

Samples from: MINITAB BOOK

Quality and Six Sigma Tools using MINITAB Statistical Software: A complete Guide to Six Sigma DMAIC Tools using MINITAB®

Prof. Amar Sahay, Ph.D.

One of the major objectives of this text is to teach quality, data analysis and statistical tools used in the Six Sigma DMAIC (Define, Measure, Analyze, Improve, and Control) process. The chapters in this book provide concepts, understanding, and computer applications of Six Sigma DMAIC tools. The statistical tools used in the DMAIC process are discussed with step-wise MINITAB computer applications. The following are samples from the book randomly selected from different chapters:

CHAPTER 5: Calculating Distributions: Discrete Distributions and Applications in Six Sigma Computer Applications and Simulations

Chapter Outline

The major objective of this chapter is to gain insight into several of the discrete probability distributions and their properties. A good knowledge and understanding of probability distributions is critical in being able to apply these distributions in data analysis, modeling, quality and process control, and computer simulation. This chapter deals with the following topics:

PROBABILITY DISTRIBUTIONS, DISCRETE PROBABILITY DISTRIBUTION, AND THE DIFFERENCE BETWEEN A PROBABILITY DISTRIBUTION AND A FREQUENCY DISTRIBUTION

- The mean or expected value and the variance of a discrete probability distribution.
- Probabilities from several discrete distributions.
- Random data generation from discrete distributions.
- Experiments and simulations that provide insight into several probability distributions.

BINOMIAL DISTRIBUTION

- Understand the binomial distribution and be able to calculate binomial probabilities.
- Observe how the shape of the binomial distribution changes as we change the characteristic parameters of the distribution.
- Simulate data using a binomial distribution.

POISSON DISTRIBUTION

- Understand the Poisson distribution and be able to calculate the Poisson probabilities.
- Observe how the shape of the Poisson distribution changes as we change the characteristic parameter of the distribution.
- Approximate a Binomial distribution by a Poisson distribution.
- Fit a Poisson distribution to a given set of data.

OTHER DISCRETE DISTRIBUTIONS

- Hypergeometric distribution.
- Negative binomial or Pascal distribution.
- Geometric distribution.
- Multinomial distribution.
- Discrete Uniform distribution.

RANDOM VARIABLES AND DISCRETE PROBABILITY DISTRIBUTION

The concepts of random variable and discrete probability distributions are important in statistical decision making and data analysis. A random variable is a variable that takes on different values as a result of the outcomes of a random experiment. It is a variable that assumes numerical values governed by chance so that a particular value cannot be predicted in advance.

:
:

If the random variable x is either finite or countably infinite, it is a **discrete random variable**. On the other hand, if a random variable takes any value within a given range, it is a **continuous random variable**.

A discrete random variable is a list of all possible outcomes of a random variable, x and their probabilities. There are three methods of describing a discrete random variable: (1) list each value of x – the outcome, and the corresponding probability of x in a table form, (2) use a histogram that shows the outcome of an experiment x on the x-axis and the corresponding probabilities of outcomes on the y-axis, and (3) use a function (or a formula) that assigns a probability to each outcome, x .

PROBABILITY DISTRIBUTION AND FREQUENCY DISTRIBUTION

The probability distribution is a model that relates the value of a variable with the probability of occurrence of that value. *The probability distribution describes the frequencies that occur theoretically; whereas, the relative frequency distribution describes the frequencies that have actually occurred.*

Consider rolling two dice and observing the sum of the numbers on the top faces. Suppose x is a **random variable** that denotes the sum of the numbers on the top faces. The theoretical probability distribution of this experiment is known and is shown in Table 5.1 below. This is an example of a discrete probability distribution.

:

Table 5.1

x	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

:

:

