Probability

Nate Wells

Math 141, 3/26/21

Outline

In this lecture, we will. . .

Outline

In this lecture, we will...

- Discuss the Law of Total Probability and Bayes' Rule
- Define and investigate Random Variables

Section 1

Conditional Probability

The Law of Total Probability

Recall the **conditional probability** of an event A given another event B is

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

The Law of Total Probability

Recall the **conditional probability** of an event A given another event B is

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

One useful trick for computing probabilities is the following:

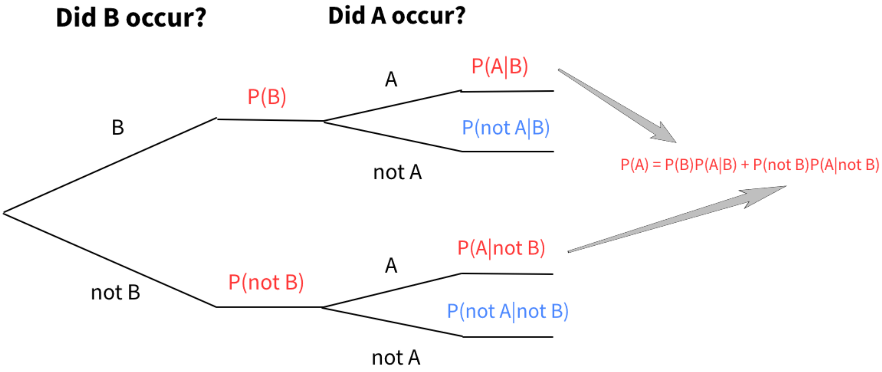
Theorem (The Law of Total Probability)

Let A and B be events. Then

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

- We can often represent the Law of Total Probability using a Tree Diagram:

Tree Diagrams



Lost Marbles

Two boxes contain a different number of red and blue marbles. The first box contains 20% red marbles while the second contains 80% red marbles. Suppose we select a marble from box 1 25% of the time and a marble from box 2 75% of the time. **What is the probability that a red marble is selected?**

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event "the first flip is heads" and let B be the event "at least one flip is heads".

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event “the first flip is heads” and let B be the event “at least one flip is heads”.
  - The event B occurs if we get one of HH, HT, TH. So $P(B) = \frac{3}{4}$

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event "the first flip is heads" and let B be the event "at least one flip is heads".
  - The event B occurs if we get one of HH, HT, TH. So $P(B) = \frac{3}{4}$
  - The event A occurs if we get one of HT or HH, so $P(A) = \frac{1}{2}$.

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event "the first flip is heads" and let B be the event "at least one flip is heads".
  - The event B occurs if we get one of HH, HT, TH. So $P(B) = \frac{3}{4}$
  - The event A occurs if we get one of HT or HH, so $P(A) = \frac{1}{2}$.
  - The events A and B both occur if we get one of HT or HH, so $P(A \text{ and } B) = \frac{1}{2}$.

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event "the first flip is heads" and let B be the event "at least one flip is heads".
  - The event B occurs if we get one of HH, HT, TH. So $P(B) = \frac{3}{4}$
  - The event A occurs if we get one of HT or HH, so $P(A) = \frac{1}{2}$.
  - The events A and B both occur if we get one of HT or HH, so $P(A \text{ and } B) = \frac{1}{2}$.
  - Then

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$$

Is conditional probability symmetric?

Consider two events A and B. Is it always true that $P(A|B) = P(B|A)$?

- Suppose we flip two coins. Let A be the event "the first flip is heads" and let B be the event "at least one flip is heads".
  - The event B occurs if we get one of HH, HT, TH. So $P(B) = \frac{3}{4}$
  - The event A occurs if we get one of HT or HH, so $P(A) = \frac{1}{2}$.
  - The events A and B both occur if we get one of HT or HH, so $P(A \text{ and } B) = \frac{1}{2}$.
  - Then

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$$

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

- Why is this rule true?

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

- Why is this rule true?
- Under what circumstances will $P(A|B) = P(B|A)$?

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

- Why is this rule true?
- Under what circumstances will $P(A|B) = P(B|A)$?
- Under what circumstances will $P(A|B)$ be much larger than $P(B|A)$? Much smaller?

