Chapter 15
Random Variables

click to add subtitles
Objectives:

- Define a random variable.
- Define a discrete random variable.
- Explain what is meant by a probability distribution or model.
- Construct the probability model for a discrete random variable.
- Construct a probability histogram.
Objectives continued,

- Expected value
- Variance of a random variable
- Standard deviation of a random variable
- Linear transformations of random variables
- Define a continuous random variable.
- Probability distribution for a continuous random variable.
Random Variables

- A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.

- For example: Flip three coins and let $X$ represent the number of heads. $X$ is a random variable.

- The sample space $S$ lists the possible values of the random variable $X$
Random Variable

- A random variable assumes a value based on the outcome of a random event.
  - We use a capital letter, like $X$, to denote a random variable.
  - A particular value of a random variable will be denoted with a lower case letter, in this case $x$.
  - Example: $P(X = x)$
Random Variables

- A **random variable** is a function that assigns a numerical value to each simple event in a sample space $S$.
- If these numerical values are only integers (no fractions or irrational numbers), it is called a **discrete random variable**.
- Note that a random variable is neither random nor a variable - it is a function with a numerical value, and it is defined on a sample space.
Random Variables

- There are two types of random variables:
  - **Discrete** random variables can take one of a finite number of distinct outcomes.
    - Example: Number of credit hours
  - **Continuous** random variables can take any numeric value within a range of values.
    - Example: Cost of books this term
Examples of Random Variables

1. A function whose range is the number of speeding tickets issued on a certain stretch of I 95 S.

2. A function whose range is the number of heads which appear when 4 dimes are tossed.

3. A function whose range is the number of passes completed in a game by a quarterback.

These examples are all discrete random variables.
The simple events in a sample space $S$ could be anything: heads or tails, marbles picked out of a bag, playing cards.

The point of introducing random variables is to associate the simple events with numbers, with which we can calculate.

We transfer the probability assigned to elements or subsets of the sample space to numbers. This is called the probability distribution of the random variable $X$. It is defined as

$$p(x) = P(X = x)$$
Discrete Random Variable

- A **discrete random variable** $X$ has a countable number of possible values.

- For example: Flip three coins and let $X$ represent the number of heads. $X$ is a discrete random variable. (In this case, the random variable $X$ can equal 0, 1, 2, or 3.)

- We can use a table to show the probability distribution of a discrete random variable.
Example - Discrete Random Variable

- For example, the number of days it rained in your community during the month of March is an example of a discrete random variable.
- If $X$ is the number of days it rained during the month of March, then the possible values for $X$ are $x = 0, 1, 2, 3, \ldots, 31$. 
Discrete Probability Distribution

Table

| Value of X: |   |   |   |   |   |
|            | $x_1$ | $x_2$ | $x_3$ | $\ldots$ | $x_n$ |
| Probability: | $p_1$ | $p_2$ | $p_3$ | $\ldots$ | $p_n$ |
# Probability Distribution Table: Number of Heads Flipping 4 Coins

<table>
<thead>
<tr>
<th></th>
<th>TTTT</th>
<th>TTTTH</th>
<th>TTHH</th>
<th>THHH</th>
<th>HHHH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TTTT</td>
<td>TTTTH</td>
<td>TTHH</td>
<td>THHH</td>
<td>HHHH</td>
</tr>
<tr>
<td>X</td>
<td>TTTT</td>
<td>TTTTH</td>
<td>TTHH</td>
<td>THHH</td>
<td>HHHH</td>
</tr>
<tr>
<td>P(X)</td>
<td>1/16</td>
<td>4/16</td>
<td>6/16</td>
<td>4/16</td>
<td>1/16</td>
</tr>
</tbody>
</table>
Discrete Probability Distributions

Can also be shown using a histogram

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

\[ \bar{x} = \frac{0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16}}{16} = \frac{12}{16} = \frac{3}{4} \]

Probability Distribution Table: Number of Heads Flipping 4 Coins

<table>
<thead>
<tr>
<th>x</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>
Example - Probability Distributions for Discrete Random Variables

