

The Binomial Probability Distribution

- The number of independent trials is denoted n .
- The probability of a 'success' is p
- The expected number of 'successes' from n trials is $E(X) = np$

Characteristics of a Poisson Experiment A Poisson experiment is a statistical experiment that has the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (m) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the region.
- The probability that a success will occur in an extremely small region is virtually zero.

Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

Probability Tables

- For some value r the tables record the probability of $P(X \geq r)$.
- The Student is required to locate the appropriate column based on the parameter values for the distribution in question.
- A copy of the Murdoch Barnes Tables will be furnished to the student in the End of Year Exam. The Tables are not required for the first mid-term exam.
- Knowledge of the sample space, partitioning of the sample points, and the complement rule are advised.

Binomial Distribution : Using Tables It is estimated by a particular bank that 25% of credit card customers pay only the minimum amount due on their monthly credit card bill and do not pay the total amount due. 50 credit card customers are randomly selected.

1. (3 marks) What is the probability that 9 or more of the selected customers pay only the minimum amount due?
2. (3 marks) What is the probability that less than 6 of the selected customers pay only the minimum amount due?
3. (3 marks) What is the probability that more than 5 but less than 10 of the selected customers pay only the minimum amount due?

Binomial Distribution : Using Tables Demonstration on Blackboard re: how to use tables in class.

1. $P(X \geq 9) = 0.9084$
2. $P(X < 6) = 1 - P(X \geq 6) = 1 - 0.9930 = 0.0070$
3. $P(6 \leq X \leq 9) = P(X \geq 6) - P(X \geq 10) = 0.9930 - 0.8363 = 0.1567$

```

n=60000 Y=numeric(n) for ( i in 1:n)
X=floor(runif(100,1,7)) Y[i]=sum(X)
Y hist(Y,breaks=seq(300,400,by=10),main=c("Totals of 100 Die Throws"),cex.lab=1.4,font.lab=2,xlab=c("Total
Score"))
hist(Y,breaks=seq(300,400,by=20),main=c("Totals of 100 Die Throws"),cex.lab=1.4,font.lab=2,xlab=c("Total
Score"))
Z=seq(1:n) Y/Z
plot(Y/Z,type="l",col="red",main=c("Die Rolls: Running Average"),font.lab=2,ylab="Average Value",
xlab=" Number of Throws") abline(h=3.5,col="green")
plot(Z,Z.y,pch=16,col="red",ylim=c(2.5,5.5),main=c("Variance"),font.lab=2,ylab=" ", xlab="X: Green Y:
Blue Z: Red" )
points(Y,Y.y,pch=16,col="blue" ) points(X,X.y,pch=16,col="green" ) points(c(1000,1000,1000),c(3,4,5),pch=18,cex=1.2)
lines(c(1000,1000),c(2.75,5.25),lty=3)
n=100000 Y=numeric(n) for ( i in 1:n)
X=floor(runif(100,1,7)) Y[i]=mean(X)
Y hist(Y,breaks=seq(270,430,by=2),main=c("Mean of 100 Die Throws (n= 100,000)"),cex.lab=1.4,font.lab=2,xlab=c("M
of 100 throws"))

```

```

n=60000 Y=numeric(n) for ( i in 1:n)
X=floor(runif(100,1,7)) Y[i]=sum(X)
Y hist(Y,breaks=seq(300,400,by=10),main=c("Totals of 100 Die Throws"),cex.lab=1.4,font.lab=2,xlab=c("Total
Score"))
hist(Y,breaks=seq(300,400,by=20),main=c("Totals of 100 Die Throws"),cex.lab=1.4,font.lab=2,xlab=c("Total
Score"))
Z=seq(1:n) Y/Z
plot(Y/Z,type="l",col="red",main=c("Die Rolls: Running Average"),font.lab=2,ylab="Average Value",
xlab=" Number of Throws") abline(h=3.5,col="green")
plot(Z,Z.y,pch=16,col="red",ylim=c(2.5,5.5),main=c("Variance"),font.lab=2,ylab=" ", xlab="X: Green Y:
Blue Z: Red" )
points(Y,Y.y,pch=16,col="blue" ) points(X,X.y,pch=16,col="green" ) points(c(1000,1000,1000),c(3,4,5),pch=18,cex=1.2)
lines(c(1000,1000),c(2.75,5.25),lty=3)
n=100000 Y=numeric(n) for ( i in 1:n)
X=floor(runif(100,1,7)) Y[i]=mean(X)
Y hist(Y,breaks=seq(270,430,by=2),main=c("Mean of 100 Die Throws (n= 100,000)"),cex.lab=1.4,font.lab=2,xlab=c("M
of 100 throws"))