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

- Why is this rule true?
- Under what circumstances will $P(A|B) = P(B|A)$?
- Under what circumstances will $P(A|B)$ be much larger than $P(B|A)$? Much smaller?
- Suppose $P(B|A) = 1$.
  - What does this suggest about A and B?

Bayes' Rule

To relate $P(A|B)$ and $P(B|A)$, we use the following theorem:

Theorem (Bayes' Rule)

Let A and B be events. Then

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

- Why is this rule true?
- Under what circumstances will $P(A|B) = P(B|A)$?
- Under what circumstances will $P(A|B)$ be much larger than $P(B|A)$? Much smaller?
- Suppose $P(B|A) = 1$.
  - What does this suggest about A and B?
  - What is $P(A|B)$ in this case?

Bayes' Rule and Hypothesis Testing

Consider a hypothesis test for a population mean with

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0$$

Suppose we obtain a sample with mean $\bar{x} = 2$.

Bayes' Rule and Hypothesis Testing

Consider a hypothesis test for a population mean with

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0$$

Suppose we obtain a sample with mean $\bar{x} = 2$.

- Let A be the event "The null hypothesis is true" and let B be the event "A random sample has mean larger than $\bar{x} = 2$.

Bayes' Rule and Hypothesis Testing

Consider a hypothesis test for a population mean with

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0$$

Suppose we obtain a sample with mean $\bar{x} = 2$.

- Let A be the event “The null hypothesis is true” and let B be the event "A random sample has mean larger than $\bar{x} = 2$.
- Express the p-value for the sample mean $\bar{x} = 2$ in terms of conditional probabilities.

Bayes' Rule and Hypothesis Testing

Consider a hypothesis test for a population mean with

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0$$

Suppose we obtain a sample with mean $\bar{x} = 2$.

- Let A be the event "The null hypothesis is true" and let B be the event "A random sample has mean larger than $\bar{x} = 2$.
- Express the p-value for the sample mean $\bar{x} = 2$ in terms of conditional probabilities.
- In a typical research setting, we are usually interested in questions of the type:
  - What is the probability that the null hypothesis is true?
  - What is the probability that the null hypothesis is true, given that I observed a particular sample?

Bayes' Rule and Hypothesis Testing

Consider a hypothesis test for a population mean with

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0$$

Suppose we obtain a sample with mean $\bar{x} = 2$.

- Let A be the event "The null hypothesis is true" and let B be the event "A random sample has mean larger than $\bar{x} = 2$.
- Express the p-value for the sample mean $\bar{x} = 2$ in terms of conditional probabilities.
- In a typical research setting, we are usually interested in questions of the type:
  - What is the probability that the null hypothesis is true?
  - What is the probability that the null hypothesis is true, given that I observed a particular sample?
- Express each of these in terms of the p-value and conditional probabilities.

Here there be monsters

Let H denote a hypothesis and O denote some observation.

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).
- The probability $P(O|H)$ for his observation is actually very high.
  - If there were monsters partying in the closet, they would certainly be very noisy!

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).
- The probability $P(O|H)$ for his observation is actually very high.
  - If there were monsters partying in the closet, they would certainly be very noisy!
- But the fact that a noise was heard doesn't make it very probable that there are party monsters in the closet.
  - I try to reassure him that even though he heard noises, $P(H|O)$ is still low.

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).
- The probability $P(O|H)$ for his observation is actually very high.
  - If there were monsters partying in the closet, they would certainly be very noisy!
- But the fact that a noise was heard doesn't make it very probable that there are party monsters in the closet.
  - I try to reassure him that even though he heard noises, $P(H|O)$ is still low.
- In this case, $P(O|H)$ is high and $P(H|O)$ is low. Why?

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).
- The probability $P(O|H)$ for his observation is actually very high.
  - If there were monsters partying in the closet, they would certainly be very noisy!
- But the fact that a noise was heard doesn't make it very probable that there are party monsters in the closet.
  - I try to reassure him that even though he heard noises, $P(H|O)$ is still low.
- In this case, $P(O|H)$ is high and $P(H|O)$ is low. Why?
  - $P(H)$ is very, very low compared to $P(O)$.

