Objectives

By the end of this set of slides, you should be able to:

1. Recognize and understand discrete probability distribution functions
2. Calculate and interpret expected values
3. Recognize the binomial probability distribution and apply it appropriately
Random Variables

- **Random Variable**: A random variable describes the outcomes of a statistical experiment
- Random variables can be of two types:
  - **Discrete random variables**: can take only a finite or countable number of outcomes. Example: number of patients with a particular disease
  - **Continuous random variables**: take values within a specified interval or continuum. Example: height or weight
- The values of a random variable can vary with each repetition of an experiment
- We will use capital letters to denote random variables and lower case letters to denote the value of a random variable
- If $X$ is a random variable, then $X$ is written in words, and $x$ is given a number
Example #1

- Let $X =$ the number of heads you get when you toss three fair coins
- The sample space $S = \{HHH, THH, HTH, HHT, TTH, HTT, THT, TTT\}$
- Then, $x = 0, 1, 2, \text{ or } 3$. Why?
- $X$ is in words and $x$ is a number
- The $x$ values are countable (discrete) outcomes
- Because you can count the possible values that $X$ can take on and the outcomes are random (the $x$ values 0, 1, 2, 3), $X$ is a discrete random variable
A discrete **probability distribution function** has two characteristics:

1. Each probability is between zero and one, inclusive
2. The sum of the probabilities is one

The probability distribution function of a discrete random variable specifies all the possible outcomes of the random variable along with the probability that each will occur.

Example: Suppose Nancy has classes **three days** a week. She attends classes **three days** a week 80% of the time, **two days** 15% of the time, **one day** 4% of the time, and **no days** 1% of the time.

- \( X = \) the number of days Nancy attends classes
- \( x \) could be 0, 1, 2, or 3

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X=x) )</td>
<td>0.01</td>
<td>0.04</td>
<td>0.15</td>
<td>0.80</td>
</tr>
</tbody>
</table>
Example #2

- Students who live in the dormitories at a certain four year college must buy a meal plan. They must select from four meal plans: 10 meals, 14 meals, 18 meals, or 21 meals per week. The Food and Housing Office has determined that the 15% of students purchase 10 meal plan, 45% of students purchase the 14 meal plan, 30% of students purchase the 18 meal plan and 10% of students purchase the 21 meal plan.

- What is the random variable \( X \)?

- Notation: In general, \( P(X) = \) probability value

- In our example,
  - \( P(X = 10) \) is the probability that a student purchases a meal plan with 10 meals per week
  - \( P(X > 14) \) is the probability that a student purchases a meal plan with more than 14 meals per week
Example #2 Continued

- Make a table that shows the probability distribution

<table>
<thead>
<tr>
<th>$x$ = the number of meals</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.15</td>
</tr>
<tr>
<td>14</td>
<td>0.45</td>
</tr>
<tr>
<td>18</td>
<td>0.30</td>
</tr>
<tr>
<td>21</td>
<td>0.10</td>
</tr>
</tbody>
</table>

- What is the probability that a student purchases more than 14 meals? i.e.,

$$P(X > 14) = ???$$

- What is the probability that a student doesn’t purchase 21 meals?, i.e.,

$$P(X < 21) = ???$$
Mean or Expected Value

- The expected value is often referred to as the "long-term" average or mean.
- Over the long term of doing an experiment over and over, you would expect this average.
- Imagine flipping a coin, over and over and over and over...
- Karl Pearson (who?) once tossed a fair coin 24,000 times!
- He recorded the results of each coin toss, obtaining heads 12,012 times.
- \( \frac{12012}{24000} \approx 0.5 \)
- The mean or expected value of an experiment is denoted by \( \mu \)
Mean or Expected Value

• To find the expected value or long term average, $\mu$, simply multiply each value of the random variable by its probability and add the products

• What is the expected value of Example #2?