- For a two child family, what are the different possible combinations of children?
- The sample space was given as $S = \{BB, BG, GB, GG\}$ and the tree diagram is repeated on the next slide for convenience.
Example - Probability Distributions for Discrete Random Variables
Example - Probability Distributions for Discrete Random Variables

- Let $X$ represent the number of girls in the family; then the values for $X$ are $x = 0, 1, 2$. Using the classical definition of probability,
- $P(X = 0) = P(BB) = \frac{1}{4} = 0.25$
Example - Probability Distributions for Discrete Random Variables

- \( P(X = 1) = P(BG \text{ or } GB) \)
  \[ = P(BG \cup GB) \]
  \[ = P(BG) + P(GB) \quad \text{since } BG \text{ and } GB \text{ are mutually exclusive events} \]
  \[ = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = 0.5 \]

- \( P(X = 2) = P(GG) \)
  \[ = \frac{1}{4} = 0.25 \]
Example - Probability Distributions for Discrete Random Variables

We can arrange in tabular form, the values of the random variable and the associated probabilities in tabular form, as shown below.

<table>
<thead>
<tr>
<th>x</th>
<th>P(X = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[ \sum P(x) = 1 \]
Your Turn:

- A bag contains 2 black checkers and 3 red checkers.
- Two checkers are drawn without replacement from this bag and the number of red checkers is noted.
- Let $X = \text{number of red checkers drawn from this bag}$.
- Determine the probability distribution of $X$ and complete the table:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
A bag contains 2 black checkers and 3 red checkers. Two checkers are drawn without replacement from this bag and the number of red checkers is noted. Let $X =$ number of red checkers drawn from this bag.

- Possible values of $X$ are 0, 1, 2. (Why?)
- $p(x = 0) = P($black on first draw and black on second draw$) = P(B_1)P(B_2 | B_1) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$
- Now, complete the rest of the table.

**Hint:** Find $p(x = 2)$ first, since it is easier to compute than $p(x = 1)$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/10</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Continued

A bag contains 2 black checkers and 3 red checkers. Two checkers are drawn without replacement from this bag and the number of red checkers is noted. Let $X = \text{number of red checkers drawn from this bag.}$

- Possible values of $X$ are 0, 1, 2. (Why?)
- $p(x = 0) = P(\text{black on first draw and black on second draw}) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$

Hint: Find $p(x = 2)$ first, since it is easier to compute than $p(x = 1)$.

Now, complete the rest of the table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/10</td>
</tr>
<tr>
<td>1</td>
<td>6/10</td>
</tr>
<tr>
<td>2</td>
<td>3/10</td>
</tr>
</tbody>
</table>
Properties of Probability Distribution

Properties:

1. \( 0 \leq p(x_i) \leq 1 \)

2. \( \sum p(x_i) = 1 \)

The first property states that the probability distribution of a random variable \( X \) is a function which only takes on values between 0 and 1 (inclusive).

The second property states that the sum of all the individual probabilities must always equal one.
Example

\( X = \) number of customers in line waiting for a bank teller

- Verify that this describes a discrete random variable

<table>
<thead>
<tr>
<th>( x )</th>
<th>( p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Example Solution

\( X = \) number of customers in line waiting for a bank teller

- Verify that this describes a discrete random variable
- **Solution:** Variable \( X \) is discrete since its values are all whole numbers. The sum of the probabilities is one, and all probabilities are between 0 and 1 inclusive, so it satisfies the requirements for a probability distribution.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>0.32</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
</tr>
</tbody>
</table>
A probability model for a random variable consists of:

- The collection of all possible values of a random variable, and
- the probabilities that the values occur.

Of particular interest is the value we expect a random variable to take on, notated $\mu$ (for population mean) or $E(X)$ for expected value.
Discrete Random Variable
Expected Value Example

Assume $X = \text{number of heads that show when tossing three coins.}$

Sample space: HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

$X = (0, 1, 1, 1, 2, 2, 2, 3)$

If you perform this experiment many times and average the number of heads, you would expect to find a number close to

$$\frac{0 + 1 + 1 + 1 + 2 + 2 + 2 + 3}{8} = \frac{12}{8} = 1.5$$
Expected Value Example (continued)