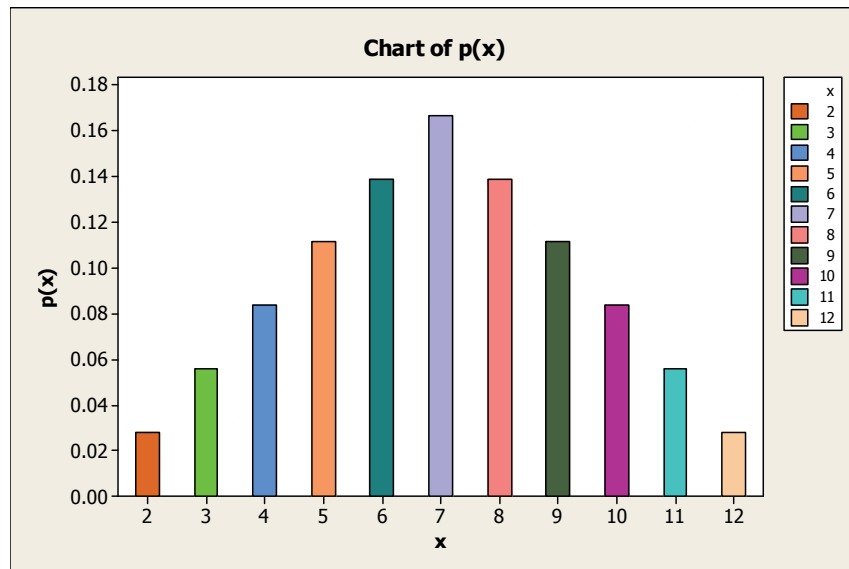


Figure 5.1: Probability Distribution of Rolling Two Dice

The following are the two requirements for the probability distribution of a discrete random variable:

1. $P(x)$ is between 0 and 1 (both inclusive) for each x , and
2. $\sum P(x) = 1.0$ (5.1)

DIFFERENCE BETWEEN A PROBABILITY DISTRIBUTION AND A FREQUENCY DISTRIBUTION

Using a statistical computer software (MINITAB) and the theoretical probability distribution (Table 5.1), we simulated 1000, 5000 and 10,000 trials of throwing two dice. We then compared the simulated results with the theoretical results in Table 5.1 and saw how the theoretical results (probability distribution) differ from the simulated results.

Steps to Perform Simulation.....

:

Results:

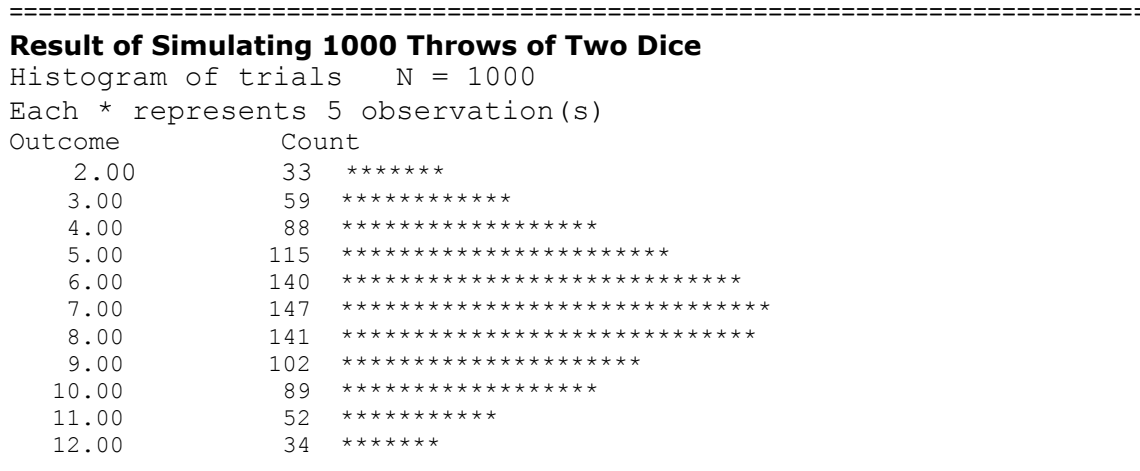


Figure 5.2: Result of Simulating 1000 Throws of Two Dice

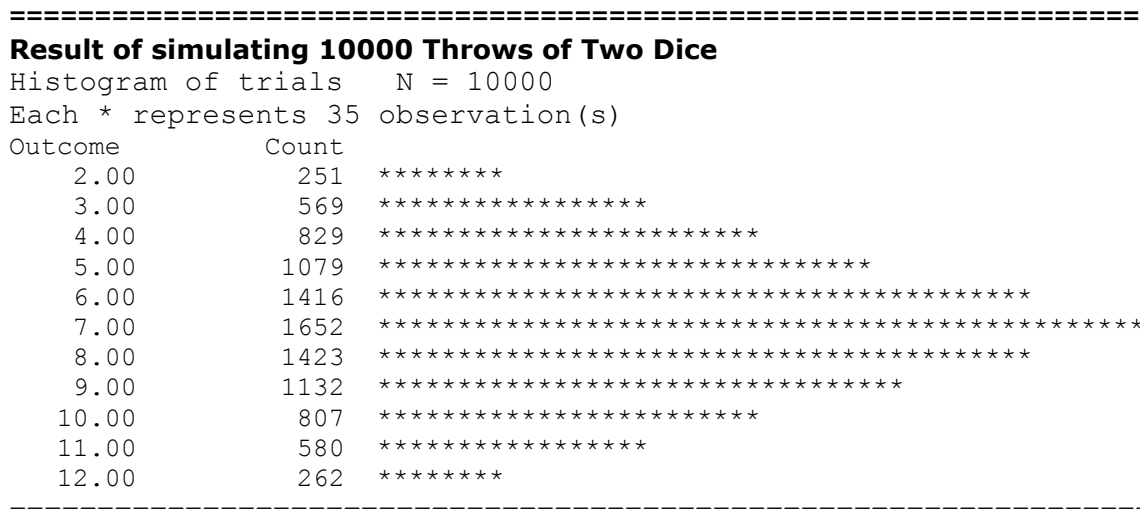


Figure 5.4: Result of Simulating 10000 Throws of Two Dice

Table 5.2: Comparing Probability and Frequency Distribution

Probability Distribution	Frequency Distribution
---------------------------------	-------------------------------

Row	x Outcome	P(x=x) Prob. Dist.	Rel.Freq1 1000 Trials	Rel.Freq2 5000 Trials	Rel.Freq3 10,000 Trials
1	2	0.0278	0.033	0.0300	0.0251
2	3	0.0560	0.059	0.0582	0.0569
3	4	0.0833	0.088	0.0858	0.0829
4	5	0.1111	0.115	0.1098	0.1079
5	6	0.1389	0.140	0.1402	0.1416
6	7	0.1670	0.147	0.1710	0.1652
7	8	0.1389	0.141	0.1370	0.1423
8	9	0.1111	0.102	0.1038	0.1132
9	10	0.0833	0.089	0.0774	0.0807
10	11	0.0560	0.052	0.0590	0.0580
11	12	0.0278	0.034	0.0278	0.0262

AN APPLICATION OF EXPECTED VALUE: SIMULATING THE GAME OF CRAPS

:
:
:
:

SOME IMPORTANT DISCRETE DISTRIBUTIONS: BERNOULLI PROCESS AND BINOMIAL DISTRIBUTION

:
:
:

BINOMIAL DISTRIBUTION

Background: A random variable that denotes x number of successes in n Bernoulli trials is said to have a Binomial distribution in which the probability of x successes is given by the following expression:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{where, } x = 0, 1, \dots, n \quad (5.6)$$

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{where, } x = 0, 1, \dots, n \quad (5.7)$$

In the above expression,

$p(x)$ = probability of x number of successes

n = number of trials

p = probability of success

$(1 - p) = q$ is the probability of failure

CALCULATING BINOMIAL PROBABILITIES USING MINITAB OR BINOMIAL PROBABILITY TABLE

To calculate the Binomial probabilities using MINITAB when $n=10$ and $p=0.05$, label the columns C1 and C2 of MINITAB work sheet as shown below

C1	C2
x	p(x)

Type the numbers 0 through 10 in column C1 or use the command sequence

Calc ⇨ **Make Patterned Data** ⇨ **Simple Set of Numbers**

:

:

Next, select the following command sequence:

Calc ⇨ **Probability distributions** ⇨ **Binomial ...**

In the Binomial distribution dialog box,

Click the circle to the left of Probability

Number of trials 10

Probability of success 0.05

Click the circle to the left of Input column and type ...

Optional storage ...

Click **OK**

The probabilities for $n=10$ and $p=0.05$ will be calculated and stored in column C2 as shown below.

Row	C1	C2
	x	p(x)
1	0	0.598737
2	1	0.315125
3	2	0.074635

4	3	0.010475
5	4	0.000965
6	5	:
		:
10	9	0.000000
11	10	0.000000

These probabilities are similar to the probabilities obtained from the Binomial table above.

You may also calculate Binomial probabilities by typing the following commands in the session window of MINITAB.

Click anywhere in the Session window of MINITAB to make it active; then use the command sequence

Editor ⇌ Enable Commands

You will see MTB > on the window. This enables you to write the commands to calculate the probabilities. Type the following commands:

MTB > pdf; (Hit enter key)

SUBC > Binomial 10 0.05. (Hit enter key)

The Binomial probabilities for $n=10$ and $p=0.05$ will be calculated and displayed on the session window.