<table>
<thead>
<tr>
<th>$x$ = the number of meals</th>
<th>$P(x)$</th>
<th>$x \cdot P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.15</td>
<td>$10 \cdot (0.15) = 1.5$</td>
</tr>
<tr>
<td>14</td>
<td>0.45</td>
<td>$14 \cdot (0.45) = 6.3$</td>
</tr>
<tr>
<td>18</td>
<td>0.30</td>
<td>$18 \cdot (0.30) = 5.4$</td>
</tr>
<tr>
<td>21</td>
<td>0.10</td>
<td>$21 \cdot (0.10) = 2.1$</td>
</tr>
</tbody>
</table>

• $\mu = \text{Expected Value} = 1.5 + 6.3 + 5.4 + 2.1 = 15.3$

• So in general,

$$\mu = \sum x \cdot P(x)$$
Example #3

- Let $X$ = the number of heads you get when you toss four fair coins
- The sample space

$$ S = \{ HHHH, HTHH, HHHT, HTTH, HHTT, HTHT, HTTT, THHH, TTHH, THTH, THHT, TTTT \} $$

- Then, $x = 0, 1, 2, 3$ or $4$. Why?
- What is the expected value of $X$?

<table>
<thead>
<tr>
<th>$x$ = the number of heads</th>
<th>$P(x)$</th>
<th>$x \cdot P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/16</td>
<td>0 \cdot (1/16) = 0</td>
</tr>
<tr>
<td>1</td>
<td>4/16</td>
<td>1 \cdot (4/16) = 0.25</td>
</tr>
<tr>
<td>2</td>
<td>6/16</td>
<td>2 \cdot (6/16) = 0.75</td>
</tr>
<tr>
<td>3</td>
<td>4/16</td>
<td>3 \cdot (4/16) = 0.75</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>4 \cdot (1/16) = 0.25</td>
</tr>
</tbody>
</table>
Example #3 Continued

- What is the expected value of $X$?

<table>
<thead>
<tr>
<th>$x$ = the number of heads</th>
<th>$P(x)$</th>
<th>$x \cdot P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1/16$</td>
<td>$0 \cdot (1/16) = 0$</td>
</tr>
<tr>
<td>1</td>
<td>$4/16$</td>
<td>$1 \cdot (4/16) = 0.25$</td>
</tr>
<tr>
<td>2</td>
<td>$6/16$</td>
<td>$2 \cdot (6/16) = 0.75$</td>
</tr>
<tr>
<td>3</td>
<td>$4/16$</td>
<td>$3 \cdot (4/16) = 0.75$</td>
</tr>
<tr>
<td>4</td>
<td>$1/16$</td>
<td>$4 \cdot (1/16) = 0.25$</td>
</tr>
</tbody>
</table>

- The expected value of $X$ is

$$\mu = \sum x \cdot P(x) = 0 + 0.25 + 0.75 + 0.75 + 0.25 = 2$$

- Is this answer intuitive?
Law of Large Numbers

- The Law of Large Numbers describes the result of performing the same experiment a large number of times
- According to the Law of Large Numbers, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed
- Where does the Law of Large Numbers come up in our every day lives?
  - Counting the number of heads in \( n \) tosses of a fair coin
  - Playing a casino game, like roulette, in Las Vegas
  - Games of chance involving rolling a die
  - Playing the lottery
  - Other examples?
Law of Large Numbers

Imagine tossing a fair coin for a very long time...

<table>
<thead>
<tr>
<th>Number Tosses</th>
<th>Number Heads</th>
<th>Expected Number Heads</th>
<th>Chance Error</th>
<th>Percentage Heads</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>70%</td>
</tr>
<tr>
<td>50</td>
<td>30</td>
<td>25</td>
<td>5</td>
<td>60%</td>
</tr>
<tr>
<td>100</td>
<td>56</td>
<td>50</td>
<td>6</td>
<td>56%</td>
</tr>
<tr>
<td>1000</td>
<td>470</td>
<td>500</td>
<td>30</td>
<td>47%</td>
</tr>
<tr>
<td>100000</td>
<td>51000</td>
<td>50000</td>
<td>1000</td>
<td>51%</td>
</tr>
</tbody>
</table>

- As I toss the coin more and more the percentage of heads approaches 0.5 -- This is the Law of Large Numbers at work!
Example #4