Notice the outcomes of \( x = 1 \) and \( x = 2 \) occur three times each, while the outcomes \( x = 0 \) and \( x = 3 \) occur once each. We could calculate the average as

\[
0 + 3 \cdot 1 + 3 \cdot 2 + 3 \cdot 3 = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}
\]

\[
= 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3)
\]

\[
= 1.5
\]
Expected Value:

- The expected value of a (discrete) random variable can be found by summing the products of each possible value by the probability that it occurs, a weighted average:

\[ \mu = E(X) = \sum x \cdot P(X = x) \]

- Note: Be sure that every possible outcome is included in the sum and verify that you have a valid probability model to start with.
Discrete Random Variable: Mean

\[ \mu_X = p_1 x_1 + p_2 x_2 + p_3 x_3 + \cdots + p_n x_n \]

\[ \mu_X = \sum p_i x_i \]
The expected value of a random variable $X$ is defined as

$$E(X) = \sum x \cdot p(x)$$

How is this interpreted? If you perform an experiment thousands of times, record the value of the random variable every time, and average the values, you should get a number close to $E(X)$. 
Computing the Expected Value

Step 1. Form the probability distribution of the random variable.

Step 2. Multiply each $x$ value of the random variable by its probability of occurrence $p(x)$.

Step 3. Add the results of step 2.
Example - Expected Value for a Discrete Random Variable

- **Example:** Find the expected number of girls in a two-child family.
- **Solution:** Let $X$ represent the number of girls in a two-child family.
- Use the formula and the information from the probability distribution given on the next slide to compute the expected
Example - Expected Value for a Discrete Random Variable

- $E(X) = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$.
- That is, if we sample from a large number of two-child families, on average, there will be one girl in each family.
Your Turn:

A rock concert producer has scheduled an outdoor concert for Saturday, March 8. If it does not rain, the producer stands to make a $20,000 profit from the concert. If it does rain, the producer will be forced to cancel the concert and will lose $12,000 (rock star’s fee, advertising costs, stadium rental, etc.)

The producer has learned from the National Weather Service that the probability of rain on March 8 is 0.4.

A) Write a probability distribution that represents the producer’s profit.

B) Find and interpret the producer’s “expected profit”.
(A) There are two possibilities: It rains on March 8, or it doesn’t. Let \( x \) represent the amount of money the producer will make. So, \( x \) can either be $20,000 (if it doesn’t rain) or \( x = -$12,000 \) (if it does rain). We can construct the following table:

<table>
<thead>
<tr>
<th></th>
<th>( x )</th>
<th>( p(x) )</th>
<th>( x \cdot p(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>-12,000</td>
<td>0.4</td>
<td>-4,800</td>
</tr>
<tr>
<td>no rain</td>
<td>20,000</td>
<td>0.6</td>
<td>12,000</td>
</tr>
</tbody>
</table>

\[
E(X) = \sum x \cdot p(x) = 7,200
\]
(B) The expected value is interpreted as a long-term average. The number $7,200 means that if the producer arranged this concert many times in identical circumstances, he would be ahead by $7,200 per concert on the average. It does not mean he will make exactly $7,200 on March 8. He will either lose $12,000 or gain $20,000.
A SRS should represent the population, so the sample mean, should be somewhere near the population mean.
Law of Large Numbers

- Draw independent observations at random from any population with finite mean $\mu$.
- Decide how accurately you would like to estimate $\mu$.
- As the number of observations drawn increases, the mean $\bar{x}$-bar of the observed values eventually approaches the mean $\mu$ of the population as closely as you specified and then stays that close.
What this means:

- The law of large numbers says that the average results of many independent observations are stable and predictable (ex: casinos, grocery stores – stock, fast food restaurants).

- Both the rules of probability and the law of large numbers describe the regular behavior of chance events in the long run.

