

Lab Module 4: Discrete Random Variables, Expected Value and Variance

Learning Objectives

- Produce an example of a discrete random variable on a finite sample space.
- Depict the probability distribution function (PDF) and cumulative distribution function (CDF) of a discrete random variable and compute its expected value and variance.
- Evaluate the expected profitability of a monetary bet or wager.
- Recognize the characteristic features of binomial and geometric random variables and compute their PDF, expected value and variance.
- Invoke the Law of Large Numbers to estimate the expected value of a discrete random variable from simulated data.

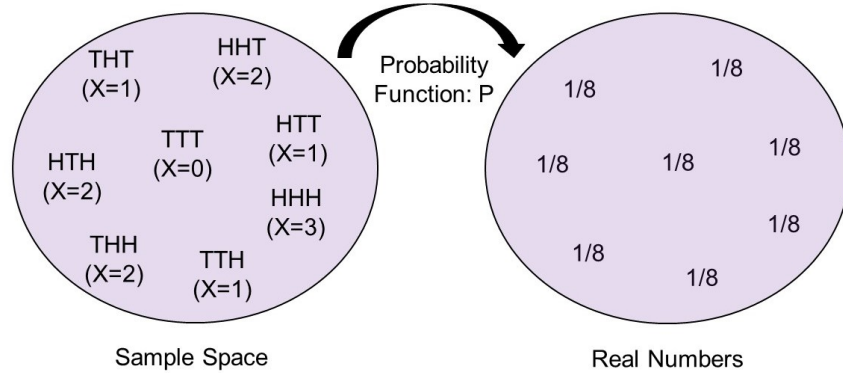
Key Concepts

Discrete Random Variables

Consider the random experiment of a three coin toss and the total number of heads that appear. The sample space consists of eight possible outcomes: HHH, HHT, HTH, THH, HTT, THT, TTH and TTT whereas there are just four possibilities for the number of heads: zero heads, one head, two heads or three heads.

A discrete random variable (X) defines a *numerical* outcome for each sample point on a discrete sample space. Let the discrete random variable X represent the number of heads resulting from a three coin toss. Each sample point in the sample space is associated with a value of the discrete random variable X as depicted below:

Random Variable ($X = \text{Number of Heads}$) on the Sample Space for a Three Coin Toss

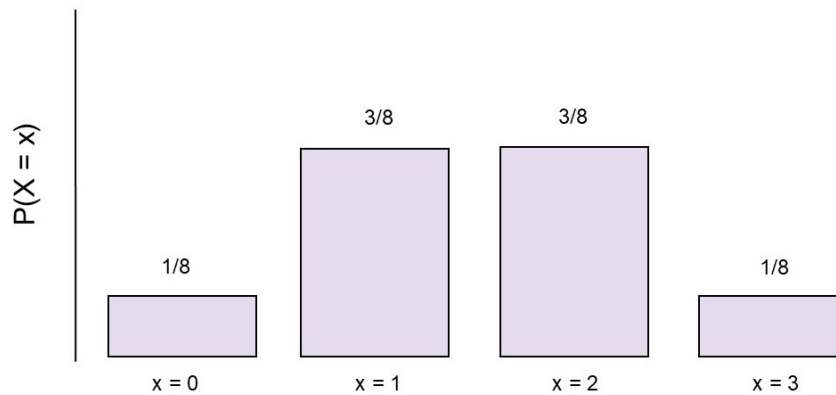


The eight equally likely random outcomes (sample points) of a three coin toss each map to a probability of $\frac{1}{8}$. The four possible values of the discrete random variable (X) are: $X = 0$, $X = 1$, $X = 2$ and $X = 3$ heads. Each value of X defines an event on the sample space. To compute the probabilities $P(X = x)$ for all x , apply the Classical Method. What is the probability that $X = 0$? There is one sample point (TTT) associated with $X = 0$, indicating that $P(X = 0) = \frac{1}{8}$. What is the probability that $X = 1$? There are three sample points (HTT, THT, TTH) associated with $X = 1$, indicating that $P(X = 1) = \frac{3}{8}$.