- Imagine you play a game where you flip a fair coin \( n \) times
- You win $100 if you obtain more than 75% heads
- **Question**: Would you want to throw the coin 10 times or 1000 times?
Standard Deviation

- We can calculate the standard deviation of a random variable $X$ as the following:

$$
\sigma = \sqrt{\sum (x - \mu)^2 \cdot P(x)}
$$

- Let $X =$ the number of heads you get when you toss four fair coins

<table>
<thead>
<tr>
<th>$x =$ the # of heads</th>
<th>$P(x)$</th>
<th>$x \cdot P(x)$</th>
<th>$(x - \mu)^2 \cdot P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/16</td>
<td>0 \cdot (1/16) = 0</td>
<td>(0 - 2)^2 \cdot (1/16) = 0.25</td>
</tr>
<tr>
<td>1</td>
<td>4/16</td>
<td>1 \cdot (4/16) = 0.25</td>
<td>(1 - 2)^2 \cdot (4/16) = 0.25</td>
</tr>
<tr>
<td>2</td>
<td>6/16</td>
<td>2 \cdot (6/16) = 0.75</td>
<td>(2 - 2)^2 \cdot (6/16) = 0</td>
</tr>
<tr>
<td>3</td>
<td>4/16</td>
<td>3 \cdot (4/16) = 0.75</td>
<td>(3 - 2)^2 \cdot (4/16) = 0.25</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>4 \cdot (1/16) = 0.25</td>
<td>(4 - 2)^2 \cdot (1/16) = 0.25</td>
</tr>
</tbody>
</table>

- The standard deviation of $X$ is

$$
\sigma = \sqrt{\sum (x - \mu)^2 \cdot P(x)} = \sqrt{0.25 + 0.25 + 0 + 0.25 + 0.25} = 1
$$
Example #5

- At the county fair, a booth has a coin flipping game
- You pay $2.50 to flip two fair coins
- If the result contains one or two heads, you win $3
- If the result is two tails then there is no prize
- **Question**: Write the probability distribution function for the amount won or lost in one game

  - What is the random variable $X$?
  - The sample space $S = \{HH, HT, TH, TT\}$

<table>
<thead>
<tr>
<th>$x$ = amount of money won</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.50$</td>
<td>$3/4$</td>
</tr>
<tr>
<td>$-2.50$</td>
<td>$1/4$</td>
</tr>
</tbody>
</table>
Example #5 Continued

- **Question:** Find the expected value for this game (Expected NET GAIN OR LOSS)

<table>
<thead>
<tr>
<th>$x$ = amount of money won</th>
<th>$P(x)$</th>
<th>$x \cdot P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.50$</td>
<td>$3/4$</td>
<td>$0.5 \cdot (3/4) = 0.375$</td>
</tr>
<tr>
<td>-$2.50$</td>
<td>$1/4$</td>
<td>$-2.5 \cdot (1/4) = -0.625$</td>
</tr>
</tbody>
</table>

$$
\mu = \sum x \cdot P(x) = 0.50 \cdot \left(\frac{3}{4}\right) + (-2.50) \cdot \left(\frac{1}{4}\right) = -0.25
$$

- So you should expect to lose, on average, 25 cents per game ($0.25) after playing the game over and over again.

- **Question:** Find the expected total net gain or loss if you play this game 100 times

- $100 \times \mu = 100 \times (-$0.25) = -$25$
Binomial Distribution

- There are four characteristics of a binomial experiment
  1. There are a fixed number of trials \( n \)
  2. Each trial has two possible outcomes that we classify as "success" or "failure." The two possible outcomes will be mutually exclusive.
  3. The outcomes of the \( n \) trials are independent. The outcome of one trial does not influence the outcome of any other trial.
  4. The probability of success \( p \) is constant for each trial.

- In a binomial experiment, the random variable \( X \) is the number of successes during the \( n \) trials.

- Examples of a binomial experiment?

- Can a binomial experiment be done without replacement?