- How large is a large number?
  - That depends on the variability of the random outcomes. The more variable the outcomes, the more trails are needed to ensure the sample mean is close to the population mean.
Example

- The distribution of the heights of all young women is close to the normal distribution with mean 64.5 inches and standard deviation 2.5 inches.
- What happens if you make larger and larger samples...
- Larger sample size means less variation and the sample statistics will get closer to the population parameters 64.5 in and 2.5 in., by the LLN.
Law of Small Numbers or Law of Averages

- Most people incorrectly believe in the law of Small Numbers.
- That is, we expect short sequences of random events to show the kind of average behavior that in fact appears only in the long run.
- “Runs” of numbers, streaks, hot hand, etc.
A Fair Game

- Definition – A game of chance is called fair if the expected value is zero.
- This means that over the long run you will not win or lose playing the game (you will brake even).
Example

At a carnival a game involves spinning a wheel that is divided into 60 equal sectors. The sectors are marked as follows:

- $20 - 1 sector
- $10 - 2 sectors
- $5 - 3 sectors
- No Prize - 54 sectors

The carnival owner wants to know the average expected payout for this game and if it is a fair game.
Solution

- Define the random variable.
  
  Let $X = \text{the amount of the payout}$

- Make a probability distribution table.

<table>
<thead>
<tr>
<th>x</th>
<th>$0$</th>
<th>$5$</th>
<th>$10$</th>
<th>$20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X=x)$</td>
<td>54/60</td>
<td>3/60</td>
<td>2/60</td>
<td>1/60</td>
</tr>
</tbody>
</table>

At a carnival a game involves spinning a wheel that is divided into 60 equal sectors. The sectors are marked as follows:

- $20$ - 1 sector
- $10$ - 2 sectors
- $5$ - 3 sectors
- No Prize - 54 sectors
Find the expected value.

\[ E(x) = 0 \cdot \frac{54}{60} + 5 \cdot \frac{3}{60} + 10 \cdot \frac{2}{60} + 20 \cdot \frac{1}{60} \]

\[ E(x) = 0.92 \]

State your conclusion.

In the long run, the carnival owner can expect a mean payout of about $0.92 on each game played. This is not a fair game.
Example

Refer to the carnival game in the previous example. Suppose the cost to play the game is $1. What are a player’s expected winnings?
Solution

- Define the random variable.
  Let $X =$ the amount won by the player. (Since it cost $1 to play, we need to subtract that amount from the amount paid if you win the game)

- Make a probability distribution table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>-$1$</th>
<th>$4$</th>
<th>$9$</th>
<th>$19$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X=x)</td>
<td>54/60</td>
<td>3/60</td>
<td>2/60</td>
<td>1/60</td>
</tr>
</tbody>
</table>
Find the expected value.

\[ E(x) = -\$1(\frac{54}{60}) + \$4(\frac{3}{60}) + \$9(\frac{2}{60}) + \$19(\frac{1}{60}) \]

\[ E(x) = -\$.08 \]

State your conclusion.

In the long run, the player can expect to lose about 8 cents for every game he plays. This is not a fair game.
Clues for Clarity

- When you are asked to calculate the expected winnings for a game of chance, don’t forget to take into account the cost to play the game.
Your Turn:

A game is set up such that you have a $\frac{1}{5}$ chance of winning $350 and a $\frac{4}{5}$ chance of losing $50$. What is your expected gain?
Solution

- Let $X$ represent the amount of gain. Note, a loss will be considered as a negative gain. The probability distribution for $X$ is given below.

<table>
<thead>
<tr>
<th>$x$ (in $$)</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>350</td>
<td>$1/5$</td>
</tr>
<tr>
<td>-50</td>
<td>$4/5$</td>
</tr>
</tbody>
</table>
Solution

- Thus, the expected value of the game is 
  \[ E(X) = 350 \times \frac{1}{5} + (-50) \times \frac{4}{5} = $30. \]
- That is, if you play the game a large number of times, on average, you will win $30 per game.
Suppose you are given the option of two investment portfolios, A and B, with potential profits and the associated probabilities displayed below. Based on expected profits, which portfolio will you choose?

