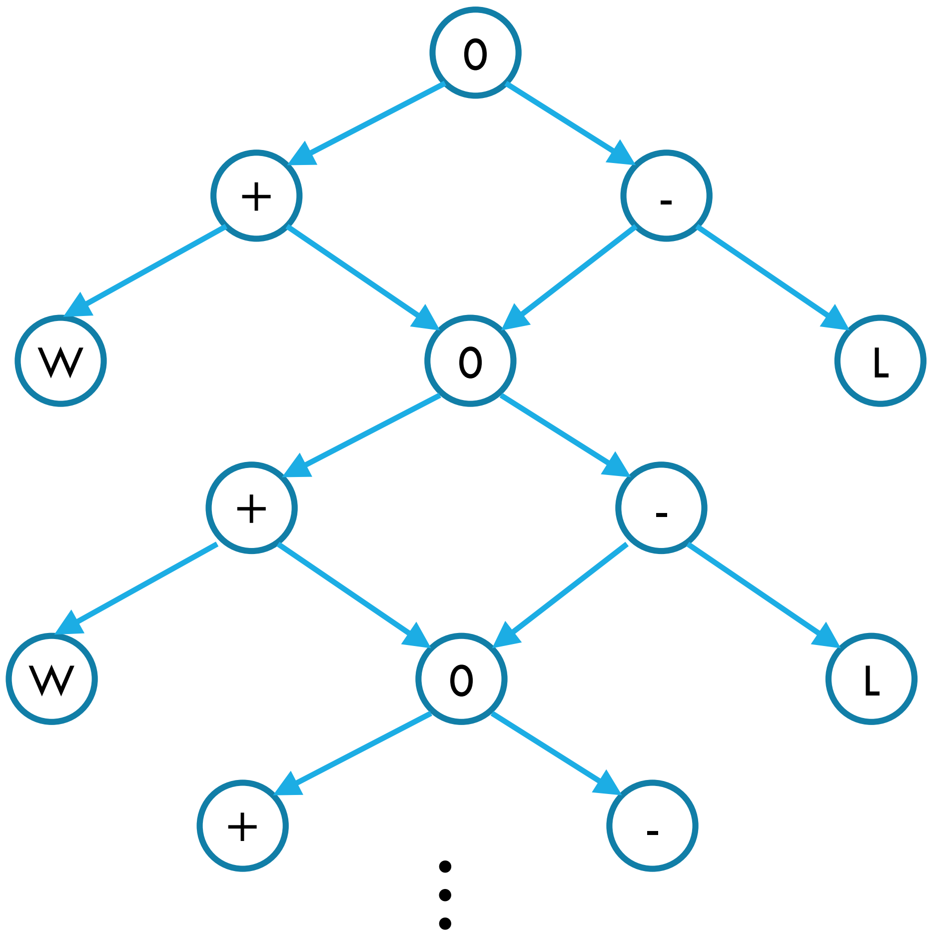
In volleyball, sets are played first team to

- Score 25 points
- Lead by at least 2

At the same time wins a set.

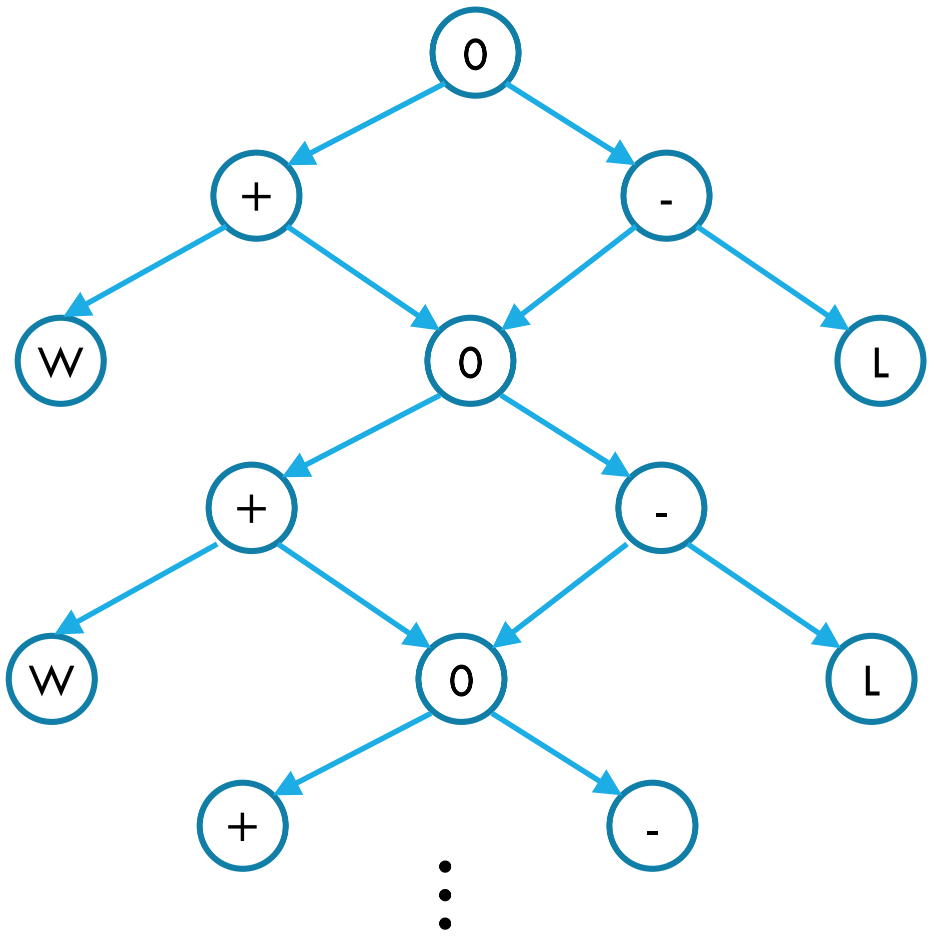
Suppose a set is 23-23. Your team wins each point independently with probability p.
What is the probability your team wins the set?

Sequential Process



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

Sequential Process



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

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

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