Here there be monsters

Let H denote a hypothesis and O denote some observation.

- Suppose my son Oliver hears a noise coming from his closet (O). He considers the hypothesis that there are monsters throwing a party (H).
- The probability $P(O|H)$ for his observation is actually very high.
  - If there were monsters partying in the closet, they would certainly be very noisy!
- But the fact that a noise was heard doesn't make it very probable that there are party monsters in the closet.
  - I try to reassure him that even though he heard noises, $P(H|O)$ is still low.
- In this case, $P(O|H)$ is high and $P(H|O)$ is low. Why?
  - $P(H)$ is very, very low compared to $P(O)$.
- By the principle of hypothesis testing, we might see that $P(O|H)$ is high and favor the monster hypothesis H.
- Hypothesis testing just tells us about consistency of data with the null hypothesis. It doesn't give us the probability that the null is true.

Section 2

Random Variables

Definitions

A **random variable** is a numeric quantity whose value depends on the outcome of a random process.

Definitions

A **random variable** is a numeric quantity whose value depends on the outcome of a random process.

- We use capital letters at the end of the alphabet (W, X, Y, Z) to denote random variables.
  - We use lowercase letters (w, x, y, z) to denote the particular values of a random variable

Definitions

A **random variable** is a numeric quantity whose value depends on the outcome of a random process.

- We use capital letters at the end of the alphabet (W, X, Y, Z) to denote random variables.
  - We use lowercase letters (w, x, y, z) to denote the particular values of a random variable
- We use equation to express events associated to random variables.
  - Let $X = 5$ denotes the event "The random variable X takes the value 5".

Definitions

A **random variable** is a numeric quantity whose value depends on the outcome of a random process.

- We use capital letters at the end of the alphabet (W, X, Y, Z) to denote random variables.
  - We use lowercase letters (w, x, y, z) to denote the particular values of a random variable
- We use equation to express events associated to random variables.
  - Let $X = 5$ denotes the event "The random variable X takes the value 5".
- Events associated to variables have probabilities of occurring.
  - $P(X = 5) = .5$ means X has 50% probability of taking the value 5.

Types of Random Variables

There are two main types of random variables:

1. **Discrete** variables can take only finitely many different values.
2. **Continuous** variables can take values equal to any real number in an interval.

Types of Random Variables

There are two main types of random variables:

1. **Discrete** variables can take only finitely many different values.
2. **Continuous** variables can take values equal to any real number in an interval.

- Examples of discrete variables:
  - The number of credits a randomly chosen Reed student is taking.
  - The number of vegetarians in a random sample of 10 people.
  - The results of a coin flip, where 0 indicates Tails and 1 indicates Heads.

Types of Random Variables

There are two main types of random variables:

1. **Discrete** variables can take only finitely many different values.
2. **Continuous** variables can take values equal to any real number in an interval.

- Examples of discrete variables:
  - The number of credits a randomly chosen Reed student is taking.
  - The number of vegetarians in a random sample of 10 people.
  - The results of a coin flip, where 0 indicates Tails and 1 indicates Heads.
- Examples of continuous variables:
  - The temperature of my office at a particular time of the day.
  - The amount of time it takes a radioactive particle to decay.

Types of Random Variables

There are two main types of random variables:

1. **Discrete** variables can take only finitely many different values.
2. **Continuous** variables can take values equal to any real number in an interval.

- Examples of discrete variables:
  - The number of credits a randomly chosen Reed student is taking.
  - The number of vegetarians in a random sample of 10 people.
  - The results of a coin flip, where 0 indicates Tails and 1 indicates Heads.
- Examples of continuous variables:
  - The temperature of my office at a particular time of the day.
  - The amount of time it takes a radioactive particle to decay.
- Some discrete variables can be well-described by continuous variables:
  - The height of a random person selected from a large population.
  - The proportion of heads in a long sequence of coin flips.