**Portfolio A**

<table>
<thead>
<tr>
<th>Profit</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1500</td>
<td>0.2</td>
</tr>
<tr>
<td>-100</td>
<td>0.1</td>
</tr>
<tr>
<td>500</td>
<td>0.4</td>
</tr>
<tr>
<td>1500</td>
<td>0.2</td>
</tr>
<tr>
<td>3500</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Portfolio B**

<table>
<thead>
<tr>
<th>Profit</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2500</td>
<td>0.2</td>
</tr>
<tr>
<td>-500</td>
<td>0.1</td>
</tr>
<tr>
<td>1500</td>
<td>0.3</td>
</tr>
<tr>
<td>2500</td>
<td>0.3</td>
</tr>
<tr>
<td>3500</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Solution

Let $X$ represent the profit for portfolio A, and let $Y$ represent the profit for portfolio B. Then,

$E(X) = (-1,500) \times 0.2 + (-100) \times 0.1 + 500 \times 0.4 + 1,500 \times 0.2 + 3,500 \times 0.1 = $540$

$E(Y) = (-2,500) \times 0.2 + (-500) \times 0.1 + 1,500 \times 0.3 + 2,500 \times 0.3 + 3,500 \times 0.1 = $1,000$. 
Solution

**Discussions:** Since, $E(Y) > E(X)$, you should invest in portfolio B based on the expected profit.

That is, in the long run, portfolio B will outperform portfolio A.

Thus, under repeated investments in portfolio B, you will, on average, gain $(1,000 - 540) = $460 over portfolio A.
First Center, Now Spread…

- For data, we calculated the **standard deviation** by first computing the deviation from the mean and squaring it. We do that with random variables as well.
- The **variance** for a random variable is:
  \[ \sigma^2 = \text{Var}(X) = \sum (x - \mu)^2 \cdot P(X = x) \]
- The **standard deviation** for a random variable is:
  \[ \sigma = \text{SD}(X) = \sqrt{\text{Var}(X)} \]
Random Variables: Variance

\[ \sigma_x^2 = p_1 (x_1 - \mu_x)^2 + p_2 (x_2 - \mu_x)^2 + \cdots + p_n (x_n - \mu_x)^2 \]

- Variance of a discrete random variable is a weighted (by the probability) average of the squared deviations \((x - \mu_x)^2\) of the variable \(x\) from its mean \(\mu_x\).
Variance for a Discrete Random Variable

- An equivalent computational formula for the variance is given as:

\[ V(X) = \sum \{x^2 \times P(x)\} - \mu^2 \]
Example - Variance and Standard Deviation for a Discrete Random Variable

What is the variance and standard deviation of a raffle with a first prize of $400, a second prize of $300, and a third prize of $200 if 1,000 tickets are sold?

Solution: If we let X represent the winnings, then $\mu = 0.9$ (Verify). Thus $V(X) = 0^2 \times 0.997 + 200^2 \times 0.001 + 300^2 \times 0.001 + 400^2 \times 0.001 - 0.9^2 = 289.19$. [Note: The units here will be (dollar)$^2$ since variance is a measure in square units].
Example - Variance and Standard Deviation for a Discrete Random Variable

Once again, one can also use the tabular presentation to help find the variance for a discrete random variable. We can work out the values for the different terms in the computational variance formula.

\[
V(X) = \sum \{x^2 \times P(x)\} - \mu^2
\]
Example - Variance and Standard Deviation for a Discrete Random Variable

From the table, \( V(X) = 290 - 0.9^2 = 289.19 \).

\[
SD(X) = \sqrt{V(X)}
\]

\[SD(X) = 17.01\]
The total number of cars to be sold next week is described by the following probability distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(x)</td>
<td>.05</td>
<td>.15</td>
<td>.35</td>
<td>.25</td>
<td>.20</td>
</tr>
</tbody>
</table>

Determine the expected value and standard deviation of \( X \), the number of cars sold.

\[
\mu_X = \sum_{i=1}^{5} x_i p(x_i) = 0(0.05) + 1(0.15) + 2(0.35) + 3(0.25) + 4(0.20) = 2.40
\]

\[
\sigma_X^2 = \sum_{i=1}^{5} (x_i - 2.4)^2 p(x_i) = (0 - 2.4)^2 (.05) + (1 - 2.4)^2 (.15)
\]

\[
+ (2 - 2.4)^2 (.35) + (3 - 2.4)^2 (.25) + (4 - 2.4)^2 (.20) = 1.24
\]

\[
\sigma_X = \sqrt{1.24} = 1.11
\]
More About Means and Variances

- Adding or subtracting a constant from data shifts the mean but doesn’t change the variance or standard deviation. The same is true for random variables.

\[ E(X \pm c) = E(X) \pm c \quad \text{Var}(X \pm c) = \text{Var}(X) \]
Example

- Couples dining at the *Quiet Nook* restaurant can expect Lucky Lovers discounts averaging $5.83 with a standard deviation of $8.62. Suppose that for several weeks the restaurant has been distributing coupons worth $5 off any one meal (one discount per table).

