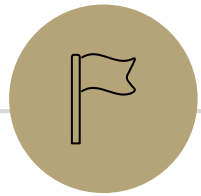
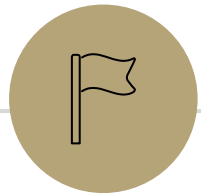


Random Variables

CSE 312 Summer 25
Lecture 8



Quiz 2 – Probability



Setting the stage: Random Variables

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).

T

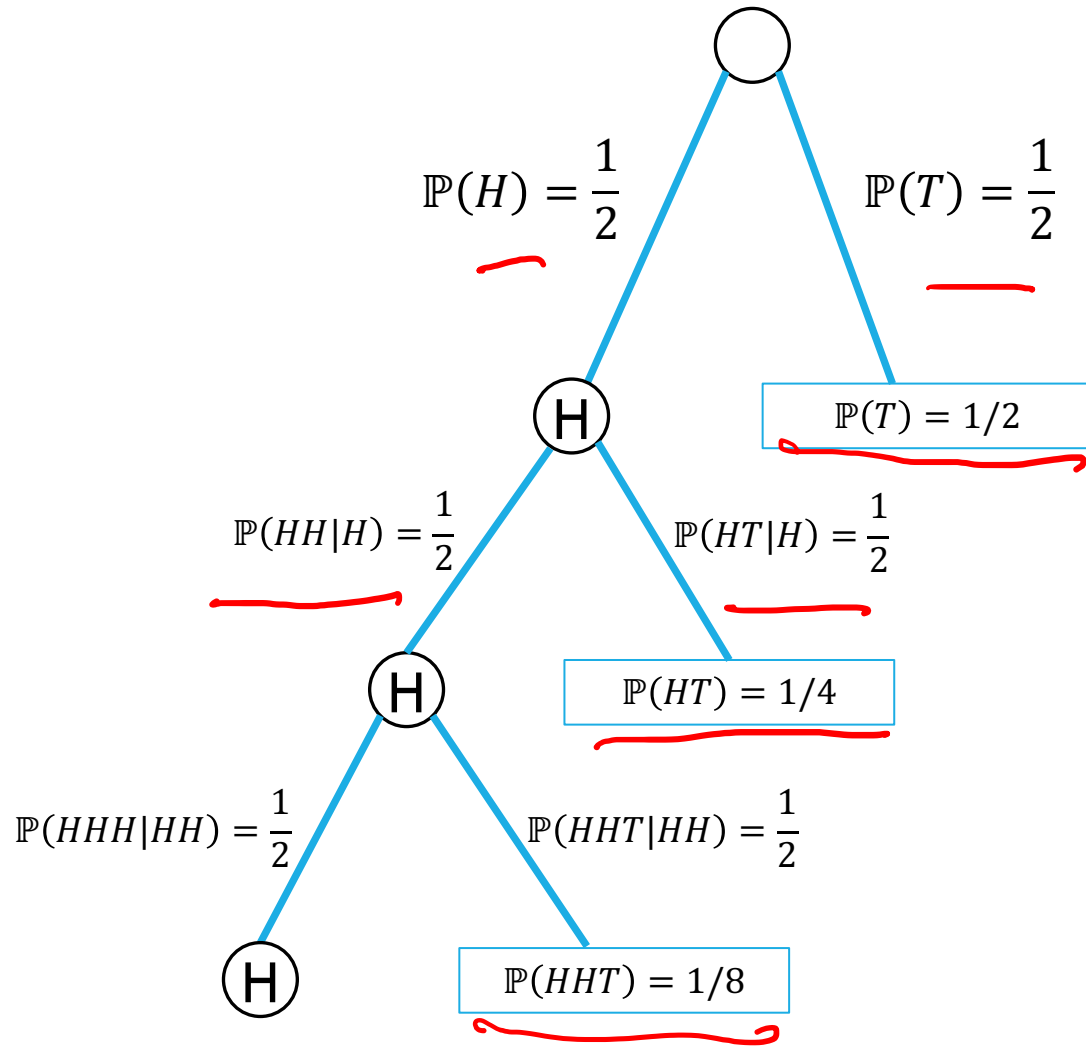
HT

HH T

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"

From every node, the children look identical (H with probability $\frac{1}{2}$, continue pattern; T to a leaf with probability $\frac{1}{2}$)

⋮

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

$$\underbrace{H H H T}_{\text{bracket}} = \frac{1}{2^{3+1}}$$

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?

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

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$$

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

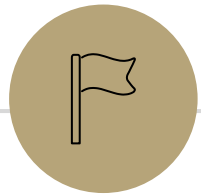
Finding $\mathbb{P}(\text{at least 3 heads})$

Method 2: Calculate the complement

0, 1, 2

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

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



Random Variables



Random Variable

$(5, 2)$

$$X((5, 2)) = 7$$

Formally

Random Variable

$X: \Omega \rightarrow \mathbb{R}$ 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"

RANDOM VARIABLE

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

X is the sum of the two dice.

And ask "which event occurs"?

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 $M(4,2) = 4$

...

Notational Notes

~~X~~

~~X~~ = x

We will always use capital letters for random variables.

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

Formally random variables are functions, so you'd think we'd write

$X(H, H, T) = 2$ \leftarrow

But we nearly never do. We just write $X = 2$ \leftarrow

Support (Ω_X)



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

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)?

$(1, 1)$



$(6, 6)$
 $(5, 6)$
 $(4, 6)$

Probability Mass Function

$P_y(y)$

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

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

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

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

Partition

A random variable partitions Ω .

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

$$p_T(0) = \frac{25}{36}$$

$$p_T(1) = \frac{10}{36}$$

$$p_T(2) = \frac{1}{36}$$

$$P_T(x) = 0$$

	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)

otherwise

Try It Yourself

$$X = 7 \quad \binom{7-1}{2} = \binom{6}{2}$$

There are 20 balls, numbered 1,2,...,20 in a jar.

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

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

Let X be the largest value among the three balls.

$$\frac{\binom{6}{2}}{\binom{20}{3}}$$

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

Write down the pmf of X ←

Try It Yourself

There are 20 balls, numbered 1,2,...,20 in a jar.

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

Let X be the largest value among the three balls.

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

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

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

Random Variable

What's a random variable?

Formally

Random Variable

$X: \Omega \rightarrow \mathbb{R}$ 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 $\mathbb{P}(X = x)$.

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)

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.

$$\sum_x p_X(x) = 1$$

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

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

CUMULATIVE DISTRIBUTION FUNCTION

Defined for all \mathbb{R} inputs.

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

Non-decreasing function

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

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

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

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...

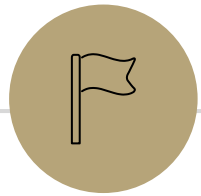
Or try for a few minutes to realize it isn't nice.

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

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



More Independence

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 \cap 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, \dots, X_n are mutually independent if for all x_1, x_2, \dots, x_n

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

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?

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

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