Probability Distribution Function

A discrete probability distribution function (PDF) is a graphical representation of the probabilities associated with the numerical outcomes (x) of a discrete random variable. The discrete PDF of X is depicted below:

Discrete PDF of X : Number of Heads Resulting from a Three Coin Toss

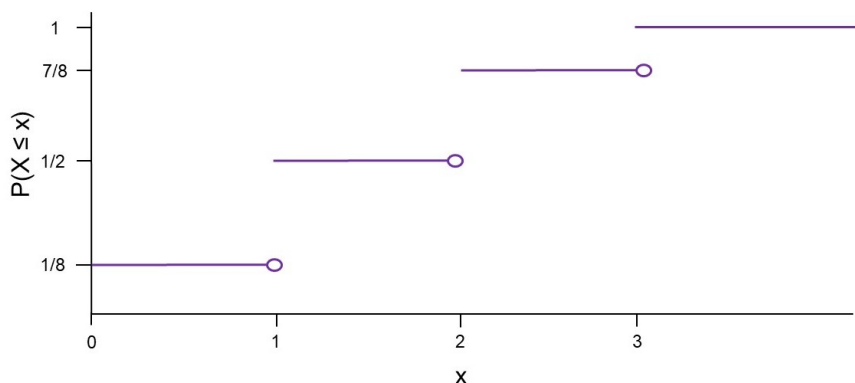


A discrete PDF has a similar appearance to a histogram. In a discrete PDF, the probabilities of the numerical outcomes of a random variable are depicted on the y-axis instead of the observed frequencies in a data set.

Cumulative Distribution Function

A cumulative probability distribution function (CDF) is a graph of $P(X \leq x)$ versus x . The CDF of X is depicted below:

Discrete CDF of X : Number of Heads Resulting from a Three Coin Toss



Since X is a discrete random variable, the CDF of X is a discontinuous function of x .

Expected Value and Variance

The outcome of a random experiment cannot be predicted with certainty. For a three coin toss, the eight random outcomes depicted in the sample space are equally likely to occur. Nevertheless, if a three coin toss is repeated many times, the average number of heads will converge on a single value. The expected value, $E(X)$, of a discrete random variable X is defined as follows:

$$E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

For a three coin toss, $E(X)$ equals:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

Notice that it is impossible to observe 1.5 heads on a single three coin toss. The expected value of a discrete random variable may not be included in its range!

The variance, $\text{Var}(X)$, of a discrete random variable X is defined as follows:

$$\text{Var}(X) = \sum_{i=1}^n (x_i - E(X))^2 P(X = x_i)$$

For a three coin toss, $\text{Var}(X)$ equals:

$$\text{Var}(X) = (0 - 1.5)^2 \times \frac{1}{8} + (1 - 1.5)^2 \times \frac{3}{8} + (2 - 1.5)^2 \times \frac{3}{8} + (3 - 1.5)^2 \times \frac{1}{8} = 0.75$$

An equivalent definition for variance is: $\text{Var}(X) = E(X^2) - E(X)^2$ which produces the identical result:

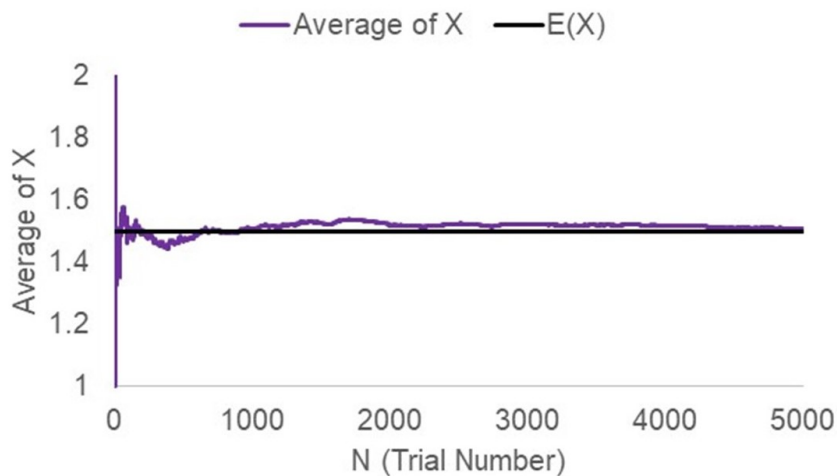
$$\text{Var}(X) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} - 1.5^2 = 0.75$$

The latter calculation is often computationally simpler.

