

Lab 9: Probability Exercise

- 1) **Purpose:** The main purpose of this lab is to make you more familiar with statistical distributions. You will go through experiments by rolling dice and tossing coins from the experiments, you will have a touching sense for the related distribution.
- 2) **Objectives:** Upon completion, you will be able to:
 - Relate probability to real-world scenarios.
 - Use Excel to generate random numbers.
- 3) **Convergence of the mean value**

In this part, you are expected to conduct 5 experiments, each of which consists of rolling a die for 8 times. Fill in the blanks of the provided tables and plot the **Mean Value** and **Sample Variance** of ALL five experiments in two separate graphs following the provided steps. The purpose of this experiments is to learn that, as the number of trials increases, how the mean value of the outcomes gradually converges to 3.5.

a. Experiment 1

Fill the second row as you roll the die. For the third row, calculate the mean of the outcome of the current trial, and all of its previous ones. For example, if you have 1 in first trial, then the mean is 1; and if you have 4 in second trial, the mean is the mean of the first and second trials (i.e. the mean of 1 and 4, which is 2.5). For the fourth row, calculate the sample variance similarly.

Trial number	1	2	3	4	5	6	7	8
Outcome								
Mean value								
Sample variance								

b. Experiment 2

Trial number	1	2	3	4	5	6	7	8
Outcome								
Mean value								
Sample variance								

c. Experiment 3

Trial number	1	2	3	4	5	6	7	8
Outcome								
Mean value								
Sample variance								

d. Experiment 4

Trial number	1	2	3	4	5	6	7	8
Outcome								
Mean value								
Sample variance								

e. Experiment 5

Trial number	1	2	3	4	5	6	7	8
Outcome								
Mean value								
Sample variance								

- f. Plot all mean values:
 1. Open an Excel file and name it “Your name_Lab9”. Copy all five tables to the first sheet of your Excel file and rename it as “Convergence”.
 2. Select **Insert>Line>2D Line**.
 3. Right click the graph, click “select data”.
 4. Click [Add]. In Series Values, erase the current content and select the Mean Values of the first experiment. In Series name, erase the content and type “Experiment 1” and click [OK].
 5. Repeat step 4 for the Mean Values of other experiments.
 6. Click [OK] in the Select Data Source window.
 7. Select **Design>Add chart element** and add Chart Title, Axis Title, and Legend to your graph.
- g. Plot all Sample Variances:
 - Repeat steps f-2 to f-7 using data from the fourth row of tables.
- h. Create a text box by selecting **Insert>Text Box** and summarize your findings in that. When you summarize, also calculate the variance of the “mean values” for the 8th trial of all experiments. You can use “=VAR.S()” function for that. What do you conclude from that?

4) Uniform distribution

In this experiment you will get familiar with the concept of Uniform distribution by rolling a die. The outcome of rolling a die follows a Discrete Uniform Distribution, which is a symmetric probability distribution, whereby a finite number of values are equally likely to be observed; every one of n values has equal probability (i.e. $1/n$). In case of rolling a die, n is equal to 6.

- a. Considering the formula for Uniform distribution, fill the second row of the table below.
- b. Fill the third row of the table below, supposing that you should roll a die for 10 times.
- c. Roll a die for 10 times and fill the fourth row with your results.
- d. Open the Excel file, which you created in part 3, and create a new worksheet and rename it as “Uniform”.
- e. Copy the table below and paste it in the worksheet you just created.
- f. Draw a column chart using “Theoretical number of outcomes” and “Actual number of observations” as your data:
 - Select **Insert>Charts>2D-Column**.
 - Right click the graph, click “select data”.
 - Click [Add]. In Series Values, erase the current content and highlight the numbers in the third row of the table. In Series name, erase the content and type “Theoretical number of outcomes” and click [OK].
 - Click [Add]. In Series Values, erase the current content and highlight the numbers in the fourth row of the table. In Series name, erase the content and type “Actual number of observations” and click [OK].
- g. Summarize your findings in one paragraph inside a text box. As you summarize, describe how the differences between the theoretical number of outcomes and actual number of observations can be used to tell if the outcome is actually from a uniform distribution.