- Notation: \( X \sim B(n, p) \)
Binomial Distribution

- Let $X$ be a binomial random variable
- We say that $X \sim B(n, p)$
- The binomial probability distribution is

$$P(X = x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot (1 - p)^{n-x}$$

▷ $n$ is the number of trials
▷ $x$ is the number of successes in $n$ trials
▷ $p$ is the probability of success
▷ $(1 - p)$ is the probability of failure

- Recall: $!$ is factorial notation
- Example: $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$
Example #6

- Let $X$ = the number of heads you get when you toss four fair coins
- The sample space

$$S = \{HHHH, HTHH, HHTH, HHTT, HTTH, HHTT, HTTT, THHH, TTHH, THTH, THHT, THTT, TTHT, TTTH, TTTT\}$$

- What is the probability of getting 2 heads?

<table>
<thead>
<tr>
<th>$x$ = the number of heads</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/16</td>
</tr>
<tr>
<td>1</td>
<td>4/16</td>
</tr>
<tr>
<td>2</td>
<td>6/16</td>
</tr>
<tr>
<td>3</td>
<td>4/16</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
</tr>
</tbody>
</table>

- $P(X = 2) = 6/16 = 0.375$
- **Question**: Is there another way to solve this?
Example #6 Continued

- Is $X$ a binomial random variable?
  - $n = 4$ so fixed number of trials $✓$
  - There are only two mutually exclusive outcomes: Heads or Tail $✓$
    - We define a "success" as getting a Head and a "failure" as getting a Tail
  - The outcomes of trials are independent $✓$
  - The probability of success $p = 0.5$ for each trial $✓$

- Yes! $X$ is a binomial random variable

- What is the probability of getting 2 heads?

$$P(X = 2) = \frac{4!}{(4 - 2)!2!} \times 0.5^2 \times (1 - 0.5)^{4-2}$$

$$= \frac{24}{4} \times 0.25 \times (0.25)$$

$$= 0.375$$
Example #7

- So why is the binomial distribution useful? Look below!
- Let \(X\) = the number of heads you get when you toss 10 fair coins
- The sample space \(S = \{HHHHHHHHHH, HTHHHHHHHH, \ldots\}\)
- Turns out there are 1024 ways to flip 10 coins
- Is the sample space easy to write down? NO!
- So what is the probability I get 8 heads?

\[
P(X = 8) = \frac{10!}{(10 - 8)!8!} \times 0.5^8 \times (1 - 0.5)^{10-8}
\]

\[
= \frac{3628800}{80640} \times 0.00390625 \times (0.25)
\]

\[
= 0.08789062
\]

- Fast and easy to calculate using binomial distribution
Number of Combinations

- So what does the value of
\[ \frac{n!}{(n - x)!x!} \]
represent?
- It is the number of ways to get a success out of the total number of trials.
- For example, think of flipping a coin 4 times and you want to get heads twice.
- The sample space is
\[ S = \{ HHHH, HTHH, HHTH, HHHT, HTTH, HHTT, HTHT, HTTT, TTHH, TTHH, THTH, THHT, TTTH, THTT, TTHT, TTHT, TTTT \} \]
- So there is 6 ways to get two heads.
For example, think of flipping a coin 4 times and you want to get heads twice.

So, \( n = 4 \) and \( x = 2 \).

So the number of combinations is

\[
\binom{n}{x} = \frac{n!}{(n-x)!x!} = \frac{4!}{(4-2)!2!} = \frac{24}{4} = 6
\]

Notation:

\[
\binom{n}{x} = \frac{n!}{(n-x)!x!}
\]
Expected Value (Mean) and Standard Deviation

- Let $X \sim B(n, p)$ then:
  - The expected value of $X$ is
    $$\mu = np$$
  - The standard deviation of $X$ is
    $$\sigma = \sqrt{np(1-p)}$$

- These shortcut formulas for $\mu$ and $\sigma$ give the same results as the definitions $\mu = \sum x \cdot P(x)$ and $\sigma = \sqrt{\sum (x - \mu)^2 \cdot P(x)}$ but with a lot less work!
Example #8

- Recall the following probability distribution table:
- Let \( X \) = the number of heads you get when you toss four fair coins