The Law of Large Numbers and Expected Value

The Law of Large Numbers guarantees that if you repeat a random experiment many times and compute the average of a discrete random variable (X), the average will come very close to $E(X)$, the expected value of X . The Law of Large Numbers guarantees that the average number of heads over repeated three coin tosses will eventually come very close to 1.5. In the computer simulation depicted below, a three coin toss is repeated over N trials, the average number of heads for each N is recorded, and this average is plotted as a function of N trials.

Average Number of Heads as a Function of N Repeated Trials of a Three Coin Toss



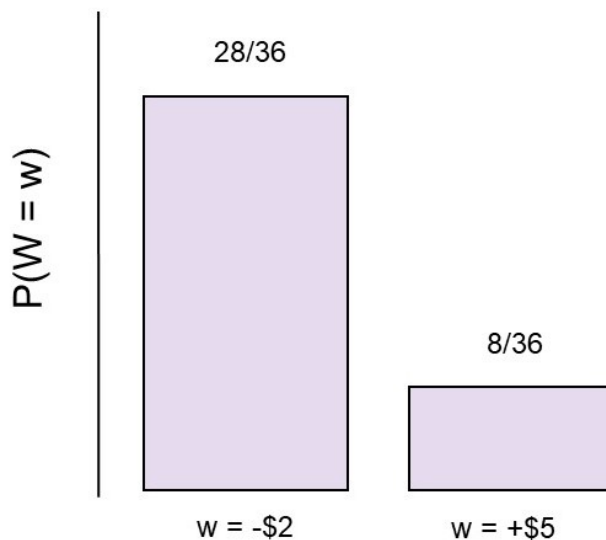
As the number of trials increases, the average number of heads gets very close to the expected value of 1.5.

Expected Value of a Bet

Suppose your friend offers you the following bet: He will pay you \$5 if on a single roll of two six-sided dice, the sum of the two dice equals 7 or 11. You will pay him \$2 if the sum of the two dice does not equal 7 or 11. Since a roll of two six-sided dice is a random experiment, the outcome of a single bet cannot be predicted with certainty. However, by computing the expected winnings, you can determine whether taking this bet would be to your advantage "in the long run", if you were to accept the bet many times.

Let the discrete random variable W represent your winnings for a single roll of two six-sided dice. Since the probability of rolling a sum of 7 or 11 (and winning \$5) equals $\frac{8}{36}$, the discrete PDF of W can be depicted as follows:

Discrete PDF of W : Winnings for the Outcome of a Bet



The expected value of W ($E(W)$) equals:

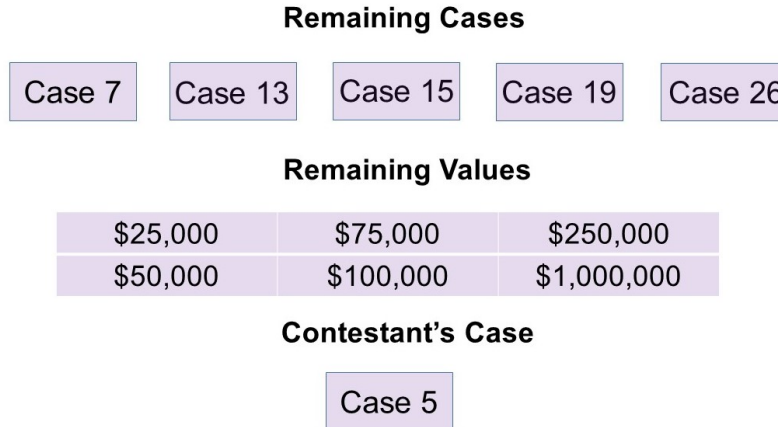
$$E(X) = -\$2 \times \frac{28}{36} + \$5 \times \frac{8}{36} = -\$ \frac{16}{36} = -\$0.44$$

Since $E(W)$ is negative, this bet is not profitable in the long run!

