Econ 103: Probability and Statistics

Manu Navjeevan

UCLA

July 1, 2021

Content Outline

Single Random Variable

- Discrete and Continuous random variables
- Mean, variance, and expectations

Multiple Random Variables

- Conditional probabilities and conditional means
- Covariance and independence

The Normal Distribution

- Properties and computing probabilities

The Law of Large Numbers and the Central Limit Theorem

- The sample mean as a random variable

Random Variable

Question: What is a random variable?

Intuitively, we can think of a random variable as a model of an outcome that is uncertain.

- Example: Flipping a coin, traffic in the morning, etc.

Random Variable

Question: What is a random variable?

Intuitively, we can think of a random variable as a model of an outcome that is uncertain.

- Example: Flipping a coin, traffic in the morning, etc.

While the outcome is random, the random variable does have a *distribution*. For any subset of the outcome space, the distribution describes the probability that the random variable takes a value in that subset.

- Example: We know that our flipped coin has a 50% probability of taking a value in the set $\{H\}$, a 50% probability of taking a value in the set $\{T\}$, and a 100% probability of taking a value in the set $\{H, T\}$.

Random Variables

Question: Why do we care about random variables? What does this have to do with econometrics?

Consider the population of California. Suppose we want to know about the education levels of people in the population.

- The education level of a randomly selected person in the population is not deterministic. Different people have different levels of education.

Random Variables

Question: Why do we care about random variables? What does this have to do with econometrics?

Consider the population of California. Suppose we want to know about the education levels of people in the population.

- The education level of a randomly selected person in the population is not deterministic. Different people have different levels of education.
- We can think about the education level as a *random variable* with a distribution that describes the probability that a randomly selected person has an education level in certain range.
  - i.e 30% of people have high school diplomas, 40% of people have college degrees, etc.

Random Variables

Question: Why do we care about random variables? What does this have to do with econometrics?

Consider the population of California. Suppose we want to know about the education levels of people in the population.

- The education level of a randomly selected person in the population is not deterministic. Different people have different levels of education.
- We can think about the education level as a *random variable* with a distribution that describes the probability that a randomly selected person has an education level in certain range.
    - i.e 30% of people have high school diplomas, 40% of people have college degrees, etc.
- In general however, we may not know the exact distribution of the random variable. Econometrics is about using a random sample of data to make inferences about the underlying distribution of the random variable.

Single Random Variables: Outcome Spaces

Let's formalize the discussion above. Let X denote a random variable. All random variables come equipped with an *outcome space* that contains all the values that the random variable can take up.

.

Single Random Variables: Outcome Spaces

Let's formalize the discussion above. Let X denote a random variable. All random variables come equipped with an *outcome space* that contains all the values that the random variable can take up.

- If X represents the flipping a coin, the outcome space is $\{H, T\}$.

.

Single Random Variables: Outcome Spaces

Let's formalize the discussion above. Let X denote a random variable. All random variables come equipped with an *outcome space* that contains all the values that the random variable can take up.

- If X represents the flipping a coin, the outcome space is $\{H, T\}$.
- If X represents rolling a die, the outcome space is $\{1, 2, 3, 4, 5, 6\}$.

.

Let's formalize the discussion above. Let X denote a random variable. All random variables come equipped with an *outcome space* that contains all the values that the random variable can take up.

- If X represents the flipping a coin, the outcome space is $\{H, T\}$.
- If X represents rolling a die, the outcome space is $\{1, 2, 3, 4, 5, 6\}$.
- If X represents the 100m sprint time (in seconds) of an Olympic athlete, the outcome space may be $[9.5, 10.5]$.

.

Single Random Variables: Outcome Spaces

Let's formalize the discussion above. Let X denote a random variable. All random variables come equipped with an *outcome space* that contains all the values that the random variable can take up.

- If X represents the flipping a coin, the outcome space is $\{H, T\}$.
- If X represents rolling a die, the outcome space is $\{1, 2, 3, 4, 5, 6\}$.
- If X represents the 100m sprint time (in seconds) of an Olympic athlete, the outcome space may be $[9.5, 10.5]$.

If the outcome space of X is *countable* (think finite), then we say that X is a discrete random variable. If the outcome space of X is *uncountable* (think infinite), we say that X is a continuous random variable.

- Flipping a coin and rolling a die would be discrete random variables
- The 100m sprint time of an Olympic athlete would be a continuous random variable.

Single Random Variables: Outcome Spaces and Probability

In general in this class, we will notate the outcome space of a random variable X as $\mathcal{O}_X$. Let $2^{\mathcal{O}_X}$ denote all the subsets of $\mathcal{O}_X$.

We will typically be interested in the probability that X takes values in some $A \in 2^{\mathcal{O}_X}$ (that is $A \subseteq \mathcal{O}_X$). This probability is a number between 0 and 1 and will be notated as $\mathbb{P}_X(A)$. We will require the probability $\mathbb{P}_X(\cdot)$ to satisfy certain properties:

In general in this class, we will notate the outcome space of a random variable X as $\mathcal{O}_X$. Let $2^{\mathcal{O}_X}$ denote all the subsets of $\mathcal{O}_X$.

We will typically be interested in the probability that X takes values in some $A \in 2^{\mathcal{O}_X}$ (that is $A \subseteq \mathcal{O}_X$). This probability is a number between 0 and 1 and will be notated as $\mathbb{P}_X(A)$. We will require the probability $\mathbb{P}_X(\cdot)$ to satisfy certain properties:

- $\mathbb{P}_X(\mathcal{O}_X) = 1$ and $\mathbb{P}_X(\emptyset) = 0$.

In general in this class, we will notate the outcome space of a random variable X as $\mathcal{O}_X$. Let $2^{\mathcal{O}_X}$ denote all the subsets of $\mathcal{O}_X$.

We will typically be interested in the probability that X takes values in some $A \in 2^{\mathcal{O}_X}$ (that is $A \subseteq \mathcal{O}_X$). This probability is a number between 0 and 1 and will be notated as $\mathbb{P}_X(A)$. We will require the probability $\mathbb{P}_X(\cdot)$ to satisfy certain properties:

- $\mathbb{P}_X(\mathcal{O}_X) = 1$ and $\mathbb{P}_X(\emptyset) = 0$.
- $0 \leq \mathbb{P}_X(A) \leq 1$ for any $A \in 2^{\mathcal{O}_X}$.

In general in this class, we will notate the outcome space of a random variable X as $\mathcal{O}_X$. Let $2^{\mathcal{O}_X}$ denote all the subsets of $\mathcal{O}_X$.

We will typically be interested in the probability that X takes values in some $A \in 2^{\mathcal{O}_X}$ (that is $A \subseteq \mathcal{O}_X$). This probability is a number between 0 and 1 and will be notated as $\mathbb{P}_X(A)$. We will require the probability $\mathbb{P}_X(\cdot)$ to satisfy certain properties:

- $\mathbb{P}_X(\mathcal{O}_X) = 1$ and $\mathbb{P}_X(\emptyset) = 0$.
- $0 \leq \mathbb{P}_X(A) \leq 1$ for any $A \in 2^{\mathcal{O}_X}$.
- If $A_1, A_2, \ldots$ are pairwise disjoint, then $\mathbb{P}_X(\cup_i A_i) = \sum_i \mathbb{P}_X(A_i)$.

