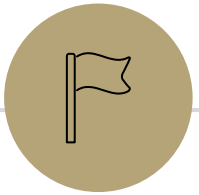


Random Variables

CSE 312 26Su

Lecture 8



Quiz 2

Where Are We?

Last time:

- Review
 - Counting
 - Uniform Probability Space
 - Conditional Probability
- Intro to Random Variables

Today:

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- 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*

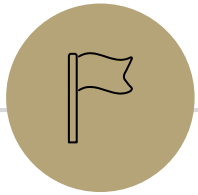
For example, if we're interested in analyzing the sum of 2 random dice...

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

We could define an event $A \sim$ 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

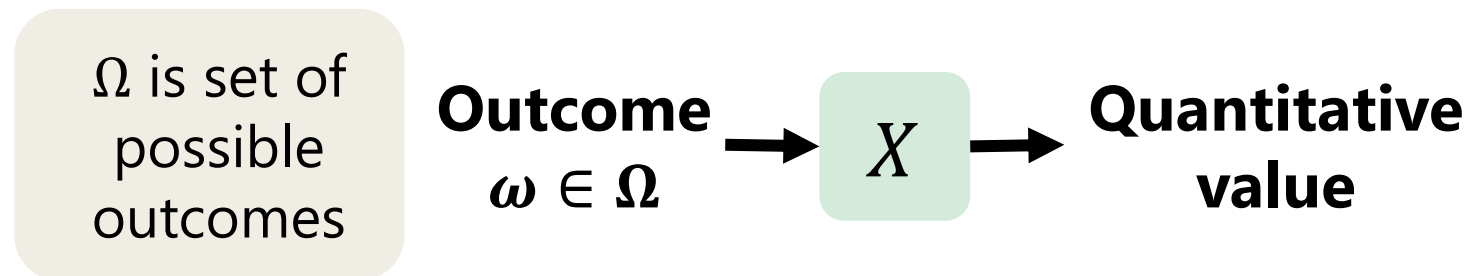
Random Variable

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

Random Variable

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

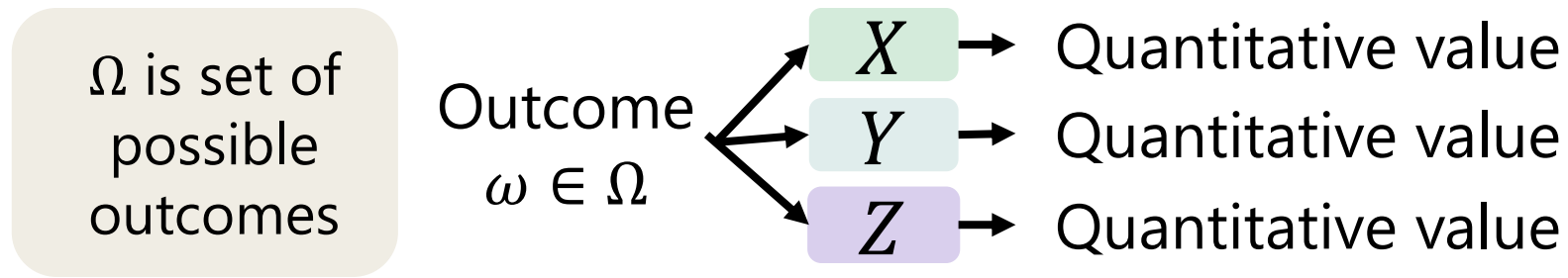
RANDOM VARIABLE

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

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 $D(\underline{4}, \underline{2}) = \underline{2}$

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

Let M be the maximum of the values $M(\underline{4}, \underline{2}) = \underline{4}$

...

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 $\{HHT, HTH, THH\}$

$$X(\underline{HHT}) = 2$$

$$X(\underline{HTH}) = 2$$

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 1st head? what about 5 flips?