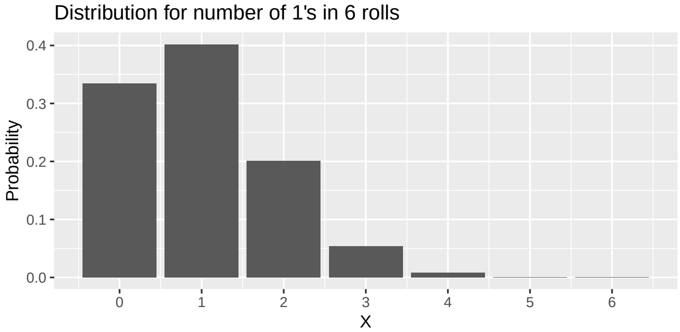
The Distribution of a Random Variable

We often use histograms or bar charts to visualize discrete random variables.

The Distribution of a Random Variable

We often use histograms or bar charts to visualize discrete random variables.

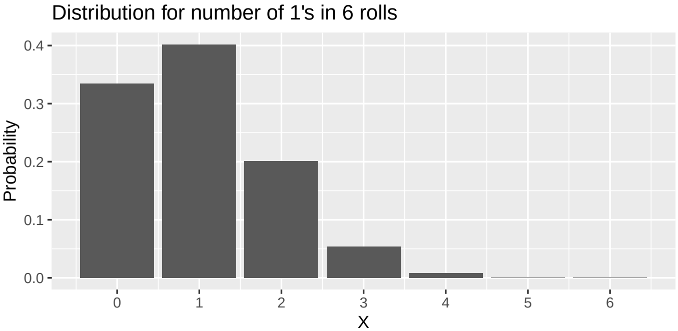
- Suppose a fair 6-sided die is rolled 6 times. Let X be the number of 1s rolled. The distribution of X is given by:



The Distribution of a Random Variable

We often use histograms or bar charts to visualize discrete random variables.

- Suppose a fair 6-sided die is rolled 6 times. Let X be the number of 1s rolled. The distribution of X is given by:

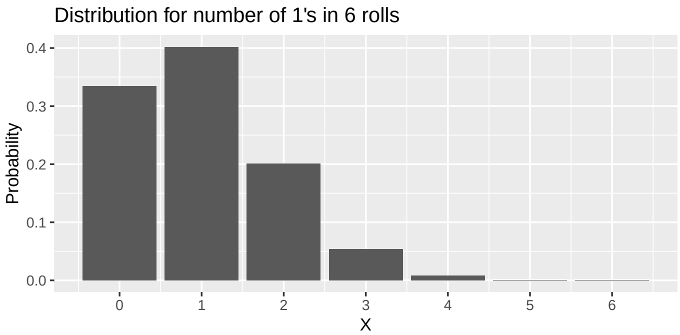


- We can use the plot to find probabilities of outcomes associated to the variable.

The Distribution of a Random Variable

We often use histograms or bar charts to visualize discrete random variables.

- Suppose a fair 6-sided die is rolled 6 times. Let X be the number of 1s rolled. The distribution of X is given by:



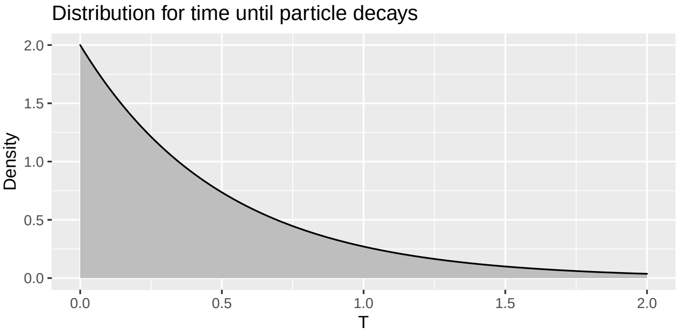
- We can use the plot to find probabilities of outcomes associated to the variable.
  - Calculate $P(X \leq 1)$. Then find x so that $P(X \leq x) \geq .75$.

The Distribution of a Continuous Variable

- We use **density plots** to visualize the distribution of a continuous variable. Areas under the plot correspond to probabilities.