Single Random Variables: Outcome Spaces and Probability

In general in this class, we will notate the outcome space of a random variable X as $\mathcal{O}_X$. Let $2^{\mathcal{O}_X}$ denote all the subsets of $\mathcal{O}_X$.

We will typically be interested in the probability that X takes values in some $A \in 2^{\mathcal{O}_X}$ (that is $A \subseteq \mathcal{O}_X$). This probability is a number between 0 and 1 and will be notated as $\mathbb{P}_X(A)$. We will require the probability $\mathbb{P}_X(\cdot)$ to satisfy certain properties:

- $\mathbb{P}_X(\mathcal{O}_X) = 1$ and $\mathbb{P}_X(\emptyset) = 0$.
- $0 \leq \mathbb{P}_X(A) \leq 1$ for any $A \in 2^{\mathcal{O}_X}$.
- If $A_1, A_2, \ldots$ are pairwise disjoint, then $\mathbb{P}_X(\cup_i A_i) = \sum_i \mathbb{P}_X(A_i)$.

When we say we are interested in the *distribution* of the random variable X, we really mean we are interested in $\mathbb{P}_X(\cdot)$ as viewed as a map from $2^{\mathcal{O}_X}$ onto $[0, 1]$.

Single Random Variables: Discrete Random Variables

If X is a discrete random variable the distribution or probability function $\mathbb{P}_X$ can be described by the *probability mass function* or *pmf*, $p_X(\cdot) : \mathcal{O}_X \to [0, 1]$.

- Recall that a discrete random variable has a countable (think finite) outcome space

.

Single Random Variables: Discrete Random Variables

If X is a discrete random variable the distribution or probability function $\mathbb{P}_X$ can be described by the *probability mass function* or *pmf*, $p_X(\cdot) : \mathcal{O}_X \to [0, 1]$.

- Recall that a discrete random variable has a countable (think finite) outcome space

For each element a of the outcome space ($a \in \mathcal{O}_X$), the probability mass function evaluated at a, $p_X(a)$, describes the probability that X takes value a. That is $p_X(a) = \mathbb{P}_X(\{a\})$.

Single Random Variables: Discrete Random Variables

If X is a discrete random variable the distribution or probability function $\mathbb{P}_X$ can be described by the *probability mass function* or *pmf*, $p_X(\cdot) : \mathcal{O}_X \to [0, 1]$.

- Recall that a discrete random variable has a countable (think finite) outcome space

For each element a of the outcome space ($a \in \mathcal{O}_X$), the probability mass function evaluated at a, $p_X(a)$, describes the probability that X takes value a. That is $p_X(a) = \mathbb{P}_X(\{a\})$.

By the last property of probability measures, the pmf can be used to recover the probability that X takes values in any subset A of the outcomes space $\mathcal{O}_X$

$$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(\{a\}) = \sum_{a \in A} p_X(a).$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

$$\mathbb{P}_X(\{2,4,6\}) = \sum_{a \in \{2,4,6\}} \mathbb{P}_X(\{a\})$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

$$\begin{aligned} \mathbb{P}_X(\{2,4,6\}) &= \sum_{a \in \{2,4,6\}} \mathbb{P}_X(\{a\}) \\ &= \sum_{a \in \{2,4,6\}} p_X(a) \end{aligned}$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

$$\begin{aligned} \mathbb{P}_X(\{2,4,6\}) &= \sum_{a \in \{2,4,6\}} \mathbb{P}_X(\{a\}) \\ &= \sum_{a \in \{2,4,6\}} p_X(a) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \end{aligned}$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

$$\begin{aligned} \mathbb{P}_X(\{2,4,6\}) &= \sum_{a \in \{2,4,6\}} \mathbb{P}_X(\{a\}) \\ &= \sum_{a \in \{2,4,6\}} p_X(a) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= 1/2 \end{aligned}$$

Single Random Variables: Discrete Random Variables

Let's see an example of this. Let X denote the outcome of a fair dice roll. We can describe the distribution of X via the probability mass function

$$p_X(a) = \begin{cases} \frac{1}{6} & \text{if } a \in \{1,2,3,4,5,6\} \\ 0 & \text{for any other value of } a \end{cases}$$

Let's use the pmf to compute $\mathbb{P}_X(A)$ for $A = \{2,4,6\}$, that is use the pmf to compute the probability that X takes on an even value.

$$\begin{aligned} \mathbb{P}_X(\{2,4,6\}) &= \sum_{a \in \{2,4,6\}} \mathbb{P}_X(\{a\}) \\ &= \sum_{a \in \{2,4,6\}} p_X(a) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= 1/2 \end{aligned}$$

Of course, this result is a bit obvious. However, if the die was not fair, we would follow the same procedure to compute this probability.

Single Random Variables: Continuous Random Variables

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

Single Random Variables: Continuous Random Variables

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

- Recall that if X is a continuous random variable then the outcome space $\mathcal{O}_X$ is (uncountably) infinite.
- What would happen if $\mathbb{P}_X(\{a\}) > 0$ for each $a \in X$?

Single Random Variables: Continuous Random Variables

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

- Recall that if X is a continuous random variable then the outcome space $\mathcal{O}_X$ is (uncountably) infinite.
- What would happen if $\mathbb{P}_X(\{a\}) > 0$ for each $a \in X$?
    - Then for any set A

$$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(\{a\}) = \infty.$$

    - So we must have $\mathbb{P}_X(\{a\}) = 0$ for all $a \in \mathcal{O}_X$.

Single Random Variables: Continuous Random Variables

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

- Recall that if X is a continuous random variable then the outcome space $\mathcal{O}_X$ is (uncountably) infinite.
- What would happen if $\mathbb{P}_X(\{a\}) > 0$ for each $a \in X$?
  - Then for any set A
  $$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(\{a\}) = \infty.$$
  - So we must have $\mathbb{P}_X(\{a\}) = 0$ for all $a \in \mathcal{O}_X$.
- Intuitively, what is the probability that an Olympic sprinters runs the 100m dash in exactly 9.8412312. . . seconds?
  - Basically zero.

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

- Recall that if X is a continuous random variable then the outcome space $\mathcal{O}_X$ is (uncountably) infinite.
- What would happen if $\mathbb{P}_X(\{a\}) > 0$ for each $a \in X$?
  - Then for any set A
  $$\mathbb{P}_X(A) = \sum_{a \in A} \mathbb{P}_X(\{a\}) = \infty.$$
  - So we must have $\mathbb{P}_X(\{a\}) = 0$ for all $a \in \mathcal{O}_X$.
- Intuitively, what is the probability that an Olympic sprinters runs the 100m dash in exactly 9.8412312. . . seconds?
  - Basically zero.

This rules out being able to use a pmf to describe the distribution of a continuous random variable.

Single Random Variables: Continuous Random Variables.

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

Single Random Variables: Continuous Random Variables.

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

Instead we will use the probability density function (pdf), $f_X(\cdot)$ to describe the distribution of X. The pdf is related to the probability measure $\mathbb{P}_X$ via the following equation:

$$\mathbb{P}_X(a \leq X \leq b) = \int_a^b f_X(x)\, dx.$$

Single Random Variables: Continuous Random Variables.

If X is a continuous random variable we cannot use a probability mass function to describe its distribution.

Instead we will use the probability density function (pdf), $f_X(\cdot)$ to describe the distribution of X. The pdf is related to the probability measure $\mathbb{P}_X$ via the following equation:

$$\mathbb{P}_X(a \leq X \leq b) = \int_a^b f_X(x)\, dx.$$

The identity above as well as the rules for the probability measure $\mathbb{P}_X$ can be used to calculate $\mathbb{P}_X(A)$ for any set $A \subseteq \mathcal{O}_X$.