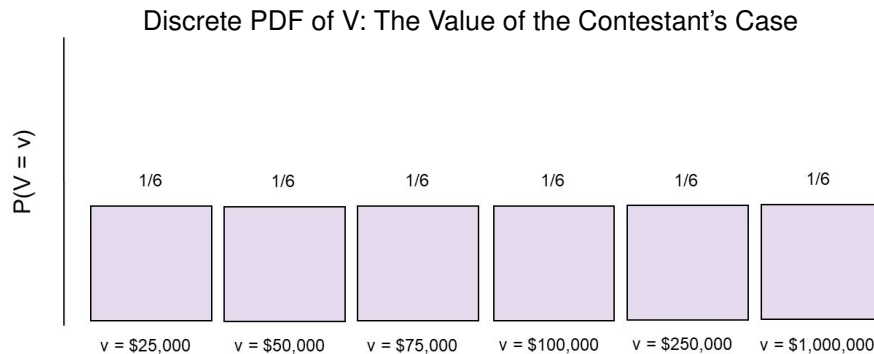
"Deal or No Deal"

In the game show, "Deal or No Deal", a contestant selects a single case from a group of 26 cases. The selected case may contain as much as \$1,000,000 or as little as \$0.01. The monetary values of all 26 cases are collectively known, but the value assigned to each individual case is unknown. The game proceeds as the contestant selects from the remaining cases and reveals their monetary value. At various times, the Bank will offer the contestant an amount of money to purchase the originally selected case.

How much is the contestant's case worth? The following graphic represents the monetary values for a set of five remaining cases in addition to the contestant's case:



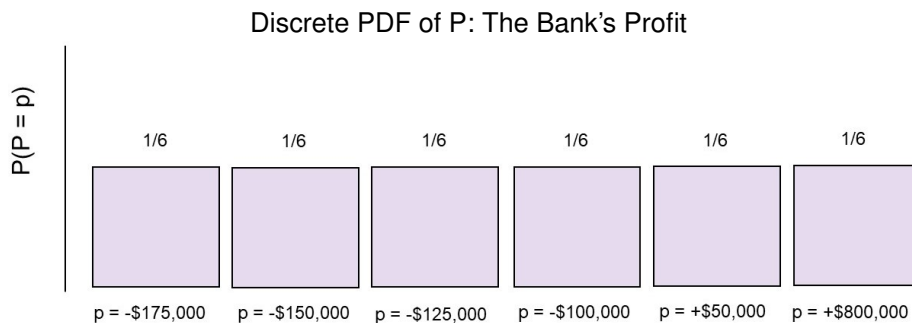
Let the discrete random variable V represent the value of the contestant's case. There are six possible monetary values for the contestant's case so that the probability of each possible value equals $\frac{1}{6}$. The discrete PDF of V is depicted below:



The expected value of the contestant's case is also its fair value. The expected value of V equals \$250,000:

$$E(V) = \$25,000 \times \frac{1}{6} + \$50,000 \times \frac{1}{6} + \$75,000 \times \frac{1}{6} + \$100,000 \times \frac{1}{6} + \$250,000 \times \frac{1}{6} + \$1,000,000 \times \frac{1}{6} = \$250,000$$

How much should the Bank offer to purchase the contestant's case? Let the discrete random variable P represent the bank's profit (P) whereby $P = \text{Value of the Case } (V) - \text{Offer Amount}$. Suppose the contestant accepts an offer of \$200,000. The discrete PDF of the bank's profit (P) is depicted below:



The expected value of P equals +\$50,000:

$$E(P) = (-\$175,000) \times \frac{1}{6} + (-\$150,000) \times \frac{1}{6} + (-\$125,000) \times \frac{1}{6} + (-\$100,000) \times \frac{1}{6} + (+\$50,000) \times \frac{1}{6} + (+\$800,000) \times \frac{1}{6} = +\$50,000$$

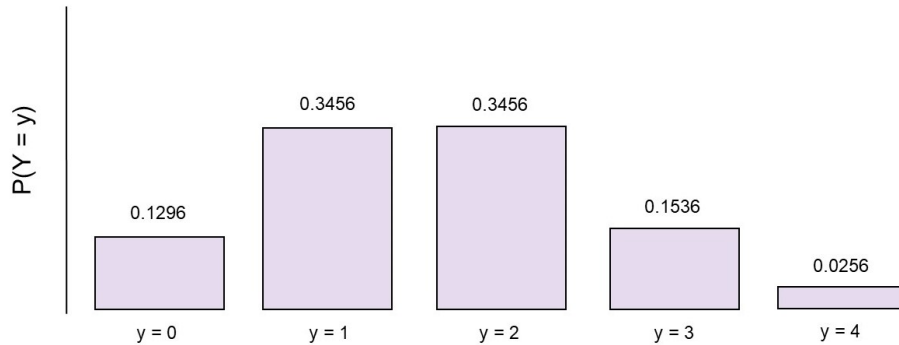
The bank expects a profit of +\$50,000 for this particular offer scenario. The Law of Large Numbers does not guarantee that the bank will profit on any individual offer, but can guarantee that the bank will average a +\$50,000 profit over many identical offer scenarios.

