3.1 Random Variables and Histograms

A random variable is a rule that assigns a number to each outcome of an experiment. We usually denote a random variable by $X$.

There are 3 types of random variables: finite discrete, infinite discrete, and continuous. We will define each of them below.

We can find probability distributions of random variables the same way we did with experiments. List all the values that are possible for $X$ and calculate their probabilities using any method we know.

Example: A coin is tossed three times and the sequence of heads and tails is observed.

1. List the outcomes of the experiment.

2. Let the random variable $X$ denote the number of tails that occur. What are the possible values of $X$?

3. Find the probability distribution of $X$.

4. What is $P(1 \leq X \leq 2)$?

5. What is $P(X > 0)$?

The above random variable is called finite discrete because it takes on only finitely many values.
Consider the experiment of rolling 2 fair 5-sided dice. We know that the sample space of this experiment is

\[ S = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (3,4), (3,5), (4,1), (4,2), (4,3), (4,4), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5) \} \]

Let \( X \) be the sum of the numbers rolled.

1. What are the possible values that \( X \) can take on?

Since \( X \) can only be a finite number of values, \( X \) is finite discrete.

2. Find the probability distribution of \( X \).

Example: A carton contains a dozen eggs, of which 5 are cracked. An experiment consists of randomly selecting 3 eggs from the carton. Let \( X \) be the number of cracked eggs that are selected.

1. Find the probability distribution for \( X \).

2. What is the probability that more than 1 cracked egg will be selected?
Example: An unfair die has the property that the probability of rolling a 1 is 0.3. This die is rolled 4 times. Let \( X \) be the number of times a 1 is rolled. Find the probability distribution for \( X \).

Example: A survey was conducted of families to determine the distribution of families by size. The results are:

<table>
<thead>
<tr>
<th>Family Size</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Families</td>
<td>29</td>
<td>16</td>
<td>24</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>

Let the random variable \( X \) be the number of people in a randomly chosen family. Find the probability distribution for \( X \).

Example: A die is rolled. Let the random variable \( X \) denote the number of times the die is rolled until a 5 appears. What are the values that \( X \) may assume?

This random variable is called infinite discrete because there are an infinite number of possible values and they can be arranged in a sequence of separated numbers. (We can list them all in a certain order even though there are infinitely many of them.)

Example: Suppose I have a random variable \( X = \) The distance a person travels to work in miles. What values may \( X \) take on?

This random variable is called continuous because there are infinitely many values that make up an interval of numbers. Continuous random variables often deal with time, distance, and measurements.
Examples: List the possible values of $X$ and determine whether the random variable is finite discrete, infinite discrete, or continuous.

1. A marble is pulled from a jar with 3 red, 2 blue, and 3 green marbles without replacement until a green marble is drawn. Let $X$ be the number of number of pulls needed.

2. A marble is pulled from a jar with 3 red, 2 blue, and 3 green marbles with replacement until a blue marble is drawn. Let $X$ be the number of number of pulls needed.

3. Four marbles are pulled one at a time with replacement from a jar with 3 red, 2 blue, and 3 green marbles. Let $X$ be the number of times a blue marble is pulled.

4. Same experiment as above but without replacement.

5. Let $X$ be the amount of time in minutes a student works on homework on a given day.

6. Let $X$ be the number of times a 3 is rolled in 5 rolls of a die.
A histogram is a graphical way of displaying the probability distribution of a random variable.

Plot the values of the random variable along the $x$-axis. Then draw a rectangle of width 1 centered above each value with height equal to the probability of the random variable attaining this value. (So, each rectangle has area equal to the probability.)

Remember that probability distribution values must add up to 1, which means the heights of all the rectangles (and also their areas) must add up to 1.

Draw a histogram for the probability distribution below. Shade the area representing $P(0 < X \leq 3)$ and calculate its value.

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.15</td>
<td>0.3</td>
<td>0.1</td>
<td>0.05</td>
<td>0.2</td>
</tr>
</tbody>
</table>

3.2 Measures of Central Tendency

The average, or mean, of a set of numbers $x_1, x_2, \ldots, x_n$, denoted by $\overline{x}$, or $\mu$ is

$$
\mu = \overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}
$$

The median of a set of numbers is the middle number when they are arranged in increasing order. If there are an even number of entries, there is no middle number, so we let the median be the average of the middle two numbers.

The mode of a set of numbers is the number that occurs most frequently.

What are the mean, median, and mode of the set of numbers 1, 4, 4, 5, 9, 9, 9?

What are the mean, median, and mode of the set of numbers 1, 1, 4, 5, 5, 9?
You can find mean and median on your calculator as follows: (Note: The mode is not given on the calculator.)

1. Press STAT, 1:Edit
2. Enter in the list of values under L1.
3. Press STAT, cursor right to CALC, and then press 1: 1-Var Stats
4. Then, you need to add L1, (2nd-1), so that the home screen reads “1-Var Stats L1”.
5. Press enter. \( \bar{x} \) is the mean or average. Scroll down to find the median, which is Med.

Example: The daily average temperatures were recorded for the city of College Station during a span of two weeks in the month of January. The results were: (I made these up.)

\[
70, 67, 75, 77, 69, 68, 68, 75, 68, 42, 55, 65, 69
\]

Find the mean, median, and mode for this data.