- If every couple dining there on Valentine’s Day brings a coupon, what will be the mean and standard deviation of the total discounts they’ll receive?
Solution: Couples dining at the *Quiet Nook* restaurant can expect Lucky Lovers discounts averaging $5.83 with a standard deviation of $8.62. Suppose that for several weeks the restaurant has been distributing coupons worth $5 off any one meal (one discount per table). If every couple dining there on Valentine’s Day brings a coupon, what will be the mean and standard deviation of the total discounts they’ll receive?

- Random variable $X = $Lucky Lovers Discount, then $E(X) = 5.83$ and $SD(X) = 8.62$ ($Var(X) = (8.62)^2$).
- Let the random variable $D = $total discount (lucky lovers plus the coupon), then $D = X+5$.
- $E(D)$
  - $E(D) = E(X+5) = E(X) + 5 = 5.83 + 5 = $10.83$
- $SD(D)$
  - $Var(D) = Var(X+5) = Var(X) = (8.62)^2$
  - $SD(D) = \sqrt{Var(D)} = \sqrt{(8.62)^2} = $8.62
- Couples with the coupon can expect total discounts averaging $10.83 and the standard deviation is still $8.62$ (*Adding or subtracting a constant from a random variable, adds or subtracts the constant from the mean of the random variable, but doesn’t change the variance or standard deviation.*).
More About Means and Variances

- Multiplying each value of a random variable by a constant multiplies the mean by that constant and the variance by the square of the constant.

\[ E(aX) = aE(X) \quad \text{Var}(aX) = a^2\text{Var}(X) \]
Example

On Valentine’s Day at the *Quiet Nook*, couples get a lucky lovers discount averaging $5.83 with a standard deviation of $8.62. When two couples dine together on a single check, the restaurant doubles the discount.

What are the mean and standard deviation of discounts for such foursomes?
Solution: On Valentine’s Day at the *Quiet Nook*, couples get a lucky lovers discount averaging $5.83 with a standard deviation of $8.62. When two couples dine together on a single check, the restaurant doubles the discount. What are the mean and standard deviation of discounts for such foursomes?

- Random variable $X = \text{Lucky Lovers Discount}$, then $E(X) = 5.83$ and $SD(X) = 8.62$ ($Var(X) = (8.62)^2$).
- Let the random variable $D = \text{total discount (double lucky lovers)}$, then $D = 2X$.
- $\text{E}(D)$
  - $\text{E}(D) = \text{E}(2X) = 2\text{E}(X) = 2(5.83) = $11.66$
- $\text{SD}(D)$
  - $\text{Var}(D) = \text{Var}(2X) = 2^2\text{Var}(X) = 2^2(8.62)^2 = 297.2176$
  - $\text{SD}(D) = \sqrt{\text{Var}(D)} = \sqrt{297.2176} = $17.24$
- Two couples dining together can expect to save an average of $11.66 with a standard deviation of $17.24 (Multiplying each value of a random variable by a constant multiplies the mean by that constant and the variance by the square of the constant.).
More About Means and Variances

In general,

- The mean of the sum of two random variables is the sum of the means.
  \[ E(X \pm Y) = E(X) \pm E(Y) \]
- The mean of the difference of two random variables is the difference of the means.
- If the random variables are independent, the variance of their sum or difference is always the sum of the variances.
  \[ \text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \]
Example

- Company A believes that the sales of product X is as follows.

<table>
<thead>
<tr>
<th>X</th>
<th>1000</th>
<th>3000</th>
<th>5000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>.1</td>
<td>.3</td>
<td>.4</td>
<td>.2</td>
</tr>
</tbody>
</table>

\[
\mu_X = 1000 \cdot .1 + 3000 \cdot .3 + 5000 \cdot .4 + 10000 \cdot .2
\]

\[
\mu_X = 5000 \text{ units}
\]

The expected sales of product X, E(X), is 5000 units.
Example

Also, Company A believes that the sales of product Y is as follows.

