PROBABILITY AND RANDOM PROCESSES (15B11MA301)

Department of Mathematics

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

(SECTOR – 128 NOIDA, UTTAR PRADESH)

EVEN 2021, Semester - 4

PROJECT BASED LEARING ASSIGNMENT

Topic: Probability Distribution In Health Sciences

Submitted By:

Group – 1

Siddhant Jindal 9918103084
Ashish Chauhan 9919103062
Vanshika Agarwal 9919103063
Shreya 9919103064
Naina Sharma 9919103065
Himanshu Sharma 9919103066

Submitted to:

Dr. Pinkey Chauhan

PROJECT BASED LEARNING ASSIGNMENT ELABORATING THE APPLICATION OF:

Probability Distribution in Health Sciences

DATE OF SUBMISSION- 20th MAY'2021

AN ASSIGNMENT UNDER

THE SUPERVISOR

DR. PINKEY CHAUHAN

PROBABILITY AND RANDOM PROCESSES (15B11MA301)

SUBMITTED BY:

GROUP – 1		
Name	**Enrollment no.**	**Batch**
Siddhant Jindal	9918103084	F3
Ashish Chauhan	9919103062	F3
Vanshika Agarwal	9919103063	F3
Shreya	9919103064	F3
Naina Sharma	9919103065	F3
Himanshu Sharma	9919103066	F3

ACKNOWLEDGEMENT

Presentation inspiration and motivation have always played a key role in the success of any venture.

We pay our deep sense of gratitude to**, Vice Chancellor Dr. S C Saxena, Jaypee Institute of Information Technology, Noida** to encourage us to the highest peak and to provide us the opportunity to prepare the project. We are immensely obliged to for their elevating inspiration, encouraging guidance and kind supervision in the completion of our project.

We feel to acknowledge our indebtedness and deep sense of gratitude to our guide **DR. PINKEY CHAUHAN** whose invaluable guidance and kind supervision given to us throughout the course which shaped the present work as its show.

Last, but not the least, **our parents** are also an important inspiration for time. So with due regards. We express our gratitude to them.

TABLE OF CONTENTS

1. INTRODUCTION

In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment . It is a mathematical description of a random phenomenon in terms of its sample space and the probabilities of events . Many processes taking place in the world around us can be described by a handful of distributions that have been well-researched and analyzed. Probability Distribution is a most commonly used concept in research and acquaints with the shape of the data.

Probability and probability distributions play a central part in medical statistics .We can gather information from the lab about the diagnosis, and summarize data. These data might give statistics such as the average age, gender, eating habit and smoking pattern, number of breast cancer cases and so on. All of these statistics simply describe characteristics of the sample and are therefore called descriptive statistics. The probability distributions are the part of descriptive statistics to describe the shape of the data and possibly predict the probability of an event.

Probability distribution plays a significant role in the field of health sciences. Although there are a number of probability distributions, majorly three distributions are used in medical research studies i.e. Binomial, Poisson and Gaussian/Normal distribution. Binomial distribution has very wide application and it is important as it allows us to deal with the outcome belonging to two categories such as accepted/rejected, yes/no or male/female. It is a probability model for a discrete outcome with dichotomous nature. It gives you the probability of m successes among n trials of any event. Poisson distribution describes the behavior of rare events such as patients arriving at an emergency room, decaying radioactive atoms, bank customers coming to their bank, number of suicide cases in adolescence . Normal distribution is used to identify whether to select a parametric test or non-parametric test.

2. OBJECTIVE

The aim of this assignment is to demonstrate the application of discrete/continuous distributions in health sciences. It aims to compute the probability of observing a specified number of "successes" when the process is repeated a specific number of times (e.g., in a set of patients) and the outcome for a given patient is either a success or a failure

This project has made utilization of The Binomial Distribution: A Probability Model for a Discrete Outcome for medical research studies. The binomial distribution is used when a researcher is interested in the occurrence of an event, not in its magnitude. For instance, adults with allergies might report relief with medication or not, children with a bacterial infection might respond to antibiotic therapy or not, adults who suffer a myocardial infarction might survive the heart attack or not, a medical device such as a coronary stent might be successfully implanted or not, in a clinical trial, a patient may survive or die. These are just a few examples of applications or processes in which the outcome of interest has two possible values (i.e., it is dichotomous). The two outcomes are often labeled "success" and "failure" with success indicating the presence of the outcome of interest. Note, however, that for many medical and public health questions the outcome or event of interest is the occurrence of disease, which is obviously not really a success. Nevertheless, this terminology is typically used when discussing the binomial distribution model. As a result, whenever using the binomial distribution, we must clearly specify which outcome is the "success" and which is the "failure". The researcher studies the number of survivors, and not how long the patient survives after treatment. Another example is whether a person is ambitious or not. Here, the binomial distribution describes the number of ambitious persons, and not how ambitious they are.