The Law of Large Numbers gives the bank a distinct advantage over the contestant. The contestant has just one chance to win a large sum of money, and may be inclined to accept an offer substantially lower than the fair value. In contrast, the bank can take a long-term view and absorb losses on a handful of bets while still achieving an overall net profit.

Binomial Random Variable

Winston is a labradoodle who loves to play fetch. He catches his ball mid-air with probability 0.4. What is the probability that Winston makes exactly one catch in a series

PDF of Y: The Number of Times Winston Catches the Ball Mid-Air in a Series of Four Tosses



The number of times Winston catches the ball mid-air out of four tosses is a binomial random variable. A random variable Y is binomial if the following five criteria are satisfied:

1. Y is defined as the number of "successes" occurring in N trials.
2. The number of identical trials (N) is a fixed constant.
3. There are two possible outcomes for each trial ("success" and "failure").
4. The "success" probability (p) is constant.
5. The N trials are statistically independent.

The PDF of a binomial random variable Y can be computed using the following general formula:

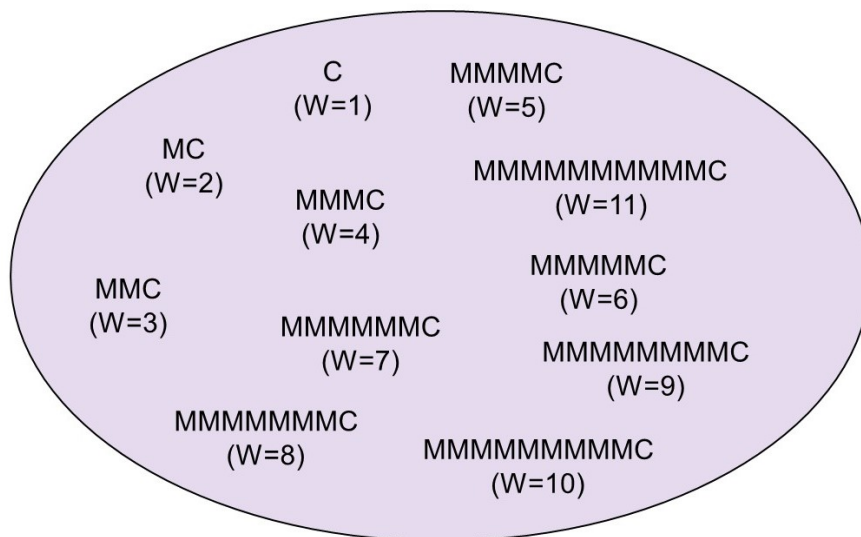
$$P(Y=y) = \binom{N}{y} p^y (1 - p)^{N-y}$$

In the above example, N equals the number of tosses (N = 4), and the constant p equals the probability of a "catch" (p = 0.4). Furthermore, the expected value of a binomial random variable Y equals $N \times p$. Winston is expected to make 1.6 mid-air catches in a series of four attempts. The variance of a binomial random variable Y equals $N \times p \times (1 - p)$.

Geometric Random Variable

What is the probability that it takes Winston just one attempt to make the first catch? Two attempts? Five attempts? Let the discrete random variable W represent the number of catch attempts that Winston makes before he catches the first ball mid-air. The sample space below depicts some possible outcomes and their association with the discrete random variable W . In theory, the sample space for W is infinite since there is no guarantee that Winston will catch the ball on any given attempt.

Sample Space W : Total Number of Catch Attempts at Winston's First Mid-Air Catch



"C" indicates a mid-air catch and "M" indicates a miss.

The number of attempts Winston makes before catching the ball mid-air for the first time is a geometric random variable. A random variable W is geometric if the following criteria are satisfied:

1. W is defined as the total number of trials at the first "success" occurrence.
2. There are two possible outcomes for each trial ("success" and "failure").
3. The "success" probability (p) is constant.
4. The trials are statistically independent.