Probability Density Function

Binomial with $n = 10$ and $p = 0.0500000$

x	P(X = x)
0	0.5987
1	0.3151
:	
:	
5	0.0001
6	0.0000

EXPLORING THE BINOMIAL DISTRIBUTION

Objective: In this section we will demonstrate how the shape of the Binomial Distribution changes when we change the characteristic parameters of the distribution. The parameters of the Binomial Distribution are the number of trials (n), and the probability of success (p).

Experiment 1: In this experiment, we generated 500 random numbers from the Binomial distribution with number of trials, $n=20$ and various values of probability of success, p ($p=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$). Observe how the shape of the distribution changes (see Figure 5.9 below).

Steps...

:

Results:

:

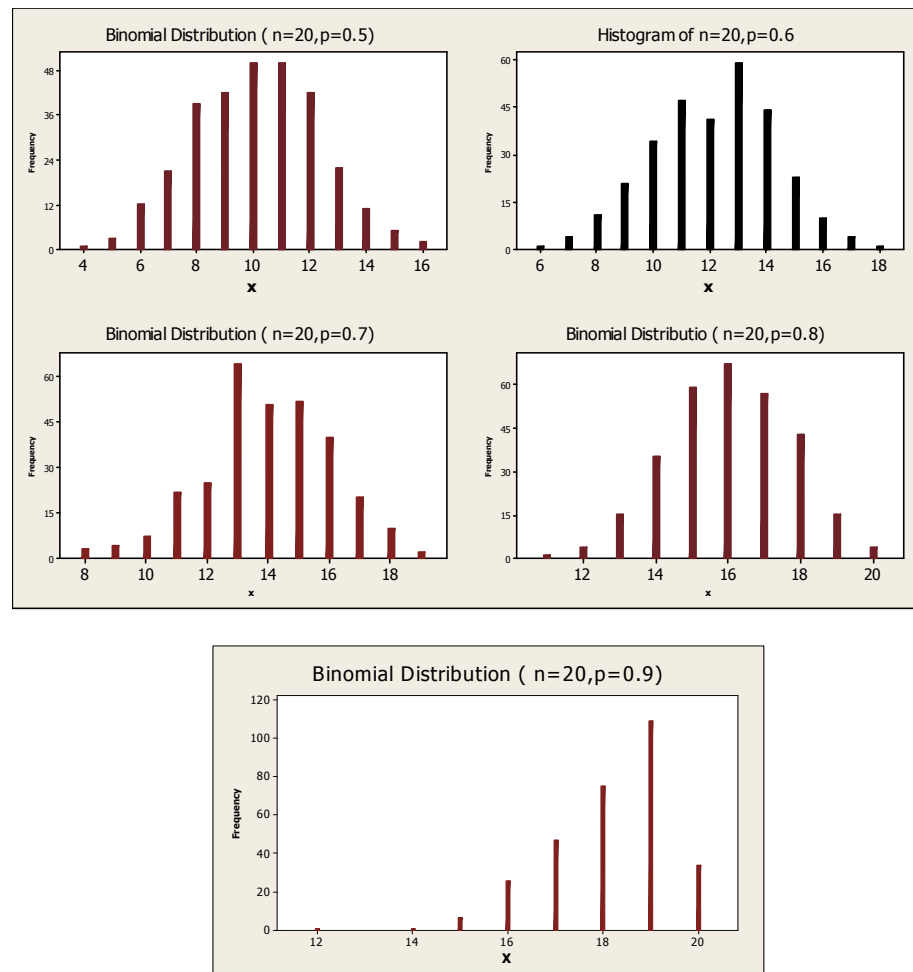


Figure 5.9: Binomial Distribution (fixed n , varying p)

Conclusion: From Figure 5.9 above, you can see that when

- the parameter p is less than 0.5, the shape of the distribution is skewed to the right.
- p exceeds 0.5, the shape of the distribution is skewed to the left, and
- $p=0.5$, the shape of the distribution is approximately symmetrical.

Experiment 2: Generate 300 random numbers from the Binomial distribution with the probability of success $p=0.3$, and various values for the number of trials n , ($n = 10, 15, 20, 25, 30, 35, 40, 50$). Observe how the shape of the distribution changes (see Figure 5.10 below).