The binomial distribution is specified by the number of observations, n, and the probability of occurrence, which is denoted by p. Binomial distribution has very wide application and it is important as it allows us to deal with the outcome belongs to two categories such as accepted/rejected, yes/no or male/female. It is a probability model for a discrete outcome with dichotomous nature. It gives us the probability of *m* successes among *n* trials of any event. For example, there are 40 kidney cancer patients and we are looking for 5-year survival of 16 patients. The individual patient outcomes are independent and if we assume that the probability of survival is p = 0.20 or 20% for all patients then the required probability will be 0.2%.

<u>This assignment will also help us in comparison of one sample proportion to an expected proportion (for small samples).Also helps in Evaluation of a new treatment and a risk factor.</u>

Similarly, we may find several options such as what is the probability of at least 16 patients survives, more than 16 patients survives etc. This report will empower the management to realize the measures to put in each phase of the project in order to make the fulfillment successful.

3. THE BINOMIAL DISTRIBUTION

The binomial distribution model is an important probability model that is used when there are two possible outcomes (hence "binomial"). In a situation in which there were more than two distinct outcomes, a multinomial probability model might be appropriate, but here we focus on the situation in which the outcome is dichotomous.

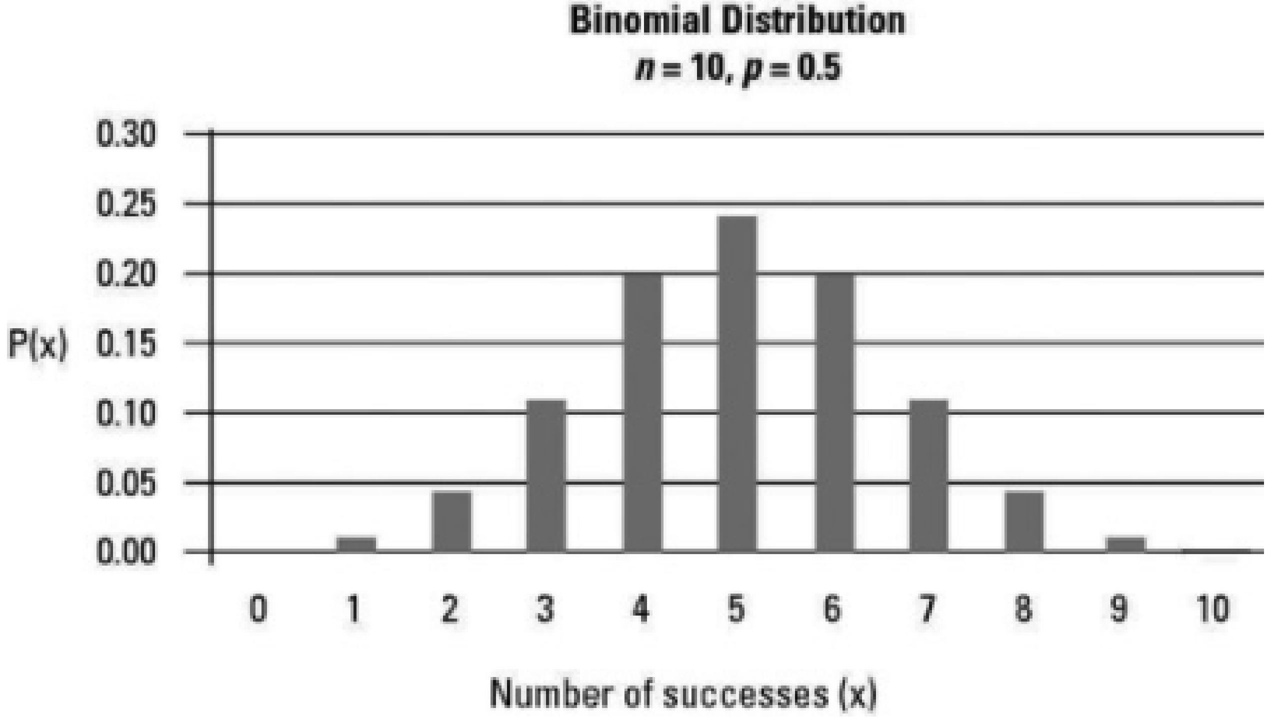
For example, adults with allergies might report relief with medication or not, children with a bacterial infection might respond to antibiotic therapy or not, adults who suffer a myocardial infarction might survive the heart attack or not, a medical device such as a coronary stent might be successfully implanted or not. These are just a few examples of applications or processes in which the outcome of interest has two possible values. The two outcomes are often labeled "success" and "failure" with success indicating the presence of the outcome of interest. Note, however, that for many medical and public health questions the outcome or event of interest is the occurrence of disease, which is obviously not really a success. As a result, whenever using the binomial distribution, we must clearly specify which outcome is the "success" and which is the "failure".

