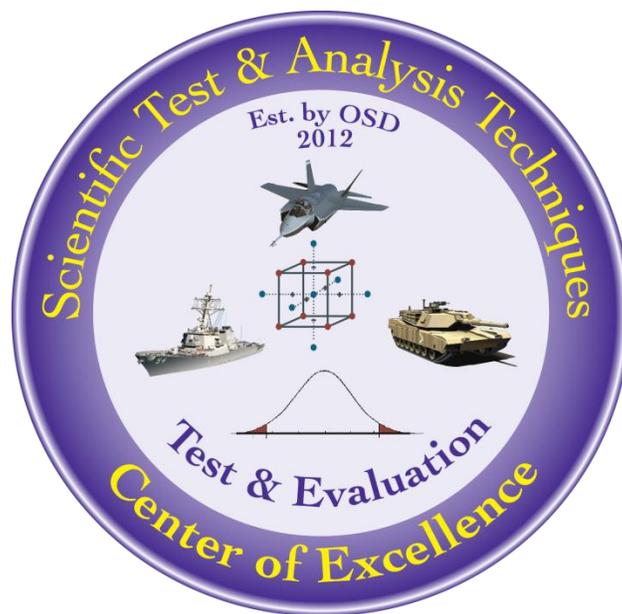


Statistics Reference Series

Part 2: Probability

Authored by: Gina Sigler

31 October 2018



The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

Table of Contents

Executive Summary.....	2
Introduction	2
Probability.....	2
Independence	2
Mutually Exclusive	2
Probability Distributions	3
Binomial	3
Poisson	4
Hypergeometric	5
Normal	6
Student's t	7
Chi-squared.....	8
F	9
Exponential	10
Weibull	11
Central Limit Theorem	12
Sample Mean vs. Population Mean	13
Conclusion.....	15
References	15
Appendix A: Distribution Formulas.....	16

Executive Summary

This is the second in a four-part series of statistics reference documents. This part is meant as a refresher on probability and a small set of commonly used probability distributions. The probability distributions described in this document are those found helpful in engineering and reliability applications. This is not meant to be an exhaustive list of probability concepts or distributions.

Keywords: probability, probability distribution, continuous, discrete

Introduction

Probability, in the simplest form, is the number of times something occurs out of the number of times it could have occurred. This is most useful to statisticians in terms of a long-run frequency where the probability is the proportion of times that something happens if an experiment or action is repeated an infinite number of times. Most often probability will be used to relate the chance of something occurring in the future. Probability distributions can often be helpful to model many different possibilities of a future event.

Probability

All probabilities have some common features. First, a probability is represented by either a fraction, percentage, or decimal. Second, a probability must always be a value between 0 and 1 or 0% to 100%. These two values represent the two extremes of future knowledge. An event with a 0% probability means the event has absolutely no chance of happening, and a 100% probability means the event is certain to happen. Finally, all probabilities belonging to the same set will sum to 1. For more information about rules of probability, please see any of the reference books located at the end of this document.

Independence

Statistical independence is a bit different than the standard definition of independence. Independence is most often explained in terms of probability. If two events are independent, the occurrence of one event has no impact on the occurrence of another; the probability of the second event occurring does not change as a result of the first event occurring. A very basic example is flipping a coin twice. The results of the first flip have no impact on the results of the second flip. Often, it is possible to assume events to be effectually independent, meaning there is so little impact of one outcome on another that strict independence is not necessary.

Mutually Exclusive

The term mutually exclusive is often mistaken for independence. Mutually exclusive means that if one event occurs, another cannot possibly occur; if one event occurs, the probability of the other event occurring is 0. Sometimes mutually exclusive events are referred to as disjoint. Continuing with the coin

flipping example, heads and tails are mutually exclusive outcomes in a single flip. If a head occurs, then a tail cannot possibly occur on that same flip.

Probability Distributions

A probability distribution is a mathematical representation of the sample space (all possible values of an experiment) and their associated probabilities. This portrayal can be in the form of a graph, table, or a mathematical function depending on whether the data is discrete or continuous. For discrete data, so called because the data can only take a finite list of values, probability distributions are referred to as probability mass functions (pmfs). These discrete distributions are often displayed as tables, graphical representations of possible values of the data and associated probabilities, or mathematical formulas. For continuous data, which can take any value in a range, probability distributions are instead referred to as probability density functions (pdfs). These are often represented as mathematical functions or graphs where probabilities are calculated on an interval instead of for a single point. Multiple examples of common types of probability distributions are given below, and formulas for a majority of the distributions can be found in Appendix A.

Binomial

The binomial distribution is an example of a discrete probability distribution function. In this case, the collected data will have only two possible outcomes and follow a binomial experiment. A binomial experiment is defined by the following properties:

1. One outcome can only take on two possible values, often labeled as a “success” or a “failure”
2. There are a fixed number of data points, identified as n , also referred to as the number of trials
3. The probability of a “success” is labeled p and is constant for each trial
4. All trials are independent

The binomial distribution is constructed around the number of successes (referred to as y) in n trials. Labeling something as a “success” doesn’t mean it is necessarily a good thing, it just marks the variable of interest or the variable you are counting. For example, the most popular color for a car is white, and approximately 25% of all cars in the United States are white. Even, though there are other car colors, this particular experiment is only interested in the binary outcomes of having a car be white or not white. If 25% of the cars are white, 75% of the cars are not white. If a random sample of 20 cars is collected, the number of white cars can be modeled by the binomial distribution as seen in Figure 1.