<table>
<thead>
<tr>
<th>Y</th>
<th>300</th>
<th>500</th>
<th>750</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Y)</td>
<td>.4</td>
<td>.5</td>
<td>.1</td>
</tr>
</tbody>
</table>

\[ \mu_Y = 300 \cdot .4 + 500 \cdot .5 + 750 \cdot .1 \]

\[ \mu_Y = 445 \text{ units} \]

The expected sales of product Y, E(Y), is 445 units.
Example

- The expected sales of product X, \( E(X) \), is 5000 units.
- The expected sales of product Y, \( E(Y) \), is 445 units.
- What are the expected sales for both products combined?
- Let the random variable \( T = \text{total product sales} \), then \( T = X + Y \).
  - \( E(T) = E(X + Y) = E(X) + E(Y) = 5000 + 445 = 5445 \)
- The expected sales for both products X and Y combined is 5445 units (The mean of the sum/diff. of two random variables is the sum/diff. of the means.).
Example

- The standard deviation of product X is 2793 units and product Y is 139 units, verify using your calculator.

<table>
<thead>
<tr>
<th>X</th>
<th>1000</th>
<th>3000</th>
<th>5000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>.1</td>
<td>.3</td>
<td>.4</td>
<td>.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>300</th>
<th>500</th>
<th>750</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Y)</td>
<td>.4</td>
<td>.5</td>
<td>.1</td>
</tr>
</tbody>
</table>

- What is the difference between the standard deviations of products X and Y?
Solution: What is the difference between the standard deviations of products X and Y?

- SD(X) = 2793 and SD(Y) = 139
- Var(X) = 2793^2 = 7800849
- Var(Y) = 139^2 = 19321
- Let the random variable D = difference in product sales, then D = X - Y.
  - Var(D) = Var(X - Y) = Var(X) + Var(Y) = 7800849 + 19321 = 7820170
  - SD(D) = \sqrt{Var(D)} = \sqrt{7820170} = 2796
- The difference between the standard deviations of products X and Y is 2796 units (If the random variables are independent, the variance of their sum or difference is always the sum of the variances.).
Continuous Random Variables

- Random variables that can take on any value in a range of values are called continuous random variables.

- Now, any single value won’t have a probability, but...

- Continuous random variables have means (expected values) and variances.

- We won’t worry about how to calculate these means and variances in this course, but we can still work with models for continuous random variables when we’re given the parameters.
Continuous Random Variables

- Good news: nearly everything we’ve said about how discrete random variables behave is true of continuous random variables, as well.
- When two independent continuous random variables have Normal models, so does their sum or difference.
- This fact will let us apply our knowledge of Normal probabilities to questions about the sum or difference of independent random variables.
A continuous random variable $X$ takes all values in an interval of numbers.
Distribution of Continuous Random Variable

- The **probability distribution** of $X$ is described by a smooth curve.
- The probability of any event is the area under the curve and above the values of $X$ that make up that event.
- All continuous probability distributions assign probability 0 to every individual outcome.
Example - Probability Distribution for a Continuous Random Variable

The shaded area under the density curve shows the proportion, or percent, of individuals in the population with values of \( X \) between \( x_1 \) and \( x_2 \).

Because the probability of drawing one individual at random depends on the frequency of this type of individual in the population, the probability is also the shaded area under the curve.
Normal distributions as probability distributions

- Suppose the continuous random variable $X$ has $N(\mu, \sigma)$ then we can use our normal distribution tools to calculate probabilities.
A variable whose value is a number resulting from a random process is a **random variable**. The probability distribution of many random variables is the **normal distribution**. It shows what values the random variable can take and is used to assign probabilities to those values.

Example: Probability distribution of women’s heights.

Here, since we chose a woman randomly, her height, \(X\), is a random variable.