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0, 1]$.

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0, 1]$.

Problem: What is $\mathbb{P}_X([0, 0.5])$?

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0, 1]$.

Problem: What is $\mathbb{P}_X([0, 0.5])$?

Answer: Use the pdf

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0, 1]$.

Problem: What is $\mathbb{P}_X([0, 0.5])$?

Answer: Use the pdf

$$\mathbb{P}_X([0, 0.5]) = \int_0^{0.5} f_X(x)\, dx$$

Single Random Variables: Continuous Random Variables

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0, 1]$.

Problem: What is $\mathbb{P}_X([0, 0.5])$?

Answer: Use the pdf

$$\begin{aligned} \mathbb{P}_X([0, 0.5]) &= \int_0^{0.5} f_X(x)\,dx \\ &= \int_0^{0.5} 1\,dx \end{aligned}$$

Single Random Variables: Continuous Random Variables

Example: Let X be a continuous random variable with pdf f_X given

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is called the *uniform distribution* on $[0,1]$.

Problem: What is $\mathbb{P}_X([0,0.5])$?

Answer: Use the pdf

$$\begin{aligned} \mathbb{P}_X([0,0.5]) &= \int_0^{0.5} f_X(x)\,dx \\ &= \int_0^{0.5} 1\,dx \\ &= 0.5 - 0 = 0.5 \end{aligned}$$

Single Random Variables: Review

Let's recap:

Random Variables:

- Describe an event whose outcome is unknown
- Instead we have probabilities that the outcome takes values in a certain set.

Single Random Variables: Review

Let's recap:

Random Variables:

- Describe an event whose outcome is unknown
- Instead we have probabilities that the outcome takes values in a certain set.

Discrete Random Variable:

- Random Variable whose outcome space is countable (think finite)
- Distribution/Probability measure completely described by pmf.
  - $p_X(a) = \mathbb{P}_X(\{a\})$

Single Random Variables: Review

Let's recap:

Random Variables:

- Describe an event whose outcome is unknown
- Instead we have probabilities that the outcome takes values in a certain set.

Discrete Random Variable:

- Random Variable whose outcome space is countable (think finite)
- Distribution/Probability measure completely described by pmf.
    - $p_X(a) = \mathbb{P}_X(\{a\})$

Continuous Random Variables:

- Random variable whose outcome space is uncountable (think infinite)
- Distribution/Probability measure completely described by pdf
    - $\mathbb{P}_X([a,b]) = \int_a^b f_X(x)\,dx$

Questions?

Single Random Variables: Expectation

One property of a random variable that we may be interested in is the expectation of a random variable.

- The expectation can be thought of as a generalized concept of a mean.
- That is it is the "average" value the random variable X will take, weighted by the probability that X takes each value

Single Random Variables: Expectation

One property of a random variable that we may be interested in is the expectation of a random variable.

- The expectation can be thought of as a generalized concept of a mean.
- That is it is the "average" value the random variable X will take, weighted by the probability that X takes each value

The expectation of X is denoted $\mathbb{E}[X]$ or μ_X and is calculated via the following

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} a \cdot p_X(a)$	$\int_{\mathcal{O}_X} a \cdot f_X(a)\, da$

Single Random Variables: Expectation

One property of a random variable that we may be interested in is the expectation of a random variable.

- The expectation can be thought of as a generalized concept of a mean.
- That is it is the "average" value the random variable X will take, weighted by the probability that X takes each value

The expectation of X is denoted $\mathbb{E}[X]$ or μ_X and is calculated via the following

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} a \cdot p_X(a)$	$\int_{\mathcal{O}_X} a \cdot f_X(a)\, da$

- Note that the difference between discrete and continuous is just summation vs. integral.

Single Random Variables: Expectation

Suppose we enter a lottery that pays us $\$100$ with probability $1/2$, $\$400$ with probability $1/4$, and $\$0$ with probability $1/4$. We can represent the payout of this lottery (in dollars) with a random variable X whose pmf is given

$$p_X(a) = \begin{cases} 1/2 & \text{if } a \in \{100\} \\ 1/4 & \text{if } a \in \{0, 400\} \\ 0 & \text{otherwise} \end{cases}.$$

Single Random Variables: Expectation

Suppose we enter a lottery that pays us $\$100$ with probability $1/2$, $\$400$ with probability $1/4$, and $\$0$ with probability $1/4$. We can represent the payout of this lottery (in dollars) with a random variable X whose pmf is given

$$p_X(a) = \begin{cases} 1/2 & \text{if } a \in \{100\} \\ 1/4 & \text{if } a \in \{0, 400\} \\ 0 & \text{otherwise} \end{cases}.$$

Let's calculate the *expected value* of this lottery, or $\mathbb{E}[X]$. This is the average amount we can expect to win by playing this lottery.

Single Random Variables: Expectation

Suppose we enter a lottery that pays us $\$100$ with probability $1/2$, $\$400$ with probability $1/4$, and $\$0$ with probability $1/4$. We can represent the payout of this lottery (in dollars) with a random variable X whose pmf is given

$$p_X(a) = \begin{cases} 1/2 & \text{if } a \in \{100\} \\ 1/4 & \text{if } a \in \{0, 400\} \\ 0 & \text{otherwise} \end{cases}.$$

Let's calculate the *expected value* of this lottery, or $\mathbb{E}[X]$. This is the average amount we can expect to win by playing this lottery. Using the formula from the slide above:

$$\mathbb{E}[X] = \sum_{a \in \{0,100,400\}} a \cdot p_X(a)$$

Single Random Variables: Expectation

Suppose we enter a lottery that pays us $\$100$ with probability $1/2$, $\$400$ with probability $1/4$, and $\$0$ with probability $1/4$. We can represent the payout of this lottery (in dollars) with a random variable X whose pmf is given

$$p_X(a) = \begin{cases} 1/2 & \text{if } a \in \{100\} \\ 1/4 & \text{if } a \in \{0, 400\} \\ 0 & \text{otherwise} \end{cases} .$$

Let's calculate the *expected value* of this lottery, or $\mathbb{E}[X]$. This is the average amount we can expect to win by playing this lottery. Using the formula from the slide above:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{a \in \{0,100,400\}} a \cdot p_X(a) \\ &= 0 \cdot \frac{1}{4} + 100 \cdot \frac{1}{2} + 400 \cdot \frac{1}{4} \end{aligned}$$

Single Random Variables: Expectation

Suppose we enter a lottery that pays us $\$100$ with probability $1/2$, $\$400$ with probability $1/4$, and $\$0$ with probability $1/4$. We can represent the payout of this lottery (in dollars) with a random variable X whose pmf is given

$$p_X(a) = \begin{cases} 1/2 & \text{if } a \in \{100\} \\ 1/4 & \text{if } a \in \{0, 400\} \\ 0 & \text{otherwise} \end{cases}.$$

Let's calculate the *expected value* of this lottery, or $\mathbb{E}[X]$. This is the average amount we can expect to win by playing this lottery. Using the formula from the slide above:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{a \in \{0,100,400\}} a \cdot p_X(a) \\ &= 0 \cdot \frac{1}{4} + 100 \cdot \frac{1}{2} + 400 \cdot \frac{1}{4} \\ &= 50 + 100 = 150 \end{aligned}$$