Steps...
Results:

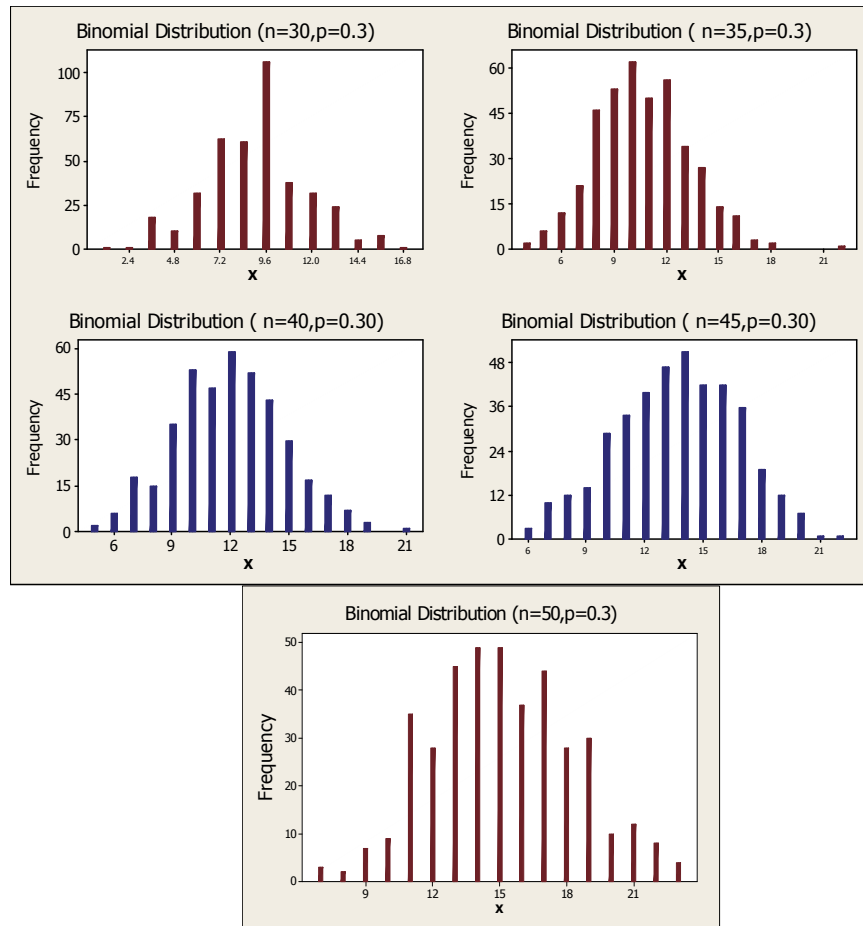


Figure 5.10: Binomial distribution (Varying n , fixed p)

Conclusion: Comparing the shape of the above distributions for the different values of n , it can be seen that if the probability of success p is held constant and the sample size n is increased, the sum of Bernoulli variables increases. The Binomial distribution becomes more and more symmetrical by virtue of the central limit theorem. Also, as n increases, it is not easy to work with the Binomial distribution (the probability calculations become difficult and time consuming). Thus, for a large n the Binomial distribution may be approximated by a Normal distribution.

More Discrete Distributions in this chapter....

THE POISSON DISTRIBUTION

Background: A random variable X is said to follow a Poisson distribution if it assumes only nonnegative values and its probability density function is given by:

$$p(x) = \frac{e^{-\mu} \mu^x}{x!} \quad \text{where, } x = 0,1,2,\dots,n \quad (5.11)$$

Where, μ represents the mean and variance of the distribution. Note that $\mu > 0$.

The Poisson distribution occurs when there are events which do *not* occur as outcomes for a fixed number of trials of an experiment (unlike that of the Binomial distribution), but which occur at random points of time and space. **The Poisson distribution is the correct distribution to apply when n is very large (that is, the area of opportunity is very large) and an event has a constant and very small probability of occurrence.**

INVESTIGATING THE POISSON DISTRIBUTION

Steps...
Results:

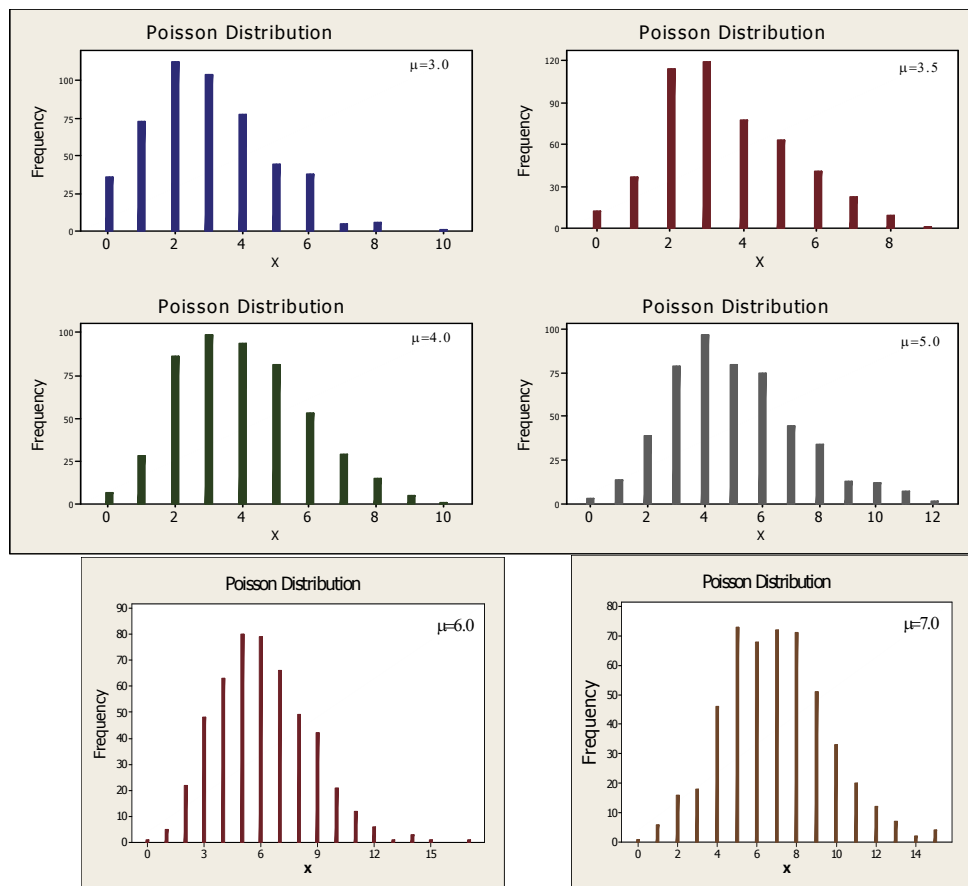


Figure 5.13: The Poisson distribution with different μ Values

Conclusion: As you can see from figures above, the Poisson distribution becomes more and more symmetrical as the value of the mean (μ), becomes 6 and higher. For the values of μ below 6, the shape is skewed to the right. Note that the parameter μ is also known as the shape parameter; a change in this parameter changes the shape of the probability density function of the Poisson distribution. Thus, for the Poisson distribution:

Shape parameter: μ
 Range of the parameter μ : $\mu > 0$
 Expected value: μ (5.12)
 Standard deviation: $\sigma = \sqrt{\mu}$ (5.

Example 15 (Calculating Poisson probabilities using MINITAB)

To calculate the Poisson probabilities using MINITAB when the average, $\mu = 3.1$, label the columns C1 and C2 of MINITAB work sheet as shown below

C1	C2
x	p(x)

Type the numbers 0 through 14 in column C1 or use the command sequence

Calc ⇨ Make Patterned Data ⇨ Simple Set of Numbers

Complete the dialog box that is displayed as shown below

Store patterned data in	x	
From first value	0	
To last value	14	
In steps of	1	
List each value	1	time
List the whole sequence	1	time

Click **OK**

The numbers 0 through 14 will be stored in column C1. Next, select the following command sequence:

Calc ⇨ Probability distributions ⇨ Poisson

In the Poisson distribution dialog box,
 Click the circle to the left of Probability
 Mean

Click the circle to the left of Input column and type x or C1
Optional storage $p(x)$ or C2
Click **OK**

The probabilities for $\mu = 3.1$ will be calculated and stored in column C2 as shown below.

Table 5.12: Poisson Probabilities for $\mu = 3.1$

Row	x	$p(x)$
1	0	0.045049
2	1	0.139653
3	2	0.216461
4	3	0.223677
5	4	0.173350
6	5	0.107477
7	6	0.055530
8	7	0.024592
9	8	0.009529
10	9	0.003282
11	10	0.001018
12	11	0.000287
13	12	0.000074
14	13	0.000018
15	14	0.000004

These probabilities are similar to the ones obtained from the Poisson table (Table 5.8).

You may also calculate Poisson probabilities by typing the following commands in the session window of MINITAB:

- Click anywhere in the Session window of MINITAB to make it active. Then use the command sequence

Editor → Enable Commands

You will see MTB > on the window. This enables you to write the commands to calculate probabilities. Type the following commands:

MTB > pdf; (Hit enter key)

SUBC > Poisson 3.1. (Hit enter key)

The Poisson probabilities for $\mu = 3.1$ as shown in Table 5.12 above will be calculated and displayed on the session window.