X	1	2	3	4	5	6
Probability						
Theoretical number of outcomes						
Actual number of observation						

5) Binomial distribution

A binomial distribution is the distribution with which we can calculate the probability of achieving a set number of successes in a predetermined number of trials. If we consider achieving 1 as success and any other number as failure in rolling a die, we can calculate the probability of having a specific number in a specific number of trials (e.g. how many 1's in ten trials of rolling a die?)

- Create a new sheet in your Excel file and rename it as "Binomial".
- You can use Excel to calculate the probability of having a specific number of successes in a specific number of trials, using the formula "`=BINOMDIST(number_s, trials, probability, cumulative)`" where:

Number_s = number of successes (e.g. Number of 1's rolled=2)

Trials = number of trials (i.e. 10)

Probability = probability of rolling a specific number in a trial (i.e. 1/6)

Cumulative = TRUE or FALSE (i.e. enter TRUE if the question asks for UP TO TWO 1's rolled in ten trials and enter FALSE if the question asks EXACTLY TWO 1's rolled in ten trials.)

- In cell A1, use the above formula to determine the probability of rolling EXACTLY two 1's in a ten-trial study.
- In cell A2, use the above formula to determine the probability of rolling UP TO two 1's in a ten-trial study. (It means, zero 1's, one 1, or two 1's.)
- Fill in the blanks of table below. You will do eight experiments, each consisting of 10 rolls of a die and count the number of 1's in each experiment.

Number of experiments with X number of 1's rolled in 8 experiments of rolling a die for 10 times.										
X=0	X=1	X=2	X=3	X=4	X=5	X=6	X=7	X=8	X=9	X=10

- In cell A4, compute the percentage of experiments where you rolled EXACTLY TWO 1's ($P=n(X=2)/8$).
- In cell A5, compute the percentage of experiments where you rolled UP TO TWO 1's ($P=n(X=0, 1, 2)/8$).
- Create a text box and answer the following questions in that: How do the experimental values (cells A4, A5) compare with the theoretical probabilities? (cells A1, A2) What could be done to bring the values of cells A4, A5 closer to the values of cells A1, A2?
- Fill in the blanks of the table below with the outcomes of step e. Copy and paste the table below in the current sheet and draw a Column graph from the path: **Insert>Chart>2D-Column**. Add data series like what you did in part 4.

Outcome of All 80 Trials	Frequencies of the outcome X (X=1, 2, 3, 4, 5, 6)					
	1	2	3	4	5	6

6) Geometric distribution

Geometric distribution is a special case of Negative Binomial distribution. The negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of Bernoulli trials before a pre-specified number of failures (denoted by r) occurs. For example, if we define a "1" as failure, all non-"1"s as successes, and we throw a dice repeatedly until the third time "1" appears ($r = 3$), then the probability distribution of the number of trials performed will be a negative binomial. If we set $r=1$, negative binomial distribution will turn to Geometric distribution.

- a. Open a new sheet in your Excel file and rename it as “Geometric”. We are going to explore the question: If you continuously roll a six-sided die, what is the probability that the first 1 shows up on the 3rd roll? Answer the following questions in a text box:
- b. What could the first roll of the die be (i.e. what numbers?) to have your first occurrence of the number “1” is on the third roll?
- c. What is the probability of rolling the number(s) you specified in question b?
- d. What could the second roll of the die be to accomplish the above situation?
- e. What is the probability of rolling the number(s) you specified?
- f. What must the third roll of the die be in order to accomplish the above situation?
- g. What is the probability of rolling the number you specified?
- h. What is the probability that your first occurrence of the number “1” is on the third roll? (intersect of the first three rolls)
- i. We will now use Excel to find the probability of rolling a 1 for the first time on the 3rd roll of a die. This is an example of a geometric distribution
- j. In cell A1, type “=NEGBINOMDIST(2,1,1/6)” When you type in “NEGBINOMDIST” you will be prompted to type in “(number_f, number_s, probability_s)”. number_f=number of failures before sth success, number_s=number of successes... if the question asked how many rolls of the die before the SECOND 1 is rolled, then number_s would equal 2. Probability_s=the probability of success (rolling a 1).
- k. Answer the following question in a NEW text box BELOW your work from the previous step: Do the probabilities from step h and the probability in cell A1 equal each other? What does this mean?