We can expect to win about $150 by playing this lottery

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0, 1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0, 1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X]$. Using the formula

$$\mathbb{E}[X] = \int_{\mathcal{O}_X} a \cdot f_X(a)\, da$$

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0, 1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X]$. Using the formula

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathcal{O}_X} a \cdot f_X(a)\, da \\ &= \int_0^1 a \cdot 1\, da \end{aligned}$$

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0, 1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X]$. Using the formula

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathcal{O}_X} a \cdot f_X(a)\, da \\ &= \int_0^1 a \cdot 1\, da \\ &= \frac{a^2}{2}\bigg|_0^1 \end{aligned}$$

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0,1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X]$. Using the formula

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathcal{O}_X} a \cdot f_X(a)\, da \\ &= \int_0^1 a \cdot 1\, da \\ &= \left.\frac{a^2}{2}\right|_0^1 \\ &= \frac{1}{2} - 0 = \frac{1}{2} \end{aligned}$$

Single Random Variables: Expectation

Let's consider the uniform distribution from before. That is, let X be continuously distributed with outcome space $\mathcal{O}_X = [0,1]$ and pdf given by

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X]$. Using the formula

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathcal{O}_X} a \cdot f_X(a)\, da \\ &= \int_0^1 a \cdot 1\, da \\ &= \left.\frac{a^2}{2}\right|_0^1 \\ &= \frac{1}{2} - 0 = \frac{1}{2} \end{aligned}$$

So the expected value of X is $\frac{1}{2}$. There are many ways to interpret this expected value, the most straightforward for me is the average amount you can expect to win from a lottery whose payouts follow the distribution of X.

Single Random Variables: Expectation

We can generalize this concept a bit further and consider the mean of any function $g(X)$, which is typically denoted $\mathbb{E}[g(X)]$.

- Note that $g(X)$ itself is a random variable. It's outcome is not deterministic but rather follows a distribution based on X.
- As with the normal expectation, we can think about this as the average value of $g(X)$.

Single Random Variables: Expectation

We can generalize this concept a bit further and consider the mean of any function $g(X)$, which is typically denoted $\mathbb{E}[g(X)]$.

- Note that $g(X)$ itself is a random variable. It's outcome is not deterministic but rather follows a distribution based on X.
- As with the normal expectation, we can think about this as the average value of $g(X)$.

The formula for calculating $\mathbb{E}[g(X)]$ is basically the same as for calculating $\mathbb{E}[X]$.

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} g(a) \cdot p_X(a)$	$\int_{\mathcal{O}_X} g(a) \cdot f_X(a)\, da$

Single Random Variables: Expectation

We can generalize this concept a bit further and consider the mean of any function $g(X)$, which is typically denoted $\mathbb{E}[g(X)]$.

- Note that $g(X)$ itself is a random variable. It's outcome is not deterministic but rather follows a distribution based on X.
- As with the normal expectation, we can think about this as the average value of $g(X)$.

The formula for calculating $\mathbb{E}[g(X)]$ is basically the same as for calculating $\mathbb{E}[X]$.

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} g(a) \cdot p_X(a)$	$\int_{\mathcal{O}_X} g(a) \cdot f_X(a)\, da$

- All that has changed here is that we are multiplying the pmf/pdf by $g(a)$ instead of by a.

Single Random Variables: Expectation

The formulas from this generalization also gives us a nice property that we will use later. Recall from the last slide we can calculate $\mathbb{E}[g(X)]$ using:

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} g(a) \cdot p_X(a)$	$\int_{\mathcal{O}_X} g(a) \cdot f_X(a)\, da$

Single Random Variables: Expectation

The formulas from this generalization also gives us a nice property that we will use later. Recall from the last slide we can calculate $\mathbb{E}[g(X)]$ using:

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a \in \mathcal{O}_X} g(a) \cdot p_X(a)$	$\int_{\mathcal{O}_X} g(a) \cdot f_X(a)\, da$

It is then straightforward to see the following:

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$

for any $a, b \in \mathbb{R}$. We will refer to this property as the linearity of the expectation. Later on when we consider multiple or joint random variables, we will see that this can be naturally extended and for two different random variables X and Y:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X^2]$. This is like calculating $\mathbb{E}[g(X)]$ where $g(a) = a^2$. Using the formula from before

$$\mathbb{E}[X^2] = \int_{\mathcal{O}_X} a^2 \cdot f_X(a)\, da$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X^2]$. This is like calculating $\mathbb{E}[g(X)]$ where $g(a) = a^2$. Using the formula from before

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{\mathcal{O}_X} a^2 \cdot f_X(a)\, da \\ &= \int_0^1 a^2 \cdot 1\, da \end{aligned}$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X^2]$. This is like calculating $\mathbb{E}[g(X)]$ where $g(a) = a^2$. Using the formula from before

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{\mathcal{O}_X} a^2 \cdot f_X(a)\, da \\ &= \int_0^1 a^2 \cdot 1\, da \\ &= \frac{a^3}{3}\bigg|_0^1 \end{aligned}$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X^2]$. This is like calculating $\mathbb{E}[g(X)]$ where $g(a) = a^2$. Using the formula from before

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{\mathcal{O}_X} a^2 \cdot f_X(a)\, da \\ &= \int_0^1 a^2 \cdot 1\, da \\ &= \frac{a^3}{3}\bigg|_0^1 \\ &= \frac{1}{3} - 0 = \frac{1}{3} \end{aligned}$$

Single Random Variables: Expectation

Let's return to the uniform distribution from before. Suppose that X follows the pdf

$$f_X(a) = \begin{cases} 1 & \text{if } 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose we want to calculate $\mathbb{E}[X^2]$. This is like calculating $\mathbb{E}[g(X)]$ where $g(a) = a^2$. Using the formula from before

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{\mathcal{O}_X} a^2 \cdot f_X(a)\, da \\ &= \int_0^1 a^2 \cdot 1\, da \\ &= \frac{a^3}{3}\bigg|_0^1 \\ &= \frac{1}{3} - 0 = \frac{1}{3} \end{aligned}$$

For an arbitrary random variable X, $\mathbb{E}[X^2]$ is typically referred to as the *second moment* of X.

Single Random Variables: Variance

Now that we can compute $\mathbb{E}[g(X)]$ for any random variable X and any function $g(\cdot)$, we are ready to talk about the variance.

The variance of a random variable X is given by the formula

$$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

The variance is a measure of the "spread" of the random variable X; it represents how far on average X is from its mean. Using linearity of the expectation the expression above can be simplified:

Single Random Variables: Variance

Now that we can compute $\mathbb{E}[g(X)]$ for any random variable X and any function $g(\cdot)$, we are ready to talk about the variance.

The variance of a random variable X is given by the formula

$$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

The variance is a measure of the "spread" of the random variable X; it represents how far on average X is from its mean. Using linearity of the expectation the expression above can be simplified:

$$\begin{aligned}
&= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Single Random Variables: Variance

Now that we can compute $\mathbb{E}[g(X)]$ for any random variable X and any function $g(\cdot)$, we are ready to talk about the variance.

The variance of a random variable X is given by the formula

$$\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

The variance is a measure of the "spread" of the random variable X; it represents how far on average X is from its mean. Using linearity of the expectation the expression above can be simplified:

$$\begin{aligned}
&= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Typically this last expression is easiest to work with. However, the first expression gives us an important property: the variance is always ≥ 0.