Approximating the Binomial distribution with the Poisson distribution (Poisson's Approximation)

Other Discrete Distributions:

Hypergeometric Distribution

Relationship between Binomial and Hypergeometric distribution

Hypergeometric distribution tends to binomial distribution as $N \rightarrow \infty$ and $D/N = p$. Also, if the sampling is done without replacement and the sample size n is less than 5% of the population ($n < 0.05N$), the binomial distribution can be used to approximate the hypergeometric distribution.

Example 25

A manufacturer of turn-indicator lights for automobiles finds out that out of a shipment of 4000 turn-indicator lights sent to a distributor, 1000 are slightly defective. If a retailer purchases 10 of these turn-indicator lights, what is the probability that 2 of these are defective?

Since the population size, $N=4000$ is large compared to the sample size, we can approximate the probability using the binomial distribution. The required probability $p(x=2)$ is calculated below using the binomial distribution.

$$n=10, \quad p = 1000/4000 = 0.25$$

The probability using MINITAB is

$$p(x=2) = 0.2818$$

To calculate the above probability, type the commands below in MINITAB session window.

Binomial using MINITAB

```
TB > PDF 'x' C2;
```

```
SUBC > Binomial 10 0.25.
```

x	p(x)
0	0.056102
1	0.187571
2	0.281826

The probability, $p(x=2)$ was also calculated using the hypergeometric distribution (see the table below). This probability is

$$p(x=2) = 0.2816$$

To calculate the above probability, the MINITAB instructions are

MTB > PDF 'x' c3;

SUBC > Hypergeometric 4000 1000 10.

x	p(x)
0	0.056314
1	0.187712
2	0.281568

Note that the probabilities obtained using the two distributions are very close.

:
:

Negative Binomial or Pascal Distribution

The negative binomial distribution can be derived from the binomial distribution with some modifications. The basic difference between the binomial and the negative binomial distribution is that the binomial distribution is used to calculate the probability of x number of successes out of n trials where the number of trials is fixed; whereas in a negative binomial distribution the trials are repeated until a fixed number of successes occur. We are interested in finding the probability that the r^{th} success occurs on the x^{th} trial.

:
:

The probability of r^{th} success occurring on the x^{th} trial is also written as

$$p(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}; x = r, r+1, r+2, \dots \quad (5.19)$$

The mean and variance of the negative binomial distribution is given by...

For the Poisson distribution discussed earlier, we have seen that the mean and the variance are equal. The equality of mean and the variance is an important characteristic of the Poisson distribution. For the binomial distribution, the mean is always greater than the variance. In some cases, the observable phenomenon gives rise to empirical distributions in which the variance is larger than the mean. *In cases where the variance is larger than the mean, the negative binomial distribution provides a good model.*

Example 28

Suppose you are tossing three coins five times and want to find the probability of getting all heads or all tails on the first trial. This probability can be calculated using the negative binomial. Note that $x = 5$, $r = 1$, $p = 1/4 = 0.25$.

$$b^-(x; r, p) = \binom{x-1}{r-1} p^r q^{x-r}; x = r, r+1, r+2, \dots$$

$$b^-(5; 1, 0.25) = \binom{5-1}{1-1} p^1 q^{5-1} = \frac{4!}{0!(4-0)!} (0.25)^1 (0.75)^4 = 0.0791$$

Thus, there is 7.9% chance of getting all heads or all tails on the first trial. We calculated the probabilities of getting all heads or all tails for the 2nd through 5th trials if the coins are tossed 5 times. These probabilities are shown below.

x	r	p(x)
5	1	0.079102
5	2	0.105469 ➤ probability of getting all heads or all tails on the 2 nd trial
5	3	0.052734
5	4	0.011719
5	5	0.000977

The probabilities are plotted in Figure 5.20. The Pascal or the negative binomial distribution is used in quality control in the area of sequential sampling.