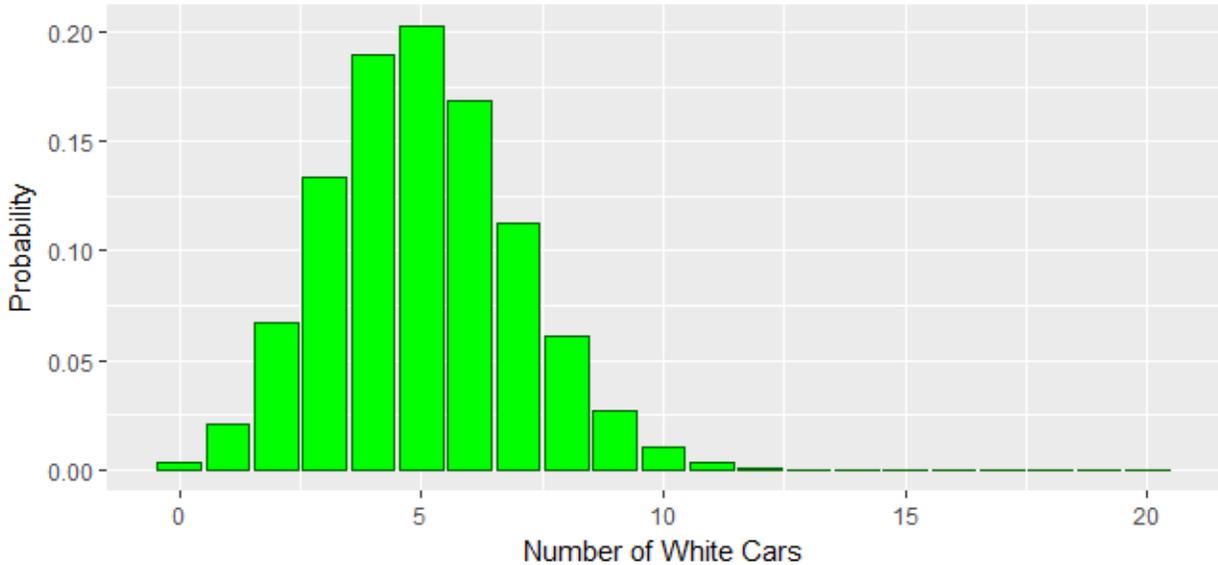


Figure 1: Binomial probability distribution showing the probability of having a certain number of white cars in a random sample of 20 cars ($n = 20$ and $p = 0.25$)

The shape of the distribution will change depending on the probability of success, p , and the number of trials, n . In a probability histogram, as in Figure 1 above, the area of the bars will always add to 1.

Poisson

The Poisson distribution is another example of a discrete probability distribution. This distribution is often used to model the number of events that occur during a given period of either time or space (although space is less common). Properties of the Poisson distribution are outlined below:

1. The number of events that occur in a given time frame or space, y , must be random and independent
2. The number of events that are expected to occur in one time period (or space) is the same as another time period (or space)
3. Two events cannot occur at the same exact time or take up the same exact space

Typically, the Poisson distribution is used to model data in the form of counts. Consider a machine that prints the “m” on top of M&Ms; it is expected to make 4 mistakes within a typical 1 hour period. The mistakes are considered to be independent from one another as the misprinting of one “m” should not change the printing on another. The rate is also considered to be constant for this machine. The distribution of the number of mistakes that the machine makes in an hour would be Poisson and can be seen in Figure 2.