Single Random Variables: Variance

From the formula $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ as well as the linearity of the expectation we can see the variance has some nice properties. For any constants $a, b \in \mathbb{R}$ and any random variable X we get that

- $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
- $\mathrm{Var}(aX) = a^2 \,\mathrm{Var}(X)$

Single Random Variables: Variance

From the formula $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ as well as the linearity of the expectation we can see the variance has some nice properties. For any constants $a, b \in \mathbb{R}$ and any random variable X we get that

- $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
- $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$

Putting these together gives us: $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.

Exercise: See if you can work this out yourselves. It shouldn't take more than a few minutes.

Single Random Variables: Variance

Intuitively, $\mathrm{Var}(X + a) = \mathrm{Var}(X)$ means that the spread around the mean is not effected by just shifting the mean of the random variable (while otherwise keeping the distribution the same).

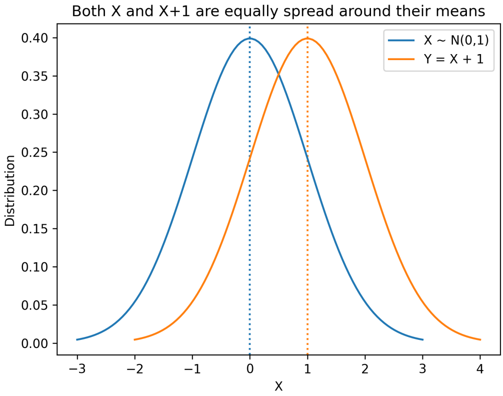

Figure 1: Variance remains the same for both variables

Single Random Variables: Variance

The square root of the variance is denoted

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathrm{Var}(X)}.$$

and is called the standard deviation of X.

Single Random Variables: Variance

Let's see this in practice. Suppose we have a lottery that pays \$200 with probability $\frac{1}{2}$ and nothing (\$0) with probability $\frac{1}{2}$. The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

It is straightforward to see that the expected payout of this lottery is \$100, $\mathbb{E}[X] = 100$.

Let's see this in practice. Suppose we have a lottery that pays \$200 with probability $\frac{1}{2}$ and nothing (\$0) with probability $\frac{1}{2}$. The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

It is straightforward to see that the expected payout of this lottery is \$100, $\mathbb{E}[X] = 100$. However, we may want to know how much we can expect our winnings to deviate from the expected value. That is, we want to know what the variance of the payouts is.

Let's compute this two ways, first using the formula $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ and the second using $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

First:

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

First:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - 100)^2] \end{aligned}$$

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

First:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - 100)^2] \\ &= \sum_{a \in \{0,200\}} (a - 100)^2 \cdot p_X(a) \end{aligned}$$

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

First:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[(X - 100)^2] \\
&= \sum_{a \in \{0,200\}} (a - 100)^2 \cdot p_X(a) \\
&= (0 - 100)^2 \cdot \frac{1}{2} + (200 - 100)^2 \cdot \frac{1}{2}
\end{aligned}$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

First:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[(X - 100)^2] \\
&= \sum_{a \in \{0,200\}} (a - 100)^2 \cdot p_X(a) \\
&= (0 - 100)^2 \cdot \frac{1}{2} + (200 - 100)^2 \cdot \frac{1}{2} \\
&= 2 \cdot \frac{100^2}{2} = 100^2 = 10000
\end{aligned}$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

Second:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

Second:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \sum_{a \in \{0,200\}} a^2 \cdot p_X(a) - 100^2 \end{aligned}$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

Second:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \sum_{a \in \{0,200\}} a^2 \cdot p_X(a) - 100^2 \\ &= 0^2 \cdot \frac{1}{2} + 200^2 \cdot \frac{1}{2} - 100^2 \end{aligned}$$

Single Random Variables: Variance

The payout of this lottery is a random variable X with pmf

$$p_X(a) = \begin{cases} \frac{1}{2} & \text{if } a \in \{0, 200\} \\ 0 & \text{otherwise} \end{cases}.$$

Second:

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \sum_{a \in \{0,200\}} a^2 \cdot p_X(a) - 100^2 \\ &= 0^2 \cdot \frac{1}{2} + 200^2 \cdot \frac{1}{2} - 100^2 \\ &= \frac{(2 \cdot 100)^2}{2} - 100^2 \\ &= \frac{4}{2} \cdot 100^2 - 100^2 = 100^2 = 10000 \end{aligned}$$

No matter how we compute it, the variance of this lottery is $100^2 = 10000$. This makes sense as no matter what happens (win or lose), we are \$100 away from the expected payout.

Single Random Variables: Questions?

Questions?

Multiple Random Variables: Introduction

Oftentimes we are not only interested in a single random variable, but the relationship between two random variables. In fact this will be the case through the rest of this course:

- We care about the relationship between education and income
- We care about the relationship between consumption of a medicine and a health outcome

Multiple Random Variables: Introduction

Oftentimes we are not only interested in a single random variable, but the relationship between two random variables. In fact this will be the case through the rest of this course:

- We care about the relationship between education and income
- We care about the relationship between consumption of a medicine and a health outcome

In the examples above notice that, not only are the random variables themselves not deterministic

- not everyone has the same education/income

but the relationship between the two random variables may not be either

- not everyone who takes a medicine will have the same health outcome

Multiple Random Variables: Introduction

Question: How do we think about joint random variables?

Multiple Random Variables: Introduction

Question: How do we think about joint random variables?

Answer: Much the same as before. Let (X, Y) be a pair of joint random variables. This means there is some outcome space $\mathcal{O}_{XY}$ that (X, Y) can take values in and a probability measure $\mathbb{P}_{XY}(\cdot) : 2^{\mathcal{O}_{XY}} \to [0, 1]$ that takes in subsets of the outcome space $A \subseteq \mathcal{O}_{XY}$ and gives the probability of both X and Y taking values in the set A.

- For example if X is income and Y is age,

$$\mathbb{P}_{XY}(\{0 \leq X \leq 100000, 40 \leq Y \leq 42\})$$

  is the probability that a randomly selected person from the population has an income between \$0 and \$100,000 and is between 40 and 42 years old.

Multiple Random Variables: Introduction

Question: How do we think about joint random variables?

Answer: Much the same as before. Let (X, Y) be a pair of joint random variables. This means there is some outcome space $\mathcal{O}_{XY}$ that (X, Y) can take values in and a probability measure $\mathbb{P}_{XY}(\cdot) : 2^{\mathcal{O}_{XY}} \to [0, 1]$ that takes in subsets of the outcome space $A \subseteq \mathcal{O}_{XY}$ and gives the probability of both X and Y taking values in the set A.

- For example if X is income and Y is age,

$$\mathbb{P}_{XY}(\{0 \leq X \leq 100000, 40 \leq Y \leq 42\})$$

  is the probability that a randomly selected person from the population has an income between \$0 and \$100,000 and is between 40 and 42 years old.

As before if $\mathcal{O}_{XY}$ is countable (finite), we say that (X, Y) are jointly discrete random variables whereas if $\mathcal{O}_{XY}$ is uncountable (infinite) we say that (X, Y) are jointly continuous random variables.

Multiple Random Variables: Discrete Random Variables

Let's quickly go over an example of a joint discrete random variables. As before, the distribution of jointly discrete random variables will be defined by a joint probability mass function.

Let's quickly go over an example of a joint discrete random variables. As before, the distribution of jointly discrete random variables will be defined by a joint probability mass function.