If data values are listed with frequencies, put the data values under L1 and their frequencies under L2.
Then, type “1-Var Stats L1, L2” and press enter. By default, if you just type “1-Var Stats,” then the calculator assumes you mean “1-Var Stats L1.”

Example: A certain puzzle manufacturer makes and sells puzzles. The data below shows how many puzzles this company makes with different numbers of pieces.

<table>
<thead>
<tr>
<th>Number of Pieces</th>
<th>500</th>
<th>1000</th>
<th>750</th>
<th>2000</th>
<th>300</th>
<th>1500</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Puzzles</td>
<td>400</td>
<td>200</td>
<td>350</td>
<td>150</td>
<td>275</td>
<td>300</td>
<td>75</td>
</tr>
</tbody>
</table>

What is the average number of pieces in a puzzle made by this company? What are the median and mode for the number of pieces in a puzzle?

In the language of probability, we use the term **expected value** instead of average or mean, although the concepts are fairly similar. With expected value, we are essentially finding the “weighted average” of a random variable \( X \) where each value is weighted by its probability.

The expected value of a random variable \( X \) can be thought of graphically as the value along the \( x \)-axis where the histogram of \( X \) would be balanced. The mode is the \( x \)-value that has the tallest rectangle. (The median is not as intuitive visually in a histogram.)
The expected value of a random variable $X$, denoted $E(X)$, is given by

$$E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n$$

where $x_1, x_2, \ldots, x_n$ are the values that $X$ may assume, and $p_1, p_2, \ldots, p_n$ are the probabilities of each of these values.

You can find expected value on the calculator by using the exact same method above where the values of $X$ are typed in L1 and their probabilities are typed in L2.

Example: Find the expected value (by hand and using the calculator) of a random variable $X$ having the following probability distribution. Also find the mode(s).

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{3}{16}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{16}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>

Two cards are selected at random from a standard deck of cards. What is the expected number of hearts?

Expected values are often used in games to determine whether the game is “fair.” In a game situation, we let $X$ be the net winnings (profit) of the player. A game is considered “fair” when the expected net winnings are 0, i.e., when $E(X) = 0$.

Ben is playing a game at a carnival. The game costs $1. Ben selects a card from a standard deck. If the card is an ace, then he wins $3. If the card is a face card, he wins $2. Find the expected net winnings.
Example: A raffle is held. 2000 people buy a ticket for $3 each. First prize is $1500. Second prize is $750. Then, four $100 consolation prizes will be given. What are the expected net winnings for a person who buys a ticket?

A game consists of rolling a pair of dice. The game costs $4 to play. If you roll the same number on both dice (a double), you win $a. Otherwise, you get nothing. What value of a would make this game fair?

Expected value can often be used in insurance.

Example: A man purchased a $25,000 life insurance policy from his employer for $200/yr. The probability that he lives another year is 0.9935. What is the life insurance company’s expected gain?

If the probability that the man lives drops to 0.98, what is the minimum amount of money, $a, he can expect to pay for his policy?
3.3 Measures of Spread

The mean, or expected value, of a random variable measured the central tendency of the variable (where the histogram would be balanced). However, this gives us no indication of how spread out the distribution is.

The variance of a random variable is a measure of how spread out the distribution is. If the variance is large, then the distribution has a larger spread about the mean. If the variance is small, then the distribution is more clustered around the mean.

Another measure of spread is the standard deviation of a random variable. Again, the larger the standard deviation, the more spread out the distribution is from the mean.

\[ \sigma = \sqrt{Var(X)} \] which means that \[ \sigma^2 = Var(X) \]

Standard deviation has the same units as the values of the random variable, whereas variance would have these units squared. (So if \( X \) is in cm, then standard deviation is in cm, whereas variance would be in cm\(^2\).)

Standard deviation is also found using 1-Var Stats. Standard deviation is the 5th entry down, \( \sigma x \). Then, just square the standard deviation to find variance.

Example: Find the variance and standard deviation of the random variable \( X \), whose probability distribution is given.

<table>
<thead>
<tr>
<th>( X )</th>
<th>-4</th>
<th>-2</th>
<th>0</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Example: The number of jelly beans in 150 bags of jelly beans were counted. The results are given below.

<table>
<thead>
<tr>
<th>Number of Bags</th>
<th>30</th>
<th>40</th>
<th>65</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Jelly Beans</td>
<td>25</td>
<td>32</td>
<td>28</td>
<td>34</td>
</tr>
</tbody>
</table>

Find the mean, standard deviation, and variance for the number of jelly beans in a bag.
Binomial Statistics
For a binomial distribution, we can use the following formulas to find the expected value (mean), standard deviation, and variance.

\[
\mu = E(X) = np
\]
\[
\sigma = \sqrt{npq}
\]
\[
Var(X) = \sigma^2 = npq
\]

Reminder: \( q \) is the probability of a failure: \( q = 1 - p \).

Example: At a certain university the probability that an entering freshman will graduate within 4 yr is 0.6. From an incoming class of 2000 freshmen, find

- The expected number of students who will graduate within 4 yr.
- The standard deviation and variance of the number of students who will graduate within 4 yr.