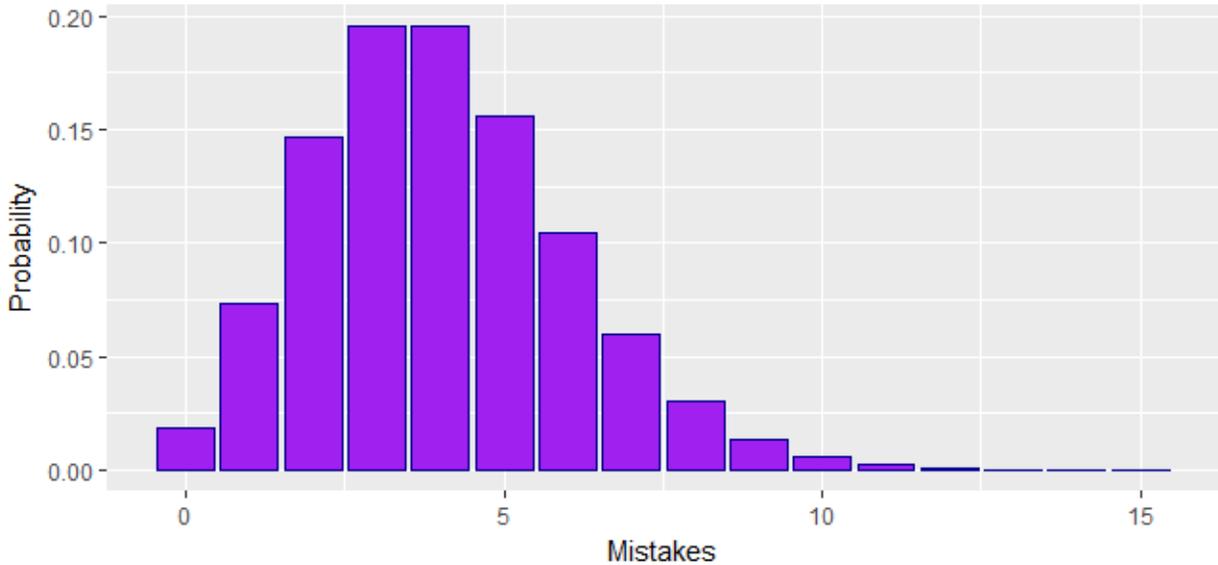


Figure 2: Poisson probability distribution showing the probability of having a certain number of mistakes by the machine in 1 hour, for an expected number of 4

The shape of the distribution will change depending on the number of events expected to occur in that given time frame. However, the area of the bars on the probability histogram will still always add to 1.

Hypergeometric

The hypergeometric distribution is another example of a discrete probability distribution. This distribution is similar to the binomial distribution in that data are recorded in two distinct categories, with the count of interest defined as y . However, the hypergeometric distribution assumes that the data points gathered are not independent and are collected from a relatively small population. Properties necessary for a hypergeometric distribution are outlined below:

1. There are only a certain number of successes possible out of a finite population
2. One outcome is labeled as a “success” and the other as a “failure”
3. The probability of a success changes for each data point collected

The hypergeometric distribution is again constructed around the number of successes, y . Part 1 of the statistics reference series showed a bar chart featuring M&M colors. In that example, the bag of M&Ms contained 104 M&Ms; 18 of those M&Ms were red. If a handful of 10 M&Ms is taken, a hypergeometric distribution can be constructed around the number of red M&Ms in that handful as seen in Figure 3. Because a handful of M&Ms were taken from a finite population (the bag of M&Ms), each M&M was sampled without replacement.

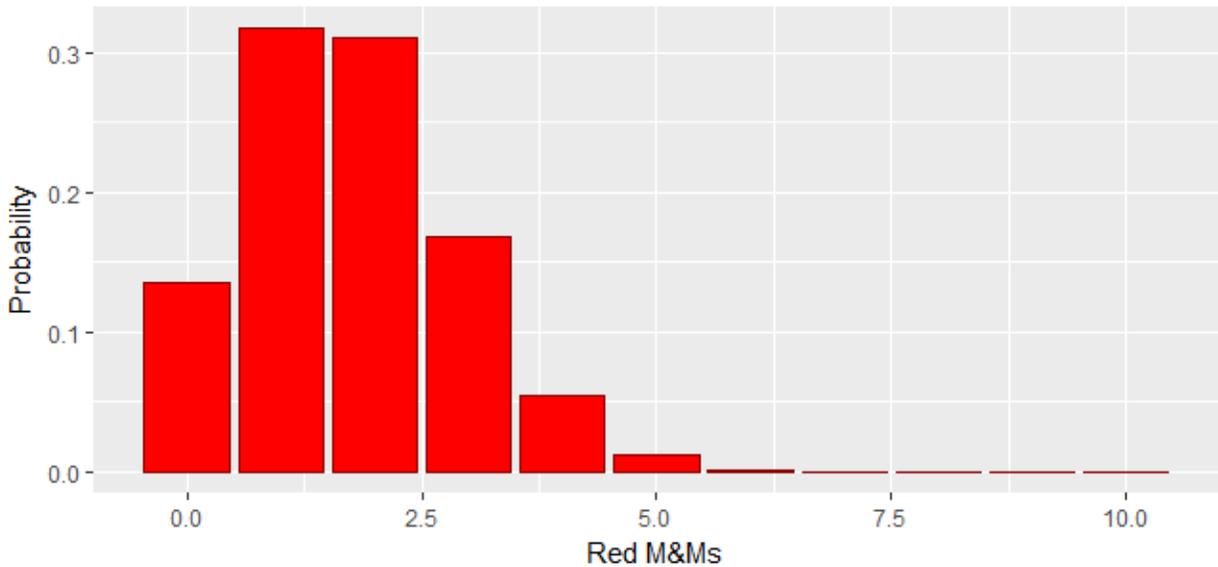


Figure 3: Hypergeometric probability distribution showing the probability of having a certain number of red M&Ms in a handful of 10 taken from a bag of 104 M&Ms with 18 red total

The shape of the distribution will change depending on the total number of items in the population, the number of possible successes in the population, and the number of data points in the sample. In every case though, the area of the bars in a probability histogram will still add to 1.