- Because we are considering two random variables, X, Y, we can represent the probability mass function as a table

Let X be a random variable that describes whether a person gets 4 hours of sleep a night, 8 hours a sleep a night, or 12 hours of sleep a night. Let Y be a random variable that describes whether a person drinks 1 or 2 cups of coffee a day.

The joint pmf of X and Y can be described with the table below

$p(x,y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Using this table we can see that the probability that a randomly selected person gets 8 hours of sleep and drinks 1 cup of coffee a day is $1/3$.

Exercise: What is the probability that a randomly selected person gets 8 hours of sleep?

Multiple Random Variables: Continuous Random Variables

Now we'll go over an example of a joint continuously distributed random variable. Just like with single random variables, the distribution of a continuous random variable is defined by a probability density function, $f_{XY}(x, y)$.

Now we'll go over an example of a joint continuously distributed random variable. Just like with single random variables, the distribution of a continuous random variable is defined by a probability density function, $f_{XY}(x, y)$.

- As before, the joint pdf will be related to the joint probability measure $\mathbb{P}_{XY}(\cdot)$ via the relation

$$\mathbb{P}_{XY}(\{a \leq X \leq b, c \leq Y \leq d\}) = \int_a^b \int_c^d f_{XY}(x, y)\, dy\, dx.$$

Multiple Random Variables: Continuous Random Variables

Let's consider two sprinters in the 100m dash. Let X denote the finish time of the UCLA sprinter and Y denote the finish time of the USC sprinter. Suppose their times follow the following joint pdf

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's consider two sprinters in the 100m dash. Let X denote the finish time of the UCLA sprinter and Y denote the finish time of the USC sprinter. Suppose their times follow the following joint pdf

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try to find the probability that the UCLA sprinter runs faster than 10 seconds and that the USC sprinter runs faster than 10.5 seconds. That is we want $\mathbb{P}_{XY}(\{X \leq 10, Y \leq 10.5\})$.

Multiple Random Variables: Continuous Random Variables

Let's consider two sprinters in the 100m dash. Let X denote the finish time of the UCLA sprinter and Y denote the finish time of the USC sprinter. Suppose their times follow the following joint pdf

$$f_{XY}(x, y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try to find the probability that the UCLA sprinter runs faster than 10 seconds and that the USC sprinter runs faster than 10.5 seconds. That is we want $\mathbb{P}_{XY}(\{X \leq 10, Y \leq 10.5\})$.

$$\begin{aligned} \mathbb{P}_{XY}(\{X \leq 10, Y \leq 10.5\}) &= \int_{9.5}^{10} \int_{10}^{10.5} f_{XY}(x, y)\, dy\, dx \\ &= \int_{9.5}^{10} \int_{10}^{10.5} 1\, dy\, dx \\ &= \int_{9.5}^{10} 0.5\, dx \\ &= 0.5 \cdot 0.5 = 0.25 \end{aligned}$$

Just as before we may want to consider the average or expected value that some function, $g(X, Y)$, of our joint random variables may take. That is, we want to calculate $\mathbb{E}[g(X, Y)]$.

Multiple Random Variables: Expectations

Just as before we may want to consider the average or expected value that some function, $g(X, Y)$, of our joint random variables may take. That is, we want to calculate $\mathbb{E}[g(X, Y)]$.

Here are some examples of functions $g(x, y)$ for which we may be interested in $\mathbb{E}[g(X, Y)]$:

- $g(x, y) = x \implies \mathbb{E}[g(X, Y)] = \mathbb{E}[X]$, the expected value of X
- $g(x, y) = x - y \implies \mathbb{E}[g(X, Y)] = \mathbb{E}[X - Y]$, the average difference between X and Y.
- $g(x, y) = \mathbb{1}\{x \leq a, y \leq b\} \implies \mathbb{E}[g(X, Y)] = \mathbb{P}_{XY}(\{X \leq a, Y \leq b\})$.
- $g(x, y) = (x - \mu_X)(y - \mu_Y) \implies \mathbb{E}[g(X, Y)] = \mathrm{Cov}(X, Y)$, the covariance between X and Y.

Multiple Random Variables: Expectations

The formula for calculating expected value is the same as before:

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a,b \in \mathcal{O}_{XY}} g(a,b) p_{XY}(a,b)$	$\int_X \int_Y g(a,b) f_{XY}(a,b)\, db\, da$

Intuition: We are evaluating the function at each point in the outcome space and weighting by the probability of that outcome occurring.

Multiple Random Variables: Expectations

The formula for calculating expected value is the same as before:

Type	Discrete R.V	Continuous R.V
Formula	$\sum_{a,b\in\mathcal{O}_{XY}} g(a,b)p_{XY}(a,b)$	$\int_X \int_Y g(a,b) f_{XY}(a,b)\, db\, da$

Intuition: We are evaluating the function at each point in the outcome space and weighting by the probability of that outcome occurring.

Again note that we have the following linearity of the expectation. For any two functions $g(x,y)$ and $h(x,y)$ and any $a,b\in\mathbb{R}$:

$$\mathbb{E}[a\cdot g(X,Y)+b\cdot h(X,Y)] = a\mathbb{E}[g(X,Y)]+b\mathbb{E}[h(X,Y)].$$

in particular, $\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$.

Multiple Random Variables: Expectations

Let's return to the 100m dash example from before, with X representing the finishing time of the UCLA and Y representing the finishing time of the USC sprinter.

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try and calculate $\mathbb{E}[X - Y]$, the expected difference in finishing times between the UCLA and USC sprinters.

Multiple Random Variables: Expectations

Let's return to the 100m dash example from before, with X representing the finishing time of the UCLA and Y representing the finishing time of the USC sprinter.

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \le x \le 10.5, 10 \le y \le 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try and calculate $\mathbb{E}[X - Y]$, the expected difference in finishing times between the UCLA and USC sprinters.

$$\mathbb{E}[X - Y] = \int_{9.5}^{10.5} \int_{10}^{11} (x - y) f_{XY}(x,y)\, dy\, dx$$

Multiple Random Variables: Expectations

Let's return to the 100m dash example from before, with X representing the finishing time of the UCLA and Y representing the finishing time of the USC sprinter.

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try and calculate $\mathbb{E}[X - Y]$, the expected difference in finishing times between the UCLA and USC sprinters.

$$\begin{aligned} \mathbb{E}[X - Y] &= \int_{9.5}^{10.5} \int_{10}^{11} (x - y) f_{XY}(x,y)\, dy\, dx \\ &= \int_{9.5}^{10.5} \int_{10}^{11} x\, dy\, dx - \int_{9.5}^{10.5} \int_{10}^{11} y\, dy\, dx \end{aligned}$$

Multiple Random Variables: Expectations

Let's return to the 100m dash example from before, with X representing the finishing time of the UCLA and Y representing the finishing time of the USC sprinter.

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try and calculate $\mathbb{E}[X-Y]$, the expected difference in finishing times between the UCLA and USC sprinters.

$$\begin{aligned}
\mathbb{E}[X-Y] &= \int_{9.5}^{10.5} \int_{10}^{11} (x-y) f_{XY}(x,y)\, dy\, dx \\
&= \int_{9.5}^{10.5} \int_{10}^{11} x\, dy\, dx - \int_{9.5}^{10.5} \int_{10}^{11} y\, dy\, dx \\
&= \int_{9.5}^{10.5} x\, dx - \int_{9.5}^{10.5} \frac{11^2 - 10^2}{2}\, dx
\end{aligned}$$

Let's return to the 100m dash example from before, with X representing the finishing time of the UCLA and Y representing the finishing time of the USC sprinter.