The binomial distribution model allows us to compute the probability of observing a specified number of "successes" when the process is repeated a specific number of times (e.g., in a set of patients) and the outcome for a given patient is either a success or a failure. We must first introduce some notation which is necessary for the binomial distribution model.

First, we let "n" denote the number of observations or the number of times the process is repeated, and "x" denotes the number of "successes" or events of interest occurring during "n" observations. The probability of "success" or occurrence of the outcome of interest is indicated by "p".

3.1.The Binomial Distribution Model

$$P\left(X\,\textbf{"successes"}\right) = \frac{n!}{x!\,(n-x)!}p^{x}\left(1-p\right)^{(n-x)}$$



Use of the binomial distribution requires three assumptions:

1. Each replication of the process results in one of two possible outcomes (success or failure).
2. The probability of success is the same for each replication, and
3. The replications are independent, meaning here that a success in one patient does not influence the probability of success in another.

4. APPLICATIONS OF BINOMIAL DISTRIBUTIONS IN HEALTH SCIENCES

4.1. Relief of Allergies

Suppose that 80% of adults with allergies report symptomatic relief with a specific medication. If the medication is given to 10 new patients with allergies, what is the probability that it is effective in exactly seven? First, do we satisfy the three assumptions of the binomial distribution model?

1. The outcome is relief from symptoms (yes or no), and here we will call a reported relief from symptoms a 'success.'
2. The probability of success for each person is 0.8.
3. The final assumption is that the replications are independent, and it is reasonable to assume that this is true.

We know that:

- # observation is n=10
- # successes or events of interest is x=7
- p=0.80

The probability of 7 successes is:

$$P(7\ \textbf{successes}) = \frac{10!}{7!(10-7)!}\mathbf{0.80}^7\left(\mathbf{1-0.80}\right)^{(10-7)}$$

This is equivalent to:

$$P(7\ \textbf{successes}) = \frac{10(9)(8)(7)(6)(5)(4)(3)(2)(1)}{[7(6)(5)(4)(3)(2)(1)]\ [(3)(2)(1)]}\left(\mathbf{0.80}\right)^7\left(\mathbf{1-0.80}\right)^{10-7}$$

But many of the terms in the numerator and denominator cancel each other out,

$$= \frac{10(9)(8)\cancel{(7)}\cancel{(6)}\cancel{(5)}\cancel{(4)}\cancel{(3)}\cancel{(2)}\cancel{(1)}}{[\cancel{7}\cancel{(6)}\cancel{(5)}\cancel{(4)}\cancel{(3)}\cancel{(2)}\cancel{(1)}]\ [3(2)(1)]} \quad (0.2097)\ (0.008)$$

So this can be simplified to:

$$P(7 \text{ successes}) = \frac{10(9)(8)}{3(2)(1)} (0.2097)(0.008) = 120(0.0297)\ (0.008) = 0.2013$$

4.1.1. Interpretation

There is a 20.13% probability that exactly 7 of 10 patients will report relief from symptoms when the probability that any one reports relief is 80%.

Binomial probabilities like this can also be computed in an Excel spreadsheet using the =BINOMDIST function. Place the cursor into an empty cell and enter the following formula:

=BINOMDIST(x,n,p,FALSE)

where x= # of 'successes', n = # of replications or observations, and p = probability of success on a single observation.

What is the probability that none report relief? We can again use the binomial distribution model with n=10, x=0 and p=0.80.

$$P(0 \text{ successes}) = \frac{10!}{0!(10-0)!} 0.80^0 (1 - 0.80)^{10-0}$$

This is equivalent to

$$P(0 \text{ successes}) = \frac{10!}{(1)(10)!} 0.80^0 (0.20)^{10}$$

This simplifies to

$$P(0 \text{ successes}) = (1)\ (1)\ (0.0000001024) = 0.000001024$$

There is practically no chance that none of the 10 will report relief from symptoms when the probability of reporting relief for any individual patient is 80%.