Normal

The normal distribution is the most well-known of the continuous probability distributions. It is also one of the most widely used continuous probability distributions. This is partly due to the Central Limit Theorem, which is explained later in this document. Probabilities of the normal distribution can be represented as the area under the curve or calculated directly from the probability density function.

Properties of the normal distribution are outlined below:

1. The distribution is symmetric and bell-shaped
2. The mean, median, and mode are all equivalent and located in the middle of the bell
3. The distribution is completely defined by its mean and standard deviation
4. The total area under the curve is equal to 1
5. The curve spans values from $-\infty$ to $+\infty$

Continuous distributions are a little different from discrete distributions in that possible values span a range of numbers. With infinite possible outcomes, the probability of any single value occurring is equal to 0. Instead, the probability between two values is calculated, in which case the probability is equal to the area under the curve between those two values. Suppose it is known that the distribution for the miles per gallon (mpg) of all used motor vehicles is considered normal with a mean of 23.7 mpg and a standard deviation of 4.6 mpg. Figure 4 shows the probability distribution for the mpg of used motor vehicles.

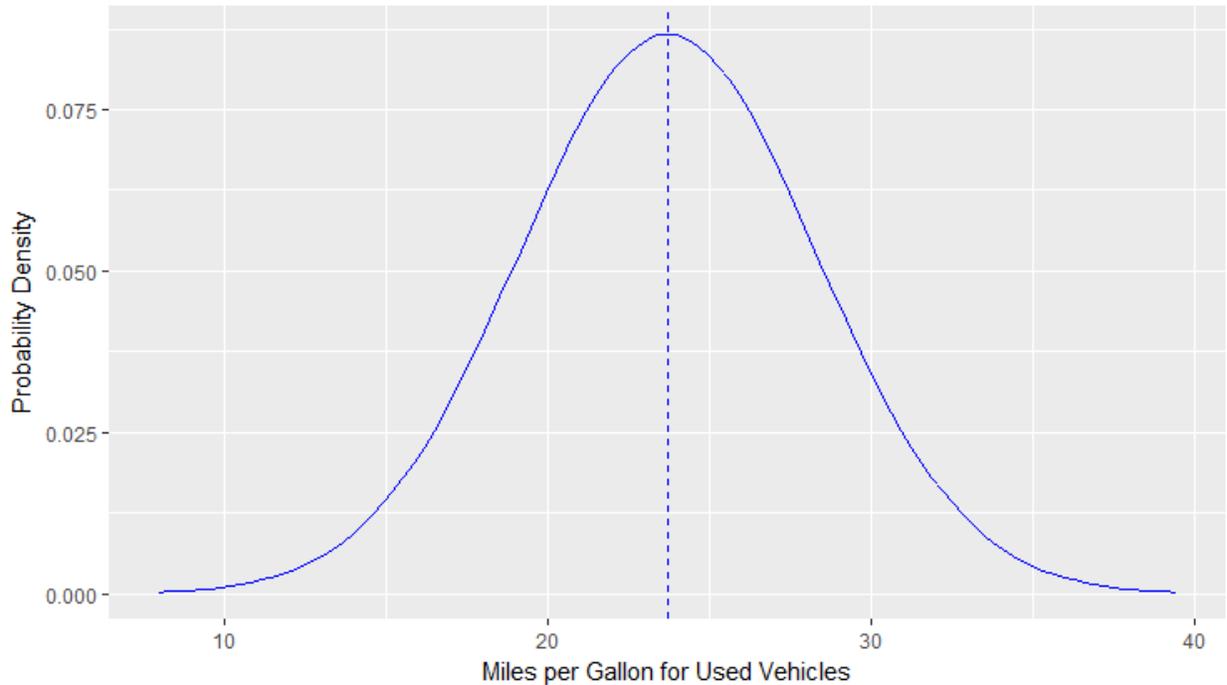


Figure 4: Normal probability distribution for the miles per gallon of used motor vehicles with a mean of 23.7 mpg and a standard deviation of 4.6 mpg

The overall shape of the normal distribution is always a bell curve no matter the values of the data, but the height and width of the bell will change depending on the standard deviation of the data. A smaller standard deviation means a narrower bell, and a larger standard deviation means a wider bell. Changing the mean only slides the bell curve left or right along the x-axis. Actual probabilities can be calculated using a computer or a z-table found in any of the statistics books listed in the reference section.

Student's *t*

One of the drawbacks of the normal distribution is that it assumes there is some previous knowledge about the population standard deviation. However, this is fairly unrealistic in practice. Researchers would have no way of knowing the standard deviation for the population; however, educated guesses can be made in the form of a sample standard deviation. There are some cases where the normal distribution may be used anyway (see Central Limit Theorem section below). But, there is an alternative for data believed to be from a normal population with an unknown standard deviation known as the student's *t*-distribution or *t*-distribution. The *t*-distribution is a theoretical distribution that is used for inferential statistics and has similar properties to the normal distribution:

1. The distribution is symmetric and bell-shaped
2. The mean, median, and mode are all equivalent and located in the middle of the bell (at 0)
3. The total area under the curve is equal to 1
4. The curve spans values from $-\infty$ to $+\infty$ (although tails appear fatter than normal distribution)

The main difference is that the t -distribution is defined by only one parameter, dependent on the sample size. The sample size will determine the shape of the bell, and the definition of degrees of freedom will change based on the use of the distribution. The t -distribution resembles a standard normal distribution that has been slightly squished on top so that the tails of the distribution become larger as seen in Figure 5. Note that a standard normal distribution is just a distribution that has been standardized so that it has a mean of 0 and a standard deviation of 1.