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

Let's try and calculate $\mathbb{E}[X-Y]$, the expected difference in finishing times between the UCLA and USC sprinters.

$$\begin{aligned}
\mathbb{E}[X-Y] &= \int_{9.5}^{10.5}\int_{10}^{11}(x-y)f_{XY}(x,y)\,dy\,dx \\
&= \int_{9.5}^{10.5}\int_{10}^{11} x\,dy\,dx - \int_{9.5}^{10.5}\int_{10}^{11} y\,dy\,dx \\
&= \int_{9.5}^{10.5} x\,dx - \int_{9.5}^{10.5} \frac{11^2-10^2}{2}\,dx \\
&= \frac{10.5^2-9.5^2}{2} - 10.5 = -0.5
\end{aligned}$$

Multiple Random Variables: Covariance

As mentioned before, a particular expectation we may be interested is the covariance between X and Y, defined as

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

The covariance measures how much the variables X and Y "move together" and will be of particular interest to us in econometrics.

Multiple Random Variables: Covariance

As mentioned before, a particular expectation we may be interested is the covariance between X and Y, defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

The covariance measures how much the variables X and Y "move together" and will be of particular interest to us in econometrics.

As before, we can simplify the expression for covariance:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x, y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

$$\mathbb{E}[XY] = \int_{9.5}^{10.5} \int_{10}^{11} xy f_{XY}(x,y)\, dy\, dx$$

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{9.5}^{10.5} \int_{10}^{11} xy f_{XY}(x,y)\, dy\, dx \\ &= \int_{9.5}^{10.5} x \int_{10}^{11} y\, dy\, dx \end{aligned}$$

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{9.5}^{10.5} \int_{10}^{11} xy f_{XY}(x,y)\, dy\, dx \\ &= \int_{9.5}^{10.5} x \int_{10}^{11} y\, dy\, dx \\ &= \int_{9.5}^{10.5} x \left(\frac{11^2 - 10^2}{2} \right) dx \end{aligned}$$

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \leq x \leq 10.5, 10 \leq y \leq 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{9.5}^{10.5} \int_{10}^{11} xy f_{XY}(x,y)\, dy\, dx \\ &= \int_{9.5}^{10.5} x \int_{10}^{11} y\, dy\, dx \\ &= \int_{9.5}^{10.5} x \left(\frac{11^2 - 10^2}{2} \right) dx \\ &= 10.5 \left(\frac{10.5^2 - 9.5^2}{2} \right) = 10.5 \cdot 10 = 105 \end{aligned}$$

Multiple Random Variables: Covariance

Let's calculate the covariance between X and Y from the example before:

$$f_{XY}(x,y) = \begin{cases} 1 & \text{if } 9.5 \le x \le 10.5, 10 \le y \le 11 \\ 0 & \text{otherwise} \end{cases}.$$

From the last exercise we know that $\mathbb{E}[X] = 10$ and $\mathbb{E}[Y] = 10.5$. What remains is to calculate $\mathbb{E}[XY]$:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{9.5}^{10.5} \int_{10}^{11} xy f_{XY}(x,y)\, dy\, dx \\ &= \int_{9.5}^{10.5} x \int_{10}^{11} y\, dy\, dx \\ &= \int_{9.5}^{10.5} x \left(\frac{11^2 - 10^2}{2} \right) dx \\ &= 10.5 \left(\frac{10.5^2 - 9.5^2}{2} \right) = 10.5 \cdot 10 = 105 \end{aligned}$$

So $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 105 - 105 = 0$. At least to a first order, there is no association between the UCLA sprinters times and the USC sprinters times.

Multiple Random Variables: Correlation

Another quantity of interest here is the correlation coefficient defined as

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

where we recall that

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$
$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\mathrm{Var}(Y)} = \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}$$

Multiple Random Variables: Correlation

Another quantity of interest here is the correlation coefficient defined as

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

where we recall that

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$$
$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{\mathrm{Var}(Y)} = \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]}$$

Calculating the correlation is no more difficult that calculating the covariance, but it is a bit tedious so we won't go over an example right now.

Multiple Random Variables: Conditioning and Independence

Finally, given two joint random variables, X and Y, we may be interested in characteristics of the distribution of Y *conditional* on X taking a certain value.

Typically this will be denoted $\mathbb{E}[Y|X = x]$ and will be called the conditional expectation of Y given $X = x$.

Finally, given two joint random variables, X and Y, we may be interested in characteristics of the distribution of Y *conditional* on X taking a certain value.

Typically this will be denoted $\mathbb{E}[Y|X = x]$ and will be called the conditional expectation of Y given $X = x$.

- We may be interested in the average income for people with college degrees, $\mathbb{E}[\text{Income} \mid \text{Education} = \text{College Graduate}]$

Finally, given two joint random variables, X and Y, we may be interested in characteristics of the distribution of Y *conditional* on X taking a certain value.

Typically this will be denoted $\mathbb{E}[Y|X = x]$ and will be called the conditional expectation of Y given $X = x$.

- We may be interested in the average income for people with college degrees, $\mathbb{E}[\text{Income} \mid \text{Education} = \text{College Graduate}]$
- We may be interested in the average sales price of a home with floor size 1200 sq. ft, $\mathbb{E}[\text{Sales Price} \mid \text{Sqft} = 1200]$.

Multiple Random Variables: Conditioning and Independence

Finally, given two joint random variables, X and Y, we may be interested in characteristics of the distribution of Y *conditional* on X taking a certain value.

Typically this will be denoted $\mathbb{E}[Y|X = x]$ and will be called the conditional expectation of Y given $X = x$.

- We may be interested in the average income for people with college degrees, $\mathbb{E}[\text{Income} \mid \text{Education} = \text{College Graduate}]$
- We may be interested in the average sales price of a home with floor size 1200 sq. ft, $\mathbb{E}[\text{Sales Price} \mid \text{Sqft} = 1200]$.
- We may be interested in the average lifespan for smokers, $\mathbb{E}[\text{Lifespan} \mid \text{Smoker} = 1]$.

Multiple Random Variables: Conditioning and Independence

Finally, given two joint random variables, X and Y, we may be interested in characteristics of the distribution of Y *conditional* on X taking a certain value.

Typically this will be denoted $\mathbb{E}[Y|X=x]$ and will be called the conditional expectation of Y given $X=x$.

- We may be interested in the average income for people with college degrees, $\mathbb{E}[\text{Income} \mid \text{Education} = \text{College Graduate}]$
- We may be interested in the average sales price of a home with floor size 1200 sq. ft, $\mathbb{E}[\text{Sales Price} \mid \text{Sqft} = 1200]$.
- We may be interested in the average lifespan for smokers, $\mathbb{E}[\text{Lifespan} \mid \text{Smoker} = 1]$.

In all of these note that knowing the conditional expectation is useful for making predictions as we typically observe the X variable before we observe the Y variable.

Calculating the conditional expectation essentially involves fixing the X variable at the point $X = x$, integrating, and then dividing by the probability that $X = x$.