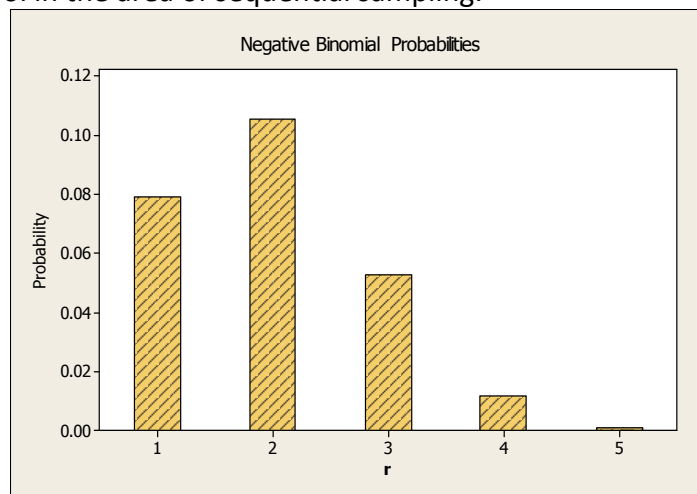


Figure 5.20: Negative Binomial Probabilities

The Geometric Distribution

The geometric distribution is also related to a sequence of Bernoulli trials in which the random variable X takes two values 0 and 1 with the probability q and p respectively, that is, $p(X=1) = p$, $p(X=0)=q$, and $q = 1-p$. To compare the Binomial distribution with the geometric distribution, recall that the Binomial distribution describes discrete data resulting from an experiment known as a **Bernoulli** process. The binomial distribution is used to calculate the probability of x successes out of n trials where the trials are independent of each other and the probability of success p remains constant from trial to trial.

In geometric distribution, the number of trials is not fixed, and the random variable of interest x is defined as the number of trials required to achieve the first success.

The geometric distribution can be derived as a special case of negative binomial distribution. In the probability density function of negative binomial distribution [equation (5.19)] above, if $r = 1$, we get the probability distribution for the number of trials required to achieve the first success.

If we have a series of independent trials that can result in a success with probability p and a failure with probability q where, $q=1-p$, then the random variable X that denotes the number of trials on which the first success occurs, is given by

$$p(x; p) = \begin{cases} pq^{x-1} & \text{where, } x = 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases} \quad (5.22)$$

The distribution is called geometric because the probabilities for $x=0, 1, 2, \dots$, are the various terms of geometric progression.

Example 29

An assembly line produces computers in which three out of every 25 computers are known to be defective. Each computer coming off the line is inspected for defects. What is the probability that every 5th computer inspected is the first defective found?

Using the geometric distribution with $x=5$ and $p = 3/25 = 0.12$, we can calculate the required probability as

$$p(x; p) = pq^{x-1}$$
$$p(5; 0.12) = 0.12(0.88)^4 = 0.0720$$

There is 7.20% chance that the 5th computer inspected will be the first defective found. Below, we have calculated the probabilities of 1 through 10 occurrences, that is; the probability that the first computer inspected was found defective; the probability that the second computer inspected was the first one defective, and so on. The results are shown below:

x	p	p(x;p)
1	0.12	0.120000
2	0.12	0.105600
3	0.12	0.092928
4	0.12	0.081777
5	0.12	0.071963
6	0.12	0.063328
7	0.12	0.055728
8	0.12	0.049041
9	0.12	0.043156
10	0.12	0.037977

These probabilities are plotted below in Figure 5.21.

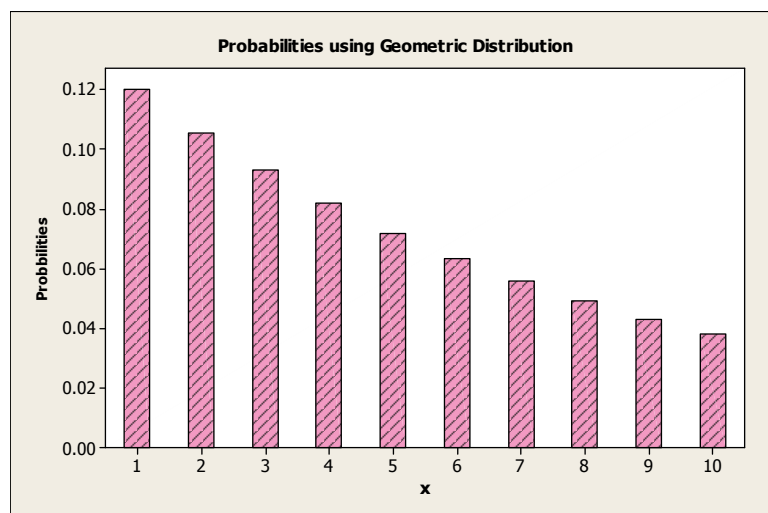


Figure 5.21: Probabilities from Geometric Distribution

Multinomial Distribution
Discrete Uniform Distribution

⋮
⋮

Continued...