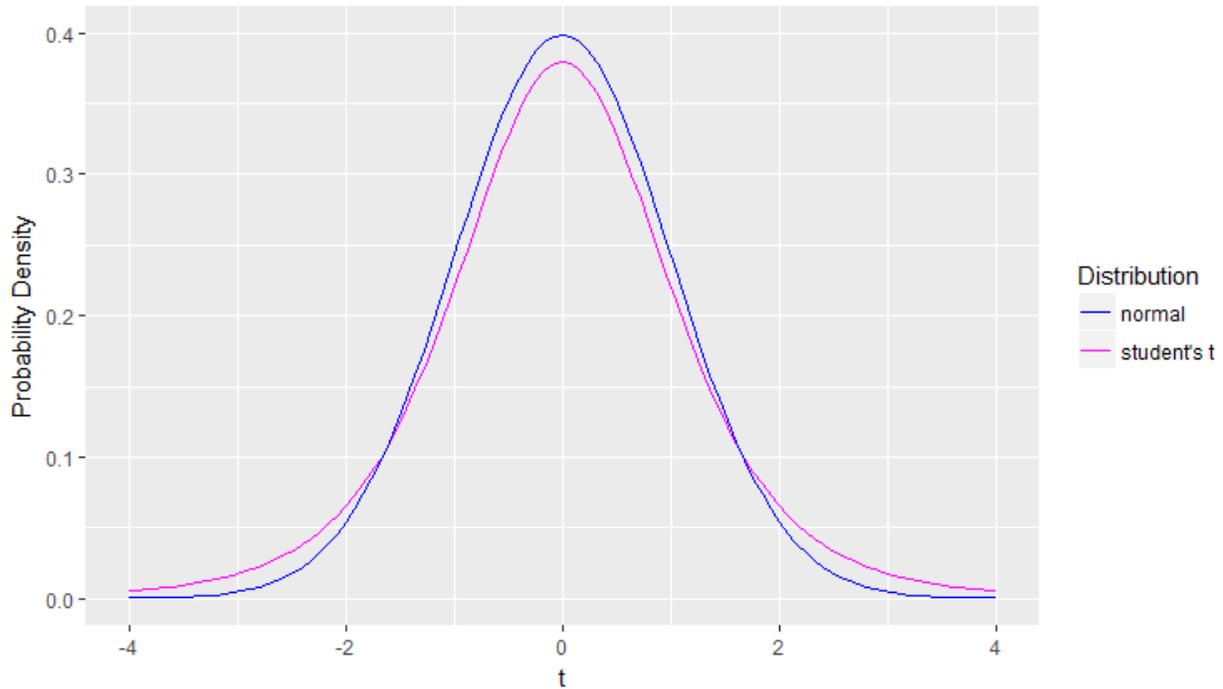


Figure 5: Student's t -distribution with 5 degrees of freedom is shown in pink along with a standard normal probability distribution in blue

This additional variation in the tails of the distribution is dependent on the degrees of freedom. As the degrees of freedom approach infinity, the t -distribution approaches the normal distribution. Probabilities are most typically calculated using a computer or a t -table.

Chi-squared

Another popular continuous distribution is the chi-squared distribution (χ^2). This distribution is often used for either hypothesis tests involving categorical data analysis or testing population variances. Probability can still be represented by the area under the curve or calculated from the associated probability density function. Properties of the chi-squared distribution are outlined below:

1. The distribution is asymmetric with its shape dependent on one parameter (called the degrees of freedom)
2. The mean is equal to the degrees of freedom
3. The standard deviation is equal to the degrees of freedom times 2

4. The total area under the curve is equal to 1
5. The curve spans values 0 and $+\infty$ (never a negative value)

As with the t -distribution, the chi-squared distribution is highly dependent on the degrees of freedom. The definition of degrees of freedom will change based on the use of the distribution. Similar to the t -distribution, when the degrees of freedom approach infinity, the chi-squared distribution begins to closely resemble the normal distribution. An example of a chi-squared distribution with 5 degrees of freedom is seen in Figure 6.