```

- Binomial Coefficients / The Choose Operator
- Definition: The Probability Mass Functions (pmf)
- Binomial Distribution : Example

Binomial Coefficients

In the last class, we came across binomial coefficients. Informally, binomial coefficients are the number of ways k items can be selected from a group of n items. The binomial coefficient indexed by n and k is usually written as nC_k or

$$\binom{n}{k}$$

. C is colloquially known as the “choose operator”.

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

(We call the operator the choose operator. We will use both notations interchangeably.)

Probability Mass Function (Formally defining something mentioned previously)

- a probability mass function (pmf) is a **function** that gives the probability that a discrete random variable is exactly equal to some value.

$$P(X = k)$$

- The probability mass function is often the primary means of defining a discrete probability distribution
- It is conventional to present the probability mass function in the form of a table.
- The p.m.f of a value k is often denoted $f(k)$.

Probability Tables In the **Sulis** workspace there are two important tables used for this part of the course. This class will feature a demonstration on how to read those tables.

- The Cumulative Binomial Tables (Murdoch Barnes Tables 1)
- The Cumulative Poisson Tables (Murdoch Barnes Tables 2)

Please get a copy of each as soon as possible.

Probability Tables

- For some value r the tables record the probability of $P(X \geq r)$.
- The Student is required to locate the appropriate column based on the parameter values for the distribution in question.
- A copy of the Murdoch Barnes Tables will be furnished to the student in the End of Year Exam. The Tables are not required for the first mid-term exam.
- Knowledge of the sample space, partitioning of the sample points, and the complement rule are advised.

Binomial Distribution : Using Tables It is estimated by a particular bank that 25% of credit card customers pay only the minimum amount due on their monthly credit card bill and do not pay the total amount due. 50 credit card customers are randomly selected.

1. (3 marks) What is the probability that 9 or more of the selected customers pay only the minimum amount due?
2. (3 marks) What is the probability that less than 6 of the selected customers pay only the minimum amount due?
3. (3 marks) What is the probability that more than 5 but less than 10 of the selected customers pay only the minimum amount due?

Binomial Distribution : Using Tables Demonstration on Blackboard re: how to use tables in class.

1. $P(X \geq 9) = 0.9084$
2. $P(X < 6) = 1 - P(X \geq 6) = 1 - 0.9930 = 0.0070$
3. $P(6 \leq X \leq 9) = P(X \geq 6) - P(X \geq 10) = 0.9930 - 0.8363 = 0.1567$

Binomial Distribution: Expected Value and Variance

If the random variable X has a binomial distribution with parameters n and p , we write

$$X \sim B(n, p)$$

Expectation and Variance If $X \sim B(n, p)$, then:

- Expected Value of X : $E(X) = np$
- Variance of X : $Var(X) = np(1 - p)$

Suppose $n=3$ and $p=0.5$ Then $E(X) = 1.5$ and $V(X) = 0.75$.

Remark: Referring to the expected value and variance may be used to validate the assumption of a binomial distribution.

The Cumulative Distribution Function

- The Cumulative Distribution Function, denoted $F(x)$, is a common way that the probabilities of a random variable (both discrete and continuous) can be summarized.
- The Cumulative Distribution Function, which also can be described by a formula or summarized in a table, is defined as:

$$F(x) = P(X \leq x)$$

- The notation for a cumulative distribution function, $F(x)$, entails using a capital "F". (The notation for a probability mass or density function, $f(x)$, i.e. using a lowercase "f". The notation is not interchangeable.