Type	Discrete R.V	Continuous R.V
$\mathbb{E}[Y \mid X = x]$	$\frac{\sum_y y \cdot p_{XY}(y,x)}{\sum_y p_{XY}(x,y)}$	$\frac{\int_Y y \cdot f_{XY}(x,y)\,dy}{\int_Y f_{XY}(x,y)\,dy}$

Calculating the conditional expectation essentially involves fixing the X variable at the point $X = x$, integrating, and then dividing by the probability that $X = x$.

Type	Discrete R.V	Continuous R.V
$\mathbb{E}[Y\|X = x]$	$\frac{\sum_y y \cdot p_{XY}(y,x)}{\sum_y p_{XY}(x,y)}$	$\frac{\int_Y y \cdot f_{XY}(x,y)\, dy}{\int_Y f_{XY}(x,y)\, dy}$

Remark

Note that $\sum_y p_{XY}(x,y) = \mathbb{P}_{XY}(\{X = x\})$. We call the quantity $\int_Y f_{XY}(x,y)\, dy$ the marginal distribution of X at x and denote it $f_X(x)$.

Calculating the conditional expectation essentially involves fixing the X variable at the point $X = x$, integrating, and then dividing by the probability that $X = x$.

Type	Discrete R.V	Continuous R.V
$\mathbb{E}[Y\|X = x]$	$\frac{\sum_y y \cdot p_{XY}(y,x)}{\sum_y p_{XY}(x,y)}$	$\frac{\int_Y y \cdot f_{XY}(x,y)\,dy}{\int_Y f_{XY}(x,y)\,dy}$

Remark

Note that $\sum_y p_{XY}(x, y) = \mathbb{P}_{XY}(\{X = x\})$. We call the quantity $\int_Y f_{XY}(x, y)\,dy$ the marginal distribution of X at x and denote it $f_X(x)$.

We see that $X = x$ is fixed in the above equations while we allow Y to vary. We then divide by the probability that $X = x$ (or the density of X at x, which is the continuous analogue).

Multiple Random Variables: Conditioning and Independence

Let's return to a previous example where Y is the number of hours of sleep one gets a night and X is the number of cups of coffee that one drinks a day.

The joint pmf of X and Y can be described with the table below

$p(x, y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Multiple Random Variables: Conditioning and Independence

Let's return to a previous example where Y is the number of hours of sleep one gets a night and X is the number of cups of coffee that one drinks a day.

The joint pmf of X and Y can be described with the table below

$p(x,y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Let's compute the number of hours of sleep we can expect someone gets conditional on them drinking 2 cups of coffee. To this, let's first calculate the probability that one drinks 2 cups of coffee.

Multiple Random Variables: Conditioning and Independence

Let's return to a previous example where Y is the number of hours of sleep one gets a night and X is the number of cups of coffee that one drinks a day.

The joint pmf of X and Y can be described with the table below

$p(x, y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Let's compute the number of hours of sleep we can expect someone gets conditional on them drinking 2 cups of coffee. To this, let's first calculate the probability that one drinks 2 cups of coffee.

$$\mathbb{P}_{XY}(\{X = x\}) = p_{XY}(x, y) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

Multiple Random Variables: Conditioning and Independence

Let's return to a previous example where Y is the number of hours of sleep one gets a night and X is the number of cups of coffee that one drinks a day.

The joint pmf of X and Y can be described with the table below

$p(x, y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Let's compute the number of hours of sleep we can expect someone gets conditional on them drinking 2 cups of coffee. To this, let's first calculate the probability that one drinks 2 cups of coffee.

$$\mathbb{P}_{XY}(\{X = x\}) = p_{XY}(x, y) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

Now let's fix $X = 2$ and calculate $\sum_y y \cdot p_{XY}(2, y)$

$$\sum_y y \cdot p_{XY}(2, y) = 4 \cdot \frac{1}{6} + 8 \cdot \frac{1}{3} + 12 \cdot 0 = \frac{10}{3}.$$

Multiple Random Variables: Conditioning and Independence

Let's return to a previous example where Y is the number of hours of sleep one gets a night and X is the number of cups of coffee that one drinks a day.

The joint pmf of X and Y can be described with the table below

$p(x, y)$	1 cup	2 cups
4 hours	0	1/6
8 hours	1/3	1/3
12 hours	1/6	0

Let's compute the number of hours of sleep we can expect someone gets conditional on them drinking 2 cups of coffee. To this, let's first calculate the probability that one drinks 2 cups of coffee.

$$\mathbb{P}_{XY}(\{X = x\}) = p_{XY}(x, y) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

Now let's fix $X = 2$ and calculate $\sum_y y \cdot p_{XY}(2, y)$

$$\sum_y y \cdot p_{XY}(2, y) = 4 \cdot \frac{1}{6} + 8 \cdot \frac{1}{3} + 12 \cdot 0 = \frac{10}{3}.$$

Putting this together we find that

$$\mathbb{E}[Y|X = 2] = \frac{10}{3} \cdot \frac{2}{1} = \frac{20}{3} \approx 6.6667.$$

Multiple Random Variables: Conditioning and Independence

This is great, but sometimes knowing X may not help us predict Y. Intuitively, knowledge of X does not give us any additional knowledge of Y.

If this is the case we say that X is independent of Y and denote $X \perp Y$.

Some examples of independent random variables:

Multiple Random Variables: Conditioning and Independence

This is great, but sometimes knowing X may not help us predict Y. Intuitively, knowledge of X does not give us any additional knowledge of Y.

If this is the case we say that X is independent of Y and denote $X \perp Y$.

Some examples of independent random variables:

- Knowing that one coin flip came up heads doesn't give you any information about the next coin flip, we can say that successive coin flips are independent

Multiple Random Variables: Conditioning and Independence

This is great, but sometimes knowing X may not help us predict Y. Intuitively, knowledge of X does not give us any additional knowledge of Y.

If this is the case we say that X is independent of Y and denote $X \perp Y$.

Some examples of independent random variables:

- Knowing that one coin flip came up heads doesn't give you any information about the next coin flip, we can say that successive coin flips are independent
- If I buy lottery tickets, whether or not I win the lottery is independent of the color of the shirt I was wearing that day

Multiple Random Variables: Conditioning and Independence

This is great, but sometimes knowing X may not help us predict Y. Intuitively, knowledge of X does not give us any additional knowledge of Y.

If this is the case we say that X is independent of Y and denote $X \perp Y$.

Some examples of independent random variables:

- Knowing that one coin flip came up heads doesn't give you any information about the next coin flip, we can say that successive coin flips are independent
- If I buy lottery tickets, whether or not I win the lottery is independent of the color of the shirt I was wearing that day
  - Unless I was wearing my lucky t-shirt

Multiple Random Variables: Conditioning and Independence

There are many characterizations of independence, but for now all we care about are the following implications. If $X \perp Y$ then:

$$\mathbb{P}_{XY}(a \leq X \leq b, c \leq Y \leq d) = \mathbb{P}_{XY}(a \leq X \leq b)\mathbb{P}_{XY}(c \leq Y \leq d) \quad \forall a, b, c, d$$

and

$$\mathbb{E}[g(Y)|X = x] = \mathbb{E}[g(Y)] \quad \forall x \in \mathcal{O}_X$$

Assuming two random variables are independent is useful as the above implications will greatly simplify calculations later on. However, we need to be careful when doing so, as it is a rather strong assumption to make.

Let's see some examples of variables that seem independent but may not be

- We may think that weather in a city is independent of average rent, is this the case?
- Is attending a Greta Thunberg rally independent of your likelihood to attend another climate rally?

Questions?