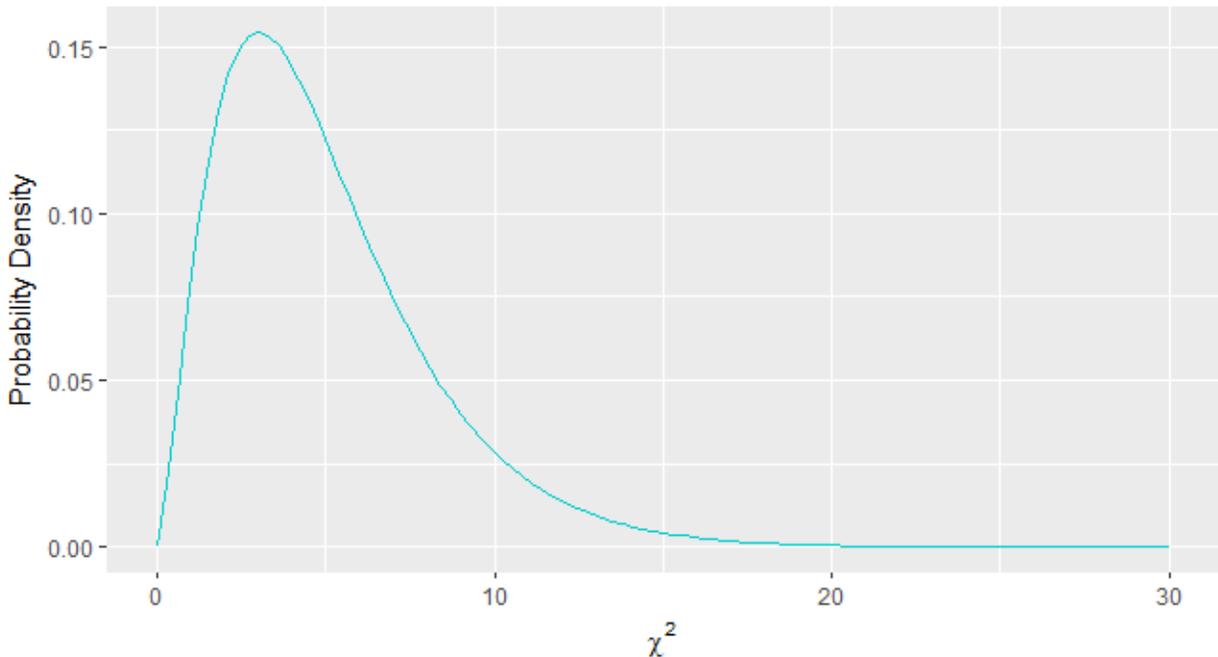


Figure 6: Chi-squared (χ^2) probability distribution with 5 degrees of freedom

The chi-squared distribution is unique in that it is only determined by the degrees of freedom, which then determine the distribution mean and variance. As usual, probabilities are calculated using a computer or a chi-square table.

F

There is another continuous distribution that is related to the chi-squared distribution known as an F -distribution. The F -distribution is most commonly used for Analysis of Variance (ANOVA) testing or comparing two population variances. As with all continuous distributions, the probability is represented by the area under the curve or calculated from the associated probability density function. Properties of the F -distribution are outlined below:

1. The distribution is asymmetric with its shape dependent on two parameters (two sets of degrees of freedom)
2. The total area under the curve is equal to 1

3. The curve spans values 0 and $+\infty$ (never a negative value)

The shape is defined by the two sets of degrees of freedom that are determined from two independent chi-squared distributions. The shape of the F -distribution can change dramatically as the two different degrees of freedom values change. One example of an F -distribution can be seen below in Figure 7.

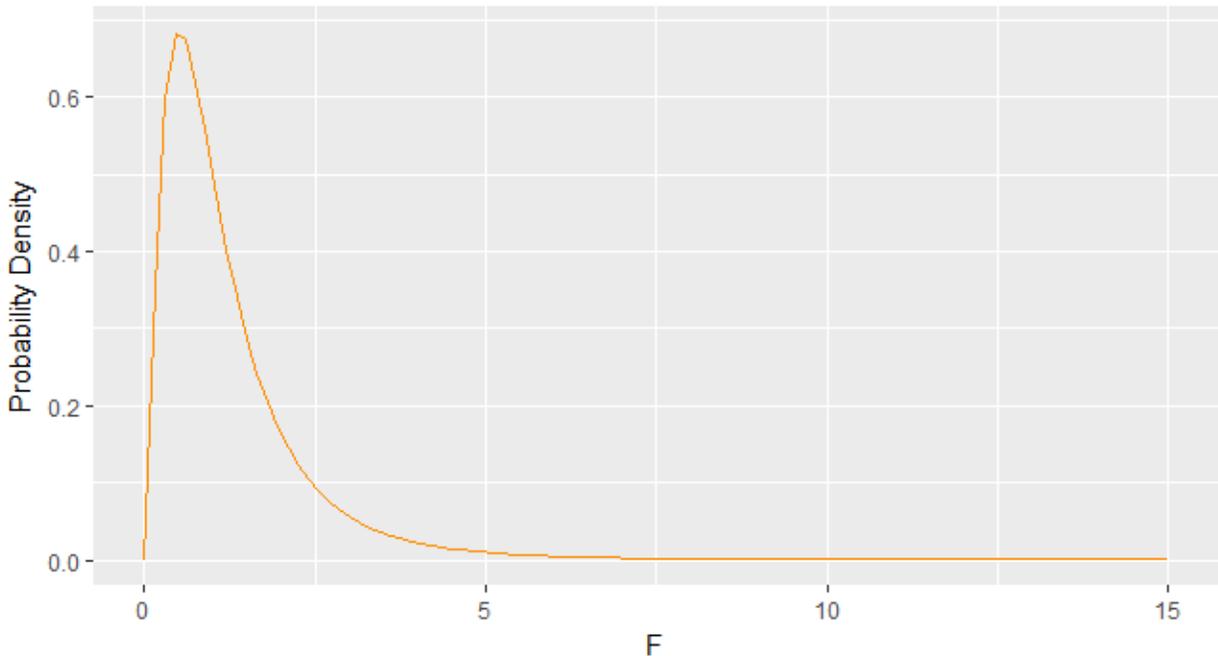


Figure 7: F probability distribution with 5 and 10 degrees of freedom respectively