What is the most likely number of patients who will report relief out of 10? If 80% report relief and we consider 10 patients, we would expect that 8 report relief. What is the probability that exactly 8 of 10 report relief?

We can use the same method that was used above to demonstrate that there is a 30.30% probability that exactly 8 of 10 patients will report relief from symptoms when the probability that any one reports relief is 80%. The probability that exactly 8 report relief will be the highest probability of all possible outcomes (0 through 10).

4.2. The Probability of Dying after a Heart Attack

The likelihood that a patient with a heart attack dies of the attack is 0.04 (i.e., 4 of 100 die of the attack). Suppose we have 5 patients who suffer a heart attack, what is the probability that all will survive? For this example, we will call a success a fatal attack ($p = 0.04$). We have $n=5$ patients and want to know the probability that all survive or, in other words, that none are fatal (0 successes).

We again need to assess the assumptions. Each attack is fatal or non-fatal, the probability of a fatal attack is 4% for all patients and the outcome of individual patients are independent. It should be noted that the assumption that the probability of success applies to all patients must be evaluated carefully. The probability that a patient dies from a heart attack depends on many factors including age, the severity of the attack, and other co morbid conditions. To apply the 4% probability we must be convinced that all patients are at the same risk of a fatal attack. The assumption of independence of events must also be evaluated carefully. As long as the patients are unrelated, the assumption is usually appropriate. Prognosis of disease could be related or correlated in members of the same family or in individuals who are co-habituating. In this example, suppose that the 5 patients being analyzed are unrelated, of similar age and free of co morbid conditions.

$$P(0 \text{ successes}) = \frac{5!}{0!(5-0)!} 0.04^0 (1-0.04)^{5-0}$$

$$P(0 \text{ successes}) = \frac{5!}{5!}(1)\ (0.96)^5 = (1)\ (1)\ (0.8154) = 0.8154$$

There is an 81.54% probability that all patients will survive the attack when the probability that any one dies is 4%. In this example, the possible outcomes are 0, 1, 2, 3, 4 or 5 successes (fatalities). Because the probability of fatality is so low, the most likely response is 0 (all patients survive). The binomial formula generates the probability of observing exactly x successes out of n.

4.2.1. Computing the Probability of a Range of Outcomes

If we want to compute the probability of a range of outcomes we need to apply the formula more than once. Suppose in the heart attack example we wanted to compute the probability that *no more than 1 person dies of the heart attack*. In other words, 0 or 1, but not more than 1. Specifically we want P(no more than 1 success) = P(0 or 1 successes) = P(0 successes) + P(1 success). To solve this probability we apply the binomial formula twice.

We already computed P(0 successes), we now compute P(1 success):

$$P(1 \text{ success}) = \frac{5!}{1!(5-1)!} 0.04^1 (1-0.04)^{5-1}$$

$$P(1 \text{ success}) = \frac{5!}{(1)(4)!}(0.04)\ (0.96)^4 = (5)(0.04)\ (0.8493) = 0.1697$$

P(no more than 1 'success') = P(0 or 1 successes) = P(0 successes) + P(1 success)
= 0.8154 + 0.1697 = 0.9851.

The probability that no more than 1 of 5 (or equivalently that at most 1 of 5) die from the attack is 98.51%.

What is the probability that 2 or more of 5 die from the attack? Here we want to compute P(2 or more successes). The possible outcomes are 0, 1, 2, 3, 4 or 5, and the sum of the probabilities of each of these outcomes is 1 (i.e., we are certain to observe either 0, 1, 2, 3, 4 or 5 successes). We just computed P(0 or 1 successes) = 0.9851, so P(2, 3, 4 or 5 successes) = 1 - P(0 or 1 successes) = 0.0149. There is a 1.49% probability that 2 or more of 5 will die from the attack.