How are probabilities assigned to the random outcomes (sample points) of the ball toss? Assume the tosses are statistically independent of one another. The sample point C ($W = 1$) has a probability of $P(C) = 0.4$. The sample point MC ($W = 2$) has a probability of $P(M) \times P(C) = 0.6 \times 0.4 = 0.24$, and so on. The PDF of a geometric random variable W can be computed using the following general formula:

$$P(W=w) = p \times (1 - p)^{w-1}.$$

Since the sample space for W is infinite, there is no upper limit for the random variable W . Nonetheless, the infinite sum of geometric probabilities equals one, as required by the Axioms of Probability, due to the convergence properties of a geometric series:

$$\sum_{w=1}^{\infty} P(W = w) = \sum_{w=1}^{\infty} p \times (1 - p)^{w-1} = \frac{p}{(1-p)} \sum_{w=1}^{\infty} (1 - p)^w = \frac{p}{(1-p)} \times \frac{(1-p)^1}{1-(1-p)} = 1$$

The expected value of a geometric random variable W equals $\frac{1}{p}$. On average, Winston is expected to make the first mid-air catch after $\frac{1}{0.4} = 2.5$ catch attempts. The variance of a geometric random variable Y equals $\frac{1-p}{p^2}$.

Assignment - Graphs & Calculations

Question 1: Deal or No Deal

Watch the Lab 4 video tutorial entitled "**Discrete Random Variables**".

Watch the Lab 4 video tutorial entitled "**Expected Value**".

Watch the Lab 4 video tutorial entitled "**Assessing the Bank's Profitability in the Game Deal or No Deal using Microsoft® Excel®**".

You are the banker in the game "Deal or No Deal". Your job is to determine the amount the bank is willing to offer to purchase the contestant's selected case. The three offer scenarios you plan to consider are:

- i) Offer an amount equal to the fair value of the contestant's selected case.
- ii) Offer an amount 10% lower than the fair value of the contestant's selected case.
- iii) Offer an amount 20% lower than the fair value of the contestant's selected case.

The data set labeled "Deal or No Deal" reports the outcomes of 6,000 "Deal or No Deal" games in which the monetary values of just six cases remain unknown. The six possible case values are listed in adjacent columns. Assume the first column is the contestant's selected case, although this value is unknown to you when you propose the bank's offer.

(a) Using Microsoft® Excel®, complete the following steps 1 - 4 for each of the 6,000 games for offer scenarios i), ii) and iii).

1. Compute the expected (fair) value of the contestant's case.
2. Determine the bank's offer for the contestant's case.
3. Determine the bank's profit for the offered amount.
4. Compute the bank's average profit/offer as a function of N offers.

(b) Report the proportion of games in which the bank had a negative profit for scenario (i). Round your answer to **four** decimal places.

(c) Report the proportion of games in which the bank had a negative profit for scenario (ii). Round your answer to **four** decimal places.

(d) Report the proportion of games in which the bank had a negative profit for scenario (iii). Round your answer to **four** decimal places.

(e) Report the average profit/offer (\$) over the 6,000 games for offer scenario i). Round your answer to **two** decimal places.

(f) Report the average profit/offer(\$) over the 6,000 games for offer scenario ii). Round your answer to **two** decimal places.

(g) Report the average profit/offer (\$) over the 6,000 games for offer scenario iii). Round your answer to **two** decimal places.

(h) Plot the average profit/offer as a function of N offers for offer scenarios i), ii) and iii).

Question 2: Skee-Ball

Watch the Lab 4 video tutorial entitled "**Assessing the Profitability of the Game Skee-Ball using Microsoft® Excel®**".

A popular summer event is Skee-Ball. For \$2, a customer purchases three balls and attempts to roll each ball into a central target. The customer wins their \$2 back and wins an additional \$1 if they hit the target once, an additional \$3 if they hit the target twice and an additional \$5 if they hit the target three times. If the customer does not hit the target at all, they lose their initial \$2. Access the data set labeled "Skee-Ball" which reports the outcomes of 2,500 games for a single day's operation of a Skee-Ball booth.

(a) Define a random variable C (Customer Score) equal to the number of times a customer hits the target in each set of three rolls. How many possible outcomes of C are there? Report your answer as an **integer**.

(b) Create a histogram for the observed values of C , the number of times a customer hit the target in each set of three rolls.

Define the booth's profit (P) as:

$$\text{Booth's Profit} = \text{Entry Cost (i.e. \$2)} - \text{Customer Winnings.}$$

(c) Report the **net profit (\$)** that the Skee-Ball booth achieved for the day. Report your answer as an **integer**.