Like the chi-squared distribution, the F -distribution is defined by two parameters - the two separate degrees of freedom values. The mean and variance of the F -distribution can be derived from the values of the degrees of freedom. Probabilities may be calculated using a computer or an F -table.

Exponential

The exponential distribution is a continuous distribution that is related to the Poisson distribution. While the Poisson distribution represents the distribution of the number of events happening in a given time frame, the exponential distribution represents the distribution of the time between two events. The exponential distribution has many common applications involving time or distance in an engineering setting. It is often used to model reliability because of its simple properties. Once again, probability can be represented by the area under the curve. Properties of the exponential distribution are outlined below:

1. The shape is asymmetric and based around λ (lambda), the rate parameter
2. The mean is the inverse of the rate parameter, λ
3. The variance is the inverse of the rate parameter squared
4. The total area under the curve is equal to 1

5. The curve spans values 0 and $+\infty$ (never a negative value)

The rate parameter tells how often an event happens in a given amount of time on average. Recall the machine that was incorrectly printing the “m” on the M&Ms. Assume now that this machine makes a mistake every 15 minutes on average. Figure 8 shows the distribution of time between mistakes for this machine.

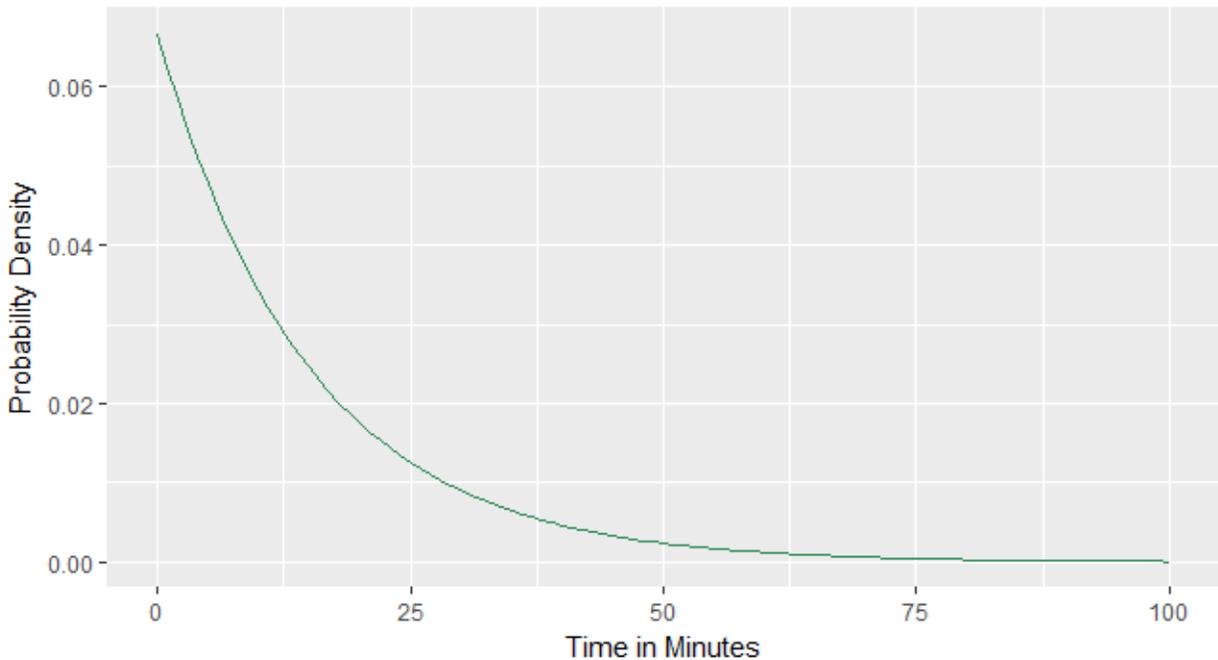


Figure 8: Exponential distribution where a machine makes a mistake every 15 minutes

The shape of the distribution will change depending on the rate parameter. Unlike the other distributions, this distribution does not have an associated table. However, probabilities can be calculated using the pdf or a computer.