<table>
<thead>
<tr>
<th>( x ) = the # of heads</th>
<th>( P(x) )</th>
<th>( x \cdot P(x) )</th>
<th>( (x - \mu)^2 \cdot P(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/16</td>
<td>0 \cdot (1/16) = 0</td>
<td>(0 - 2)^2 \cdot (1/16) = 0.25</td>
</tr>
<tr>
<td>1</td>
<td>4/16</td>
<td>1 \cdot (4/16) = 0.25</td>
<td>(1 - 2)^2 \cdot (4/16) = 0.25</td>
</tr>
<tr>
<td>2</td>
<td>6/16</td>
<td>2 \cdot (6/16) = 0.75</td>
<td>(2 - 2)^2 \cdot (6/16) = 0</td>
</tr>
<tr>
<td>3</td>
<td>4/16</td>
<td>3 \cdot (4/16) = 0.75</td>
<td>(3 - 2)^2 \cdot (4/16) = 0.25</td>
</tr>
<tr>
<td>4</td>
<td>1/16</td>
<td>4 \cdot (1/16) = 0.25</td>
<td>(4 - 2)^2 \cdot (1/16) = 0.25</td>
</tr>
</tbody>
</table>

- The expected value of \( X \) is

\[
\mu = \sum x \cdot P(x) = 0 + 0.25 + 0.75 + 0.75 + 0.25 = 2
\]

- The standard deviation of \( X \) is

\[
\sigma = \sqrt{\sum (x - \mu)^2 \cdot P(x)} = \sqrt{0.25 + 0.25 + 0 + 0.25 + 0.25} = 1
\]
Example #8 Continued

- Let \( X \) = the number of heads you get when you toss four fair coins
- Is \( X \) a binomial random variable?
- So, \( n = 4 \) and \( p = 0.5 \)
- The expected value of \( X \) is

\[
\mu = np = 4 \cdot (0.5) = 2
\]

- The standard deviation of \( X \) is

\[
\sigma = \sqrt{np(1-p)} = \sqrt{4 \cdot (0.5) \cdot (0.5)} = 1
\]

- Much faster to use the formulas for expected value and standard deviation of a binomial random variable!
Example #9

- A college claims that 70% of students receive financial aid. Suppose that 10 students at the college are randomly selected. We are interested in the number of students in the sample who receive financial aid.

- Is this a binomial experiment?
- What is the random variable $X$?
- What is $n$?
- What is $p$?
Example #9 Continued

- A college claims that 70% of students receive financial aid. Suppose that 10 students at the college are randomly selected. We are interested in the number of students in the sample who receive financial aid.

- $X$ is the number of students who received financial aid
- The number of trials is $n = 10$
- The probability of receiving financial aid is $p = 0.7$
- **Question:** What is the probability that two students received financial aid?

$$P(X = 2) = \frac{10!}{(10 - 2)!2!} \cdot (0.7^2) \cdot (1 - 0.7)^8 = 0.0014467$$
Example #9 Continued

- **Question:** What is the probability that less than two students received financial aid?

\[ P(X < 2) = P(X = 0) + P(X = 1) \]

\[ = \frac{10!}{(10 - 0)!0!}(0.7^0)(1 - 0.7)^{10} + \frac{10!}{(10 - 1)!1!}(0.7^1)(1 - 0.7)^9 \]

\[ = 0.0001436859 \]

- **Question:** What is the probability that less than or equal to 9 students received financial aid?

\[ P(X \leq 9) = P(X = 0) + P(X = 1) + \ldots + P(X = 8) + P(X = 9) \]

or

\[ P(X \leq 9) = 1 - P(X = 10) \]
Example #9 Continued

- What is the expected number of students who should receive financial aid?
  - The expected number of students who should receive financial aid is
    \[ \mu = np = 10 \cdot (0.7) = 7 \]
- What is the standard deviation?
  - \[ \sigma = \sqrt{np(1 - p)} = \sqrt{10(0.7)(0.3)} = \sqrt{2.1} = 1.449138 \]
Other Discrete Probability Distributions

- Some other well known discrete probability distributions:
  - The Poisson distribution
  - The geometric distribution
  - The hypergeometric distribution
  - The negative-binomial distribution

- And the list goes on...