Estimate the PDF of C based on the relative frequencies: $P(C = c) = \frac{\text{Frequency of } C = c}{2,500}$.

(d) Report $P(C = 0)$. Round your answer to **three** decimal places.

(e) Report $P(C = 1)$. Round your answer to **three** decimal places.

(f) Report $P(C = 2)$. Round your answer to **three** decimal places.

(g) Report $P(C = 3)$. Round your answer to **three** decimal places.

(h) Plot the *estimated* PDF of the booth's profit (P).

Question 3: Binomial Random Variables

Watch the Lab 4 video tutorial entitled "**Binomial Random Variables**".

Watch the Lab 4 video tutorial entitled "**Evaluating Binomial Data using Microsoft® Excel®**"

Winston the labradoodle loves to play fetch. He catches his ball mid-air with probability p and catch attempts are statistically independent of one another. Winston attempts to catch his ball in a series of five tosses and then he takes a water break. The data set labeled "Winston" reports the number of catches over 3,500 trials of five tosses each.

(a) Define a random variable Y equal to the number of times Winston catches the ball mid-air in a series of five tosses. How many possible outcomes of Y are there? Report your answer as an **integer**.

(b) What proportion of the time did Winston catch exactly two balls mid-air out of five tosses? Round your answer to **three** decimal places.

(c) What proportion of the time did Winston catch exactly three balls mid-air out of five tosses? Round your answer to **three** decimal places.

(d) Use the data to estimate p , the probability that Winston catches each ball where:
$$p = \frac{\text{Total Catches}}{\text{Total Catch Attempts}}$$
 Round your answer to **three** decimal places.

Use your estimate of p above to compute the PDF of Y assuming Y is a binomial random variable.

(e) Report $P(Y = 0)$. Round your answer to **three** decimal places.

(f) Report $P(Y = 1)$. Round your answer to **three** decimal places.

(g) Report $P(Y = 2)$. Round your answer to **three** decimal places.

(h) Report $P(Y = 3)$. Round your answer to **three** decimal places.

(i) Report $P(Y = 4)$. Round your answer to **three** decimal places.

(j) Report $P(Y = 5)$. Round your answer to **three** decimal places.

(k) Create a histogram for the observed values of Y and the expected values of Y (as-

suming Y is a binomial random variable).

Assignment - Short Report

Produce a short report which includes the graphs you created in Questions 1 - 3. Reflect on the results you have obtained by answering the following questions.

Question 1: Deal or No Deal

- (a) Clearly label and paste into your report the graph of the average profit/offer (Scenario i) as a function of N offers.
- (b) Clearly label and paste into your report the graph of the average profit/offer (Scenario ii) as a function of N offers.
- (c) Clearly label and paste into your report the graph of the average profit/offer (Scenario iii) as a function of N offers.
- (d) The Law of Large Numbers guarantees that the average profit/offer will converge to a constant value over a sufficiently large number of N offers. Do you observe this convergence in your graphs of average profit/offer? Which scenario (i-iii) produces the largest average profit/offer?
- (e) In your opinion, how likely is it that a contestant willingly accepts an offer lower than the fair value, thereby guaranteeing the bank a positive net profit over a sufficiently large number of N offers?
- (f) Discuss how the Law of Large Numbers benefits the bank but offers no similar benefit to the contestant.

Question 2: Skee-Ball

- (a) Clearly label and paste into your report the histogram for the observed values of C , the number of times a customer hit the target in each set of three rolls.
- (b) Clearly label and paste into your report the histogram of the observed value of P (the booth's profit) for each of the 2,500 customer bets.

(c) Consider the net profit that the booth achieves over 2,500 customer bets. Is the current pricing strategy for the booth successful?

(d) Use the estimated PDF of C to develop a revised pricing strategy for the booth so that the booth's expected profit per Skee-Ball game equals 30 cents per game.

Question 3: Binomial Random Variable

(a) Clearly label and paste into your report the histogram for the observed values of Y and the expected values of Y (assuming Y is a binomial random variable).

(b) Consider the five criteria for a binomial random variable. Comment on the fit of each of these criteria with respect to the random variable Y .

(c) Compare the observed values of Y and the expected values of Y (assuming Y is a binomial random variable). How closely do they compare?