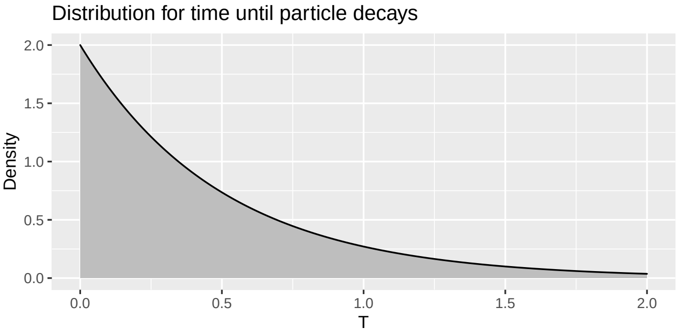
The Distribution of a Continuous Variable

- We use **density plots** to visualize the distribution of a continuous variable. Areas under the plot correspond to probabilities.
- The distribution for the amount of time T until a radioactive particle decays is given below:



The Distribution of a Continuous Variable

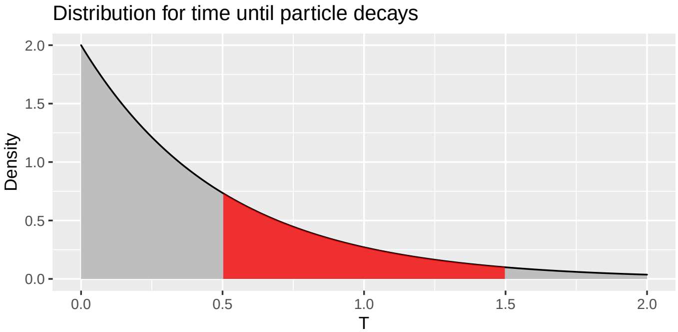
- We use **density plots** to visualize the distribution of a continuous variable. Areas under the plot correspond to probabilities.
- The distribution for the amount of time T until a radioactive particle decays is given below:



- The probability that it takes between 0.5 and 1.5 units of time to decay is the area under the curve between 0.5 and 1.5. $P(0.5 < T < 1.5) =$

The Distribution of a Continuous Variable

- We use **density plots** to visualize the distribution of a continuous variable. Areas under the plot correspond to probabilities.
- The distribution for the amount of time T until a radioactive particle decays is given below:



- The probability that it takes between 0.5 and 1.5 units of time to decay is the area under the curve between 0.5 and 1.5. $P(0.5 < T < 1.5) = 0.34$

Using Densities for Discrete Variables

If a discrete variable takes a large number of values which are close together, we can often approximate it using a continuous variable.

Using Densities for Discrete Variables

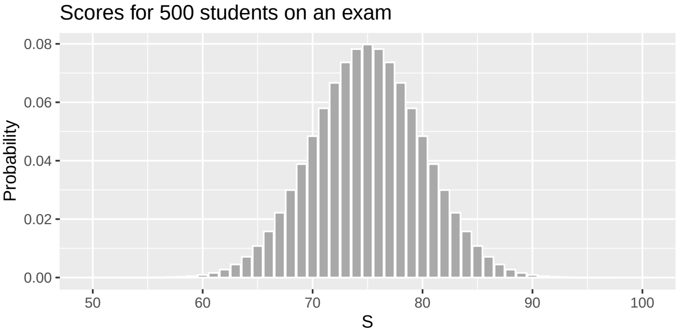
If a discrete variable takes a large number of values which are close together, we can often approximate it using a continuous variable.

- Suppose 500 students take a standardized exam, with mean 75 points. The distribution for the score S of a randomly chosen student is:

Using Densities for Discrete Variables

If a discrete variable takes a large number of values which are close together, we can often approximate it using a continuous variable.

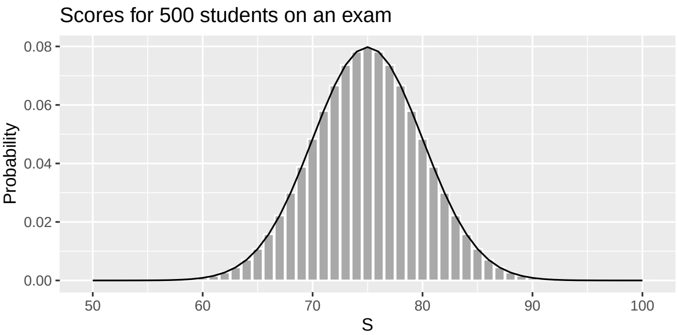
- Suppose 500 students take a standardized exam, with mean 75 points. The distribution for the score S of a randomly chosen student is:



Using Densities for Discrete Variables

If a discrete variable takes a large number of values which are close together, we can often approximate it using a continuous variable.