> Cumulative Distribution Function

e.g., what’s the probability it takes *less than* 3 coin flips till the 1st 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.

$$R = 1 \quad B = 6 \quad D = 1 - 6 = -5$$

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

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

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

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

$$r = 6 \quad B = 1$$

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 $\{-5, -4, -3, \dots, 4, 5\}$

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

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

Probability Mass Function

$$P(X) = ??$$

⚠ A random variable X is not an event.

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

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

sum of 2 dice

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

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

$R=1, B=1$

$R=1, B=6$
 $R=2, B=5$

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.



$P_S(k)$

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 ? $|E| = 2 \cdot 1 \cdot 5$

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

$P(T=0)$

$P(T=1)$

$P(T=2) = 1/36$

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

> 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 .

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

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

$T=0$
 $T=1$
 $T=2$

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.

$$p_T(0) = 25/36$$



$$p_T(1) = 10/36$$

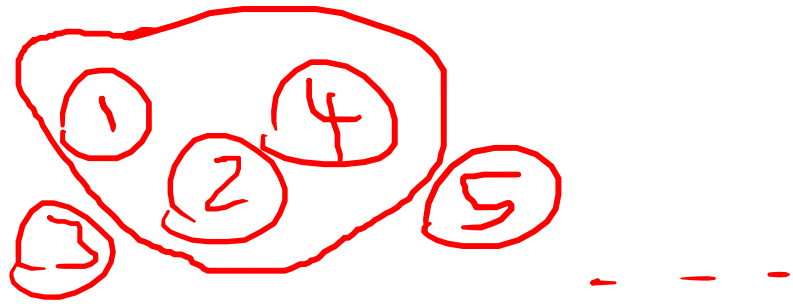


$$p_T(2) = 1/36$$



	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)

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

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 X** ?
range of values X can take on
- > Write down the **PMF of X**
i.e., what is $p_X(k) = \mathbb{P}(X = k)$?

Try It Yourself

$$\{1, 2, 3, \dots, 20\}$$

$$(1, 2, \text{---})$$

$$|\Omega| \neq |\Omega \times \Omega|$$

$$|\Omega| = \binom{20}{3}$$

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

$$1 \ 2 \ 3 \ 4 \ \text{---} \ 6 \ 7$$

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

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

Let X be the largest value among the three balls.

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

$$P(X=k) = \frac{|\Omega_k|}{|\Omega|}$$

$$p_X(k) = \begin{cases} \binom{k-1}{2} / \binom{20}{3} \\ 0 \end{cases}$$

$$3 \leq k \leq 20, k \in \mathbb{N}$$

$$k \in \{3, 4, 5, \dots, 20\}$$

otherwise

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.

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

defined for all $\mathbb{R}$

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

$$P(X = k)$$

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 $X = k$
(and is the most common way to describe a random variable)

There's a second representation:

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

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

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

$$= \mathbb{P}(X \leq k-1) + \mathbb{P}(X = k)$$

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

"sum up the probabilities of X taking all possible numbers less than k "

Try it yourself

1 2 3 4 5 K 7 8 - 20
k=6

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

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.

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

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

$P(X \leq 6) = P(X \leq 6.3)$

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

Try it yourself

There are 20 balls, numbered $1, 2, \dots, 20$ in an urn.

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.

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

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

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.

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

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

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

$$\underline{P(X \leq k)}$$

CUMULATIVE DISTRIBUTION FUNCTION

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

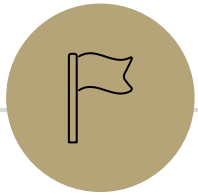
Often has "0 otherwise" and 1 otherwise" extra cases

Non-decreasing function

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

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

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



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.

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

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

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

Example 2

You roll a biased die.

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

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

Example 2

You roll a biased die.

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

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

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

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 $\mathbb{E}[Y]$?

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:
$$\begin{aligned} \mathbb{E}[X] &= 6 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

$\mathbb{E}[X]$ 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:** Ω_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!