To calculate probabilities with the normal distribution, we will standardize the random variable (\(z\)-score) and use the Z Table.
Reminder: standardizing $N(\mu, \sigma)$

We standardize normal data by calculating $z$-scores so that any Normal curve $N(\mu, \sigma)$ can be transformed into the standard Normal curve $N(0,1)$.

$$z = \frac{(y - \mu)}{\sigma}$$
Previously, we wanted to calculate the proportion of individuals in the population with a given characteristic.

Distribution of women’s heights \( \approx N(\mu, \sigma) = N(64.5, 2.5) \)

Example: What's the proportion of women with a height between 57" and 72"?

That's within \( \pm 3 \) standard deviations \( \sigma \) of the mean \( \mu \), thus that proportion is roughly 99.7%.

Since about 99.7% of all women have heights between 57" and 72", the chance of picking one woman at random with a height in that range is also about 99.7%.
**Example:** What is the probability, if we pick one woman at random, that her height will be some value \( X \)? For instance, between 68” and 70” \( P(68 < X < 70) \)?

Because the woman is selected at random, \( X \) is a random variable.

\[
z = \frac{x - \mu}{\sigma}
\]

As before, we calculate the \( z \)-scores for 68 and 70.

For \( x = 68" \), \( z = \frac{(68 - 64.5)}{2.5} = 1.4 \)

For \( x = 70" \), \( z = \frac{(70 - 64.5)}{2.5} = 2.2 \)

The area under the curve for the interval [68” to 70”] is 0.9861 − 0.9192 = 0.0669. Thus, the probability that a randomly chosen woman falls into this range is 6.69%.

\[ P(68 < X < 70) = 6.69\% \]
Example: Inverse problem

Your favorite chocolate bar is dark chocolate with whole hazelnuts. The weight on the wrapping indicates 8 oz. Whole hazelnuts vary in weight, so how can they guarantee you 8 oz. of your favorite treat? You are a bit skeptical...

To avoid customer complaints and lawsuits, the manufacturer makes sure that 98% of all chocolate bars weight 8 oz. or more.

The manufacturing process is roughly normal and has a known variability $\sigma = 0.2$ oz.

How should they calibrate the machines to produce bars with a mean $\mu$ such that $P(x < 8 \text{ oz.}) = 2\%$?
How should they calibrate the machines to produce bars with a mean \( m \) such that \( P(x < 8 \text{ oz.}) = 2\% \)?

Here, we know the area under the density curve (2\% = 0.02) and we know \( x \) (8 oz.).

We want \( \mu \).

In Table A we find that the \( z \) for a left area of 0.02 is roughly \( z = -2.05 \).

\[
\begin{align*}
z &= \frac{(x - \mu)}{\sigma} \quad \Leftrightarrow \quad \mu &= x - (z \times \sigma) \\
\mu &= 8 - (-2.05 \times 0.2) = 8.41 \text{ oz}
\end{align*}
\]

Thus, your favorite chocolate bar weighs, on average, 8.41 oz. Excellent!!!
What Can Go Wrong?

- Probability models are still just models.
  - Models can be useful, but they are not reality.
  - Question probabilities as you would data, and think about the assumptions behind your models.
- If the model is wrong, so is everything else.
What Can Go Wrong? (cont.)

- Don’t assume everything’s Normal.
- Watch out for variables that aren’t independent:
  - You can add expected values for \textit{any} two random variables, but
  - you can only add variances of \textit{independent} random variables.
What Can Go Wrong?

- Don’t forget: Variances of independent random variables add. Standard deviations don’t.
- Don’t forget: Variances of independent random variables add, even when you’re looking at the difference between them.
- Don’t forget: Don’t write independent instances of a random variable with notation that looks like they are the same variables.
What have we learned?

- We know how to work with random variables.
  - We can use a probability model for a discrete random variable to find its expected value and standard deviation.
- The mean of the sum or difference of two random variables, discrete or continuous, is just the sum or difference of their means.
- And, for independent random variables, the variance of their sum or difference is always the sum of their variances.
What have we learned?

- Normal models are once again special.
  - Sums or differences of Normally distributed random variables also follow Normal models.
Assignment

- Pg. 383 – 387:#1, 3, 5, 15, 19, 21, 25, 27, 33, 37
- Read Ch-17 pg. 388 - 400