0.1 Useful Results for Discrete Random Variables

Useful Results (Demonstration on the blackboard re: partitioning of the sample space, using examples on next slide)

- $P(X \leq 1) = P(X = 0) + P(X = 1)$
- $P(X \leq r) = P(X = 0) + P(X = 1) + \dots + P(X = r)$
- $P(X \leq 0) = P(X = 0)$
- $P(X = r) = P(X \geq r) - P(X \geq r + 1)$
- **Complement Rule:** $P(X \leq r - 1) = P(X < r) = 1 - P(X \geq r)$
- **Interval Rule:** $P(a \leq X \leq b) = P(X \geq a) - P(X \geq b + 1)$.

For the binomial distribution, if the probability of success is greater than 0.5, instead of considering the number of successes, to use the table we consider the number of failures.

a newspaper on average 1 in 10 000 characters is incorrectly printed. Suppose the paper contains 50 000 characters. Calculate the exact probability that

- (i) no printing errors are made
- (ii) at least 3 errors are made

Using the appropriate approximation, estimate these two probabilities.

Binomial Example

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution:

This is a binomial experiment in which

- a success is defined as an outcome of '4'.
- the number of trials is equal to $n = 5$,

- the number of successes is equal to $k = 2$,
- the number of failures is equal to 3,
- the probability of success on a single trial is $1/6$,
- the probability of failure on a single trial is $5/6$.

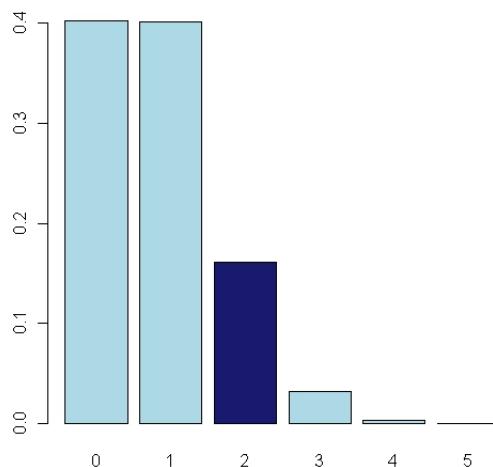
Binomial Example

Therefore, the probability of getting exactly 2 fours is:

$$P(X = 2) = {}^5C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

Remark: ${}^5C_2 = 10$

Bar plot : Number of successes from 5 throws of a die



Remark : The sum of the probabilities of each of the possible outcomes (i.e. no fours, one four etc) is equal to one.

$$P(X = 0) + P(X = 1) + \dots + P(X = 5) = 1$$

Binomial Example: At least two successes

- Suppose we were asked to find the probability of **at least** 2 fours.
- Can you suggest the most efficient way of computing this?
- Suggestion: Compute $P(X = 0)$ and $P(X = 1)$.
- Together these probabilities are the complement probability of what we require.
- $P(X \geq 2) = 1 - (P(X = 0) + P(X = 1))$.

- (We will continue with this in future classes).

Binomial Example: Sample Problem

Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.

Solution: Since only one out of five possible answers is correct, the probability of answering a question correctly by random is $1/5 = 0.2$. We can find the probability of having exactly 4 correct answers by random attempts as follows. (Blackboard. Correct Answer is 13.29%)

Binomial Example 1 (Revision from Last Class)

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution: This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is $1/6$ or about 0.167. Therefore, the binomial probability is:

$$P(X = 2) = {}^5C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

Binomial Example 2 Suppose there is a container that contains 6 items. The probability that any one of these items is defective is 0.3. Suppose all six items are inspected.

- What is the probability of 3 defective components?
- What is the probability of 4 defective components?

$$P(3 \text{ defects}) = f(3) = P(X = 3) = \binom{6}{3} 0.3^3 (1 - 0.3)^{6-3} = 0.1852$$

$$P(4 \text{ defects}) = f(4) = P(X = 4) = \binom{6}{4} 0.3^4 (1 - 0.3)^{6-4} = 0.0595$$

Binomial Distribution: Example 2

- Let the number of containers be the number of independent trials is $n = 100$.
- A success may be defined as a defective component.
- The probability of a success is approximate $p = 0.10$. (The probability of “failure” is $1 - p = 0.9$).
- The expected number of defective components is $np = 10$, which concurs with our observed data.
- The variance is computed as

$$np(1 - p) = 100 \times 0.1 \times 0.9 = 9$$

- The observed standard deviation is 3 units, i.e. a variance of 9 square units.
- Yes the binomial distribution is useful in this case.