7) Central Limit Theorem (Rolling a die)

In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, will be approximately normally distributed, regardless of the underlying distribution. That is, suppose that a sample is obtained containing a large number of observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic average of the observed values is computed. If this procedure is performed many times, the central limit theorem says that the computed values of the *average* will be distributed according to the normal distribution.

a. Experiment 1

1. Roll a die one time, and record the mean value. (If we got 2, then the mean value is 2.)
2. Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 1	$\bar{X} < 1.5$	$1.5 \leq \bar{X} < 2.5$	$2.5 \leq \bar{X} < 3.5$	$3.5 \leq \bar{X} < 4.5$	$4.5 \leq \bar{X} < 5.5$	$5.5 \leq \bar{X}$	Sum
Frequency							10
Relative Frequency							1

b. Experiment 2

- Roll a die two times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 2	$\bar{X} < 1.5$	$1.5 \leq \bar{X} < 2.5$	$2.5 \leq \bar{X} < 3.5$	$3.5 \leq \bar{X} < 4.5$	$4.5 \leq \bar{X} < 5.5$	$5.5 \leq \bar{X}$	Sum
Frequency							10
Relative Frequency							1

c. Experiment 3

- Roll a die three times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 3	$\bar{X} < 1.5$	$1.5 \leq \bar{X} < 2.5$	$2.5 \leq \bar{X} < 3.5$	$3.5 \leq \bar{X} < 4.5$	$4.5 \leq \bar{X} < 5.5$	$5.5 \leq \bar{X}$	Sum
Frequency							10
Relative Frequency							1

d. Experiment 4

- Roll a die four times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 4	$\bar{X} < 1.5$	$1.5 \leq \bar{X} < 2.5$	$2.5 \leq \bar{X} < 3.5$	$3.5 \leq \bar{X} < 4.5$	$4.5 \leq \bar{X} < 5.5$	$5.5 \leq \bar{X}$	Sum
Frequency							10
Relative Frequency							1

- e. Create a new sheet in your excel file and rename it as “CLT_die”.
- f. Copy and paste the tables above and make a line graph with **Relative Frequency** data of the four experiments (a line graph consisting of four series of data). In a text box write the CLT. Write what you found for this CLT experiment.

8) Central Limit Theorem (Tossing a coin)

a. Experiment 1

- Toss a coin one time, and record the mean value. (If you got HEAD, $X=1$ and if you got TAIL, $X=0$. In Experiment 1, you will have the mean value of 0 or 1.)
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 1	$\bar{X} < 0.2$	$0.2 \leq \bar{X} < 0.4$	$0.4 \leq \bar{X} < 0.6$	$0.6 \leq \bar{X} < 0.8$	$0.8 \leq \bar{X}$	Sum
Frequency						10
Relative Frequency						1

b. Experiment 2

- Toss a coin two times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 1	$\bar{X} < 0.2$	$0.2 \leq \bar{X} < 0.4$	$0.4 \leq \bar{X} < 0.6$	$0.6 \leq \bar{X} < 0.8$	$0.8 \leq \bar{X}$	Sum
Frequency						10
Relative Frequency						1

c. Experiment 3

- Toss a coin three times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 1	$\bar{X} < 0.2$	$0.2 \leq \bar{X} < 0.4$	$0.4 \leq \bar{X} < 0.6$	$0.6 \leq \bar{X} < 0.8$	$0.8 \leq \bar{X}$	Sum
Frequency						10
Relative Frequency						1

d. Experiment 4

- Toss a coin four times, and record the mean value.
- Repeat this 10 times to fill out the following table.

Class of $\bar{X}$ of size 1	$\bar{X} < 0.2$	$0.2 \leq \bar{X} < 0.4$	$0.4 \leq \bar{X} < 0.6$	$0.6 \leq \bar{X} < 0.8$	$0.8 \leq \bar{X}$	Sum
Frequency						10
Relative Frequency						1

- e. Create a new sheet in your excel file and rename it as “CLT_coin”.
- f. Make a line graph with **relative frequency** data of all experiments (a line graph consisting of four series of data). In a text box write what you found for this CLT experiment.

Assignment general instructions

- Upload the Excel file you worked on in this lab on D2L in folder “Lab9 Assignment”. Your Excel file should have six worksheets.