4.2.2. Mean and Standard Deviation of a Binomial Populations

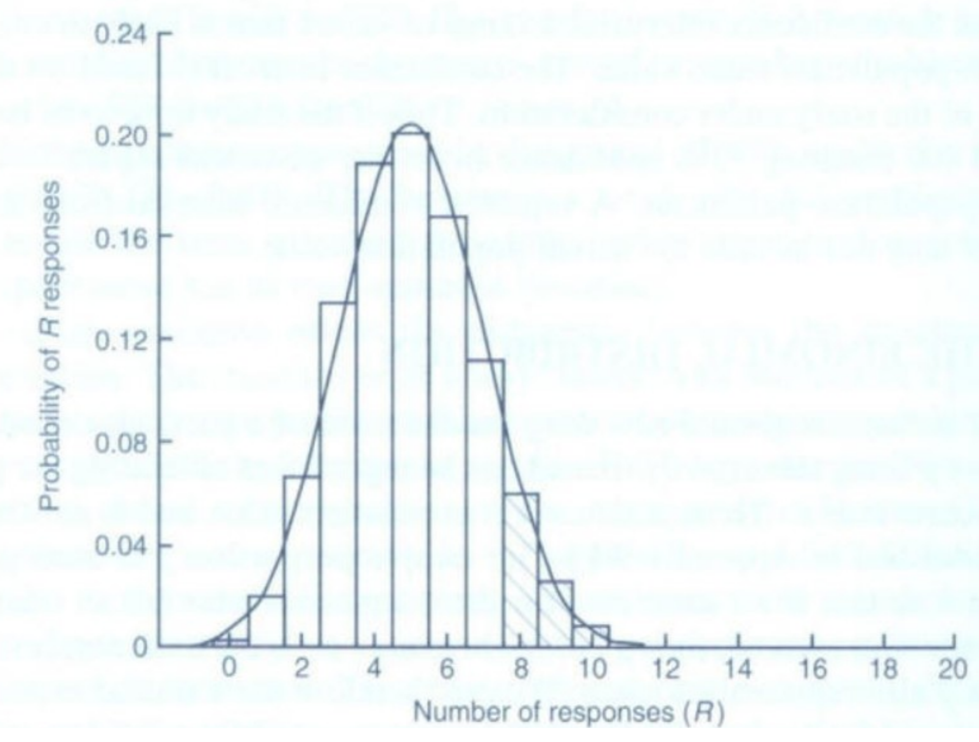
Mean number of successes: $\mu = np$

Standard Deviation: $\sigma = \sqrt{n(p)(1-p)}$

For the previous example on the probability of relief from allergies with n-10 trials and p=0.80 probability of success on each trial:

$$\mu = np = (10)(0.80) = 8$$

$$\sigma = \sqrt{n(p)(1-p)} = \sqrt{10(0.8)(0.2)} = 1.3$$



5. CONCLUSION

Health science is the application of science to health including the study of medicine, nutrition, and other health-related topics.

Like privacy, health research has high value to society. It can provide important information about disease trends and risk factors, outcomes of treatment or public health interventions, functional abilities, patterns of care, and health care costs and use. There are two parts to health science: the study, research, and knowledge of health and the application of that knowledge to improve health, cure diseases, and understanding how humans and animals function.

The binomial distribution model is an important probability model that is used when there are two possible outcomes. In a situation in which there were more than two distinct outcomes, a multinomial probability model might be appropriate, but here we focus on the situation in which the outcome is dichotomous. The binomial distribution model allows us to compute the probability of observing a specified number of "successes" when the process is repeated a specific number of times and the outcome for a given patient is either a success or a failure. We must first introduce some notation which is necessary for the binomial distribution model. The binomial equation also uses factorials. In mathematics, the factorial of a non-negative integer k is denoted by k!, which is the product of all positive integers less than or equal to k.

Use of the binomial distribution requires three assumptions:
1) Each replication of the process results in one of two possible outcomes.
2) The probability of success is the same for each replication.
3) The replications are independent, meaning here that a success in one patient does not influence the probability of success in another.

If we want to compute the probability of a range of outcomes we need to apply the formula more than once. Suppose in the heart attack example we wanted to compute the probability that no more than 1 person dies of the heart attack.

6. REFERENCES

1. https://sphweb.bumc.bu.edu/otlt/mphmodules/bs/bs704_probability/bs704_probability7.html
2. http://www.pathbliss.com/index.php/2017/11/18/distributions-medical-statistics
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4387424
4. https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1b-statistical-methods/statistical-distributions
5. https://oxfordmedicine.com/view/10.1093/med/9780198743583.001.0001/med-9780198743583-chapter-7
6. https://medicalphysics.org/documents/Nickloff_stats_Ch8.pdf
7. https://www.tandfonline.com/doi/pdf/10.1080/01605682.2019.1609878
8. https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=5149&context=etd

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY NOIDA

VALIDATION CERTIFICATE

This is to certify that **SIDDHANT JINDAL, ASHISH CHAUHAN, VANSHIKA AGARWAL, SHREYA, NAINA SHARMA, HIMANSHU SHARMA** have successfully done the PBL on

Probability Distribution In Health Sciences

On **20^{th} MAY, 2021.**

DR. PINKEY CHAUHAN