- Suppose 500 students take a standardized exam, with mean 75 points. The distribution for the score S of a randomly chosen student is:



Expected Value

The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Expected Value

The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

- The expected value of X is the sum of the value X can take, weighted by the probability it takes those values.

Expected Value

The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

- The expected value of X is the sum of the value X can take, weighted by the probability it takes those values.
- Suppose we have a data set consisting of values $\{1, 1, 2, 2, 2, 2, 3, 4, 5, 5\}$. Let X be a value chosen from this data set randomly. What is the expected value of X?

Expected Value

The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

- The expected value of X is the sum of the value X can take, weighted by the probability it takes those values.
- Suppose we have a data set consisting of values $\{1, 1, 2, 2, 2, 2, 3, 4, 5, 5\}$. Let X be a value chosen from this data set randomly. What is the expected value of X?

$$\begin{aligned} E[X] =& 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + 4P(X = 4) + 5P(X = 5) \\ =& 1\frac{2}{10} + 2\frac{4}{10} + 3\frac{1}{10} + 4\frac{1}{10} + 5\frac{2}{10} = \frac{27}{10} = 2.7 \end{aligned}$$

Expected Value

The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1P(X = x_1) + x_2P(X = x_2) + \dots x_nP(X = x_n) = \sum_{i=1}^{n} x_iP(X = x_i)$$

- The expected value of X is the sum of the value X can take, weighted by the probability it takes those values.
- Suppose we have a data set consisting of values $\{1, 1, 2, 2, 2, 2, 3, 4, 5, 5\}$. Let X be a value chosen from this data set randomly. What is the expected value of X?

$$\begin{aligned} E[X] =& 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + 4P(X = 4) + 5P(X = 5) \\ =& 1\frac{2}{10} + 2\frac{4}{10} + 3\frac{1}{10} + 4\frac{1}{10} + 5\frac{2}{10} = \frac{27}{10} = 2.7 \end{aligned}$$

- But also notice that

$$\begin{aligned} E[X] =& \frac{1}{10}\left(1 \cdot 2 + 2 \cdot 4 + 3 \cdot 1 + 4 \cdot 1 + 5 \cdot 2\right) \\ =& \frac{1}{10}\left(1 + 1 + 2 + 2 + 2 + 2 + 3 + 4 + 5 + 5\right) \end{aligned}$$

The Law of Large Numbers, again

Previously, we said that by the Law of Large numbers, the proportion of times an outcome occurs in a long sequence of trials is close to the probability for that outcome.

The Law of Large Numbers, again

Previously, we said that by the Law of Large numbers, the proportion of times an outcome occurs in a long sequence of trials is close to the probability for that outcome.

This is a generalization:

Theorem (The Law of Large Numbers)

Let X be a random variable whose value depends on a random experiment. Suppose the experiment is repeated n times and let $\bar{x}_n$ denote the arithmetic mean of the values of X in each trial. As n gets larger, the arithmetic mean $\bar{x}_n$ approaches the expected value $E[X]$ of that variable.

Gambler's Ruin

A roulette wheel consists of 37 wedge (18 black, 18 red, 1 green). A player may bet \$10 that a spun ball will land on a black wedge. If the ball lands on black, the player wins \$10. Otherwise, the player loses \$10.

Gambler's Ruin

A roulette wheel consists of 37 wedge (18 black, 18 red, 1 green). A player may bet $10 that a spun ball will land on a black wedge. If the ball lands on black, the player wins $10. Otherwise, the player loses $10.

- Assuming each wedge has equal probability, what is the expected value of the bet?

- Suppose a gambler begins with $10,000. What will the gambler's fortune look like after 1000 plays?