Binomial Example 1

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution: This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

$$P(X = 2) = {}^5C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

0.2 Question 3 : Binomial Distribution

- 10% of intended passengers dont show up. Each flight holds 50 people.
- Binomial distribution with parameters n=50 and p = 0.1
- From Murdoch Barnes Table 1 (third page of tables at back of book)
- a) What is the probability that six people or more fail to show up.

$$P(X = 6) = 0.3839$$

- b) Less than three people fail to show up (i.e. X=0,1 or 2)

$$P(X < 3) = 1 - P(X \geq 3) = 1 - 0.8883 = 0.1117$$

[ANS]

- c) More than 2 but less than 8 (i.e. X = 3,4,5,6,7)

This is equal to

$$P(X = 3) - P(X = 8) = 0.8883 - 0.1221 = 0.7662$$

Binomial Example 1

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution: This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

$$P(X = 2) = {}^5C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

0.3 Binomial Distribution: Worked Example

Suppose a gambler is playing a simple coin flip game. The gambler does not know that the coin has been tampered with such that the probability of a Head is 47%.

Suppose the gambler plays this coin flip game nine times. What is the probability that he wins precisely 3 times.

0.4 Binomial Distribution

MCQ questions - 25% chance of getting a single question right at random.

number of questions is 10

Binomial parameter values $n=10$, $p = 0.25$

X number of correct answers

$$P(X = 7) = 0.0035$$

[00.35%]

0.5 Poisson Questions

0.5.1 Question D01 - Poisson Example

A motor dealership which specializes in agricultural machinery sells on vehicle every 2 days, on average

In this question the unit period is one day. The company expects to sell, on average, 0.5 vehicles every day. The Poisson mean m is therefore 0.5.

$$P(X \geq 1)$$

Go to your Poisson tables, and search for the $m = 0.5$ column. We are interested in the probability of **exactly** one vehicle sold on a particular day. From the tables we can easily work out $P(X \geq 1)$, but this is probability of one or more vehicles being sold. This is not the same thing.

$$P(X \geq 1) = P(X = 1) + P(X = 2) + P(X = 3) + \dots P(X \geq 1) = P(X = 1) + P(X \geq 2)$$

From tables $P(X \geq 1)$ $P(X \geq 2)$

Six day working week? our unit period is now six days. How many vehicles do we expect to sell in 6 days?
answer = 3 $m = 3$

$$P(X \geq 4)$$

$$e^{-6/5} = 0.3011942e^{-4/5} = 0.449329e^{-5/5} = 0.3678794 \quad (1)$$

(c)XYZ Ltd supplies motherboards to Dell. You are a production manager with Dell. There is a constant probability of 0.4 that a board will be defective. You select 100 boards at random. Using the log tables for the binomial distribution what is the probability that

- (i) 0 boards will be defective?
- (ii) 2 or more boards will be defective?
- (iii) 5 or less boards will be defective?
- (d) Use the normal approximation to the binomial to answer (i), (ii) and (iii) in part (c) above.

0.6 Geometric Distribution

) Suppose X has a geometric distribution with parameter p . From the standard interpretation of the geometric distribution, conditioning on whether $X=1$ or not and using the memoryless property of the geometric distribution

[(i)]calculate $E(X)$ calculate $\text{Var}(X)$. [Hint: first calculate $E(X^2)$].

0.7 Exponential

2. Suppose X has an $\text{Exp}()$ distribution. i) Derive $E(X)$ and $\text{Var}(X)$. [Use the fact that $\lim_{x \rightarrow \infty} x^k e^{-x} = 0$, for any positive integer k]. ii) Using induction, show that the k -th moment of X is given by $k!/k$. iii) Show that X has the memoryless property.

The average lifespan of a laptop is 5 years. You may assume that the lifespan of laptop computers follows an exponential distribution.

1. What is the probability that the lifespan of the laptop will be at least 6 years.

$$e^{-6/5} = 0.3011942$$

2. What is the probability that the lifespan of the laptop will not exceed 4 years.

$$e^{-4/5} = 0.449329$$

3. What is the probability that the lifespan of the laptop will be between 5 years and 6 years.

$$e^{-5/5} = 0.3678794$$

0.8 Normal