Weibull

The Weibull distribution is another continuous distribution with many practical applications, particularly modeling reliability. Probability can be represented as the area under the curve. Of the distributions listed above, the Weibull is most closely related to the exponential distribution. However, it has two additional parameters that all combine to define the shape, scale, and location of the distribution. Because the Weibull distribution has so many parameters, it is a very flexible distribution that models many different aspects of a system. Properties of the Weibull distribution are outlined below:

1. Defined by three parameters: shape, scale, and location
2. The total area under the curve is equal to 1
3. The curve spans values 0 and $+\infty$ (never a negative value)

Depending on the textbook or source of the probability density function, the parameters are represented by different symbols. For ease, just think of them in terms of their functions. Many practical applications will not use the location parameter, in which case setting this value to zero will simplify the distribution to a more commonly used two-parameter Weibull distribution. Consider a system where the failure time of a critical part follows a Weibull distribution with a shape parameter equal to 2.7 and a scale parameter equal to 97 hours, seen in Figure 9.

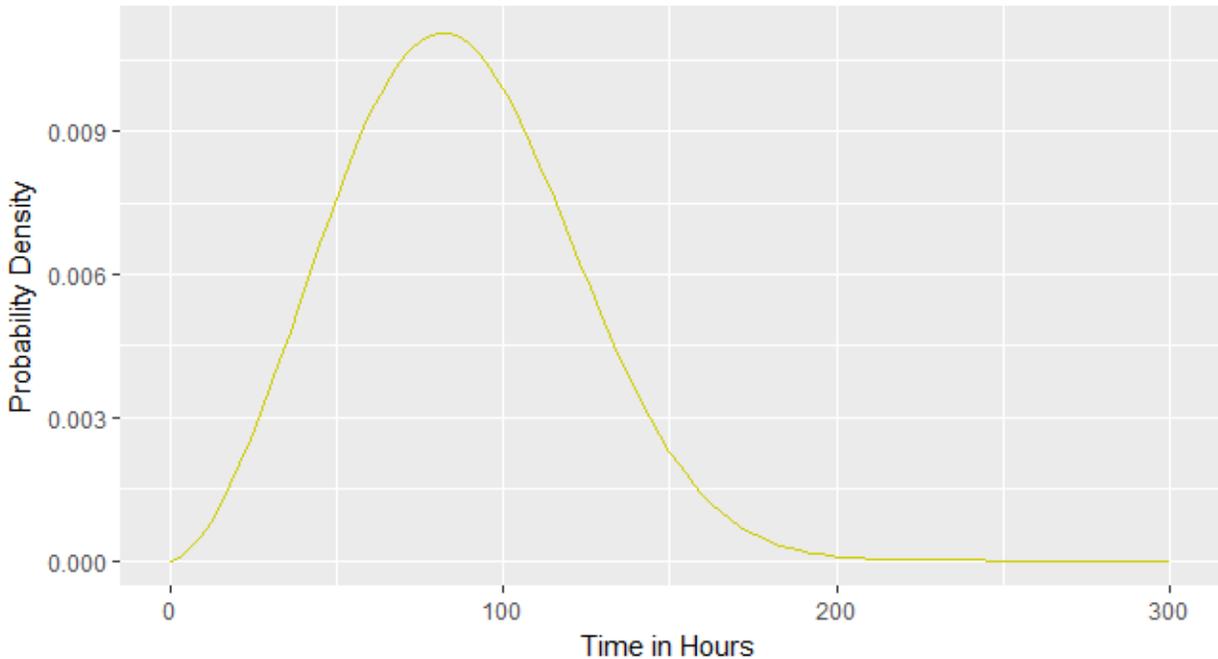


Figure 9: Weibull distribution with a shape parameter of 2.7, scale parameter of 97 hours, and a location parameter of 0

As with all the other distributions, the shape will change as the parameter values are changed. Because of the many possible combinations, the Weibull distribution does not have an associated table, but probabilities can be calculated by using the pdf or a computer.

Central Limit Theorem

The distributions described above represent only a small proportion of defined probability distributions commonly used in practice. However, there is a convenient and powerful theorem used in statistics that makes the underlying distribution of the data less important. The Central Limit Theorem states that the distribution of the **sample mean** is approximately normal as long as the sample gathered is sufficiently large. The distribution of the sample mean will have the population mean from the parent distribution and have a standard deviation of approximately σ/\sqrt{n} . This means that the overall variability of the distribution will shrink as sample size increases. Notice that the Central Limit Theorem doesn't say anything about the distribution of the data in question. The underlying distribution of the data only

matters in determining how “large” of a sample needs to be gathered for the theorem to apply. If the data comes from a distribution very different than the normal distribution (e.g., heavily skewed to one side), a larger sample is needed for the theorem to apply. The usual rule of thumb for real-world data sets is that a sample of 30 or more is sufficient to satisfy the rule of “large enough.” Being able to assume that the sample mean is normally distributed makes the completion of statistical inference, such as hypothesis testing, confidence interval, tolerance intervals, etc., much simpler.

Sample Mean vs. Population Mean

As an example of the Central Limit Theorem in action, consider the amount of time spent visiting a bank. The distribution of these times can be considered heavily skewed. Most people will spend very little time in the bank, but those wanting to open new accounts or visit safety deposit boxes may spend a great deal of time in the bank. Figure 10 shows the population distribution of the probability of visiting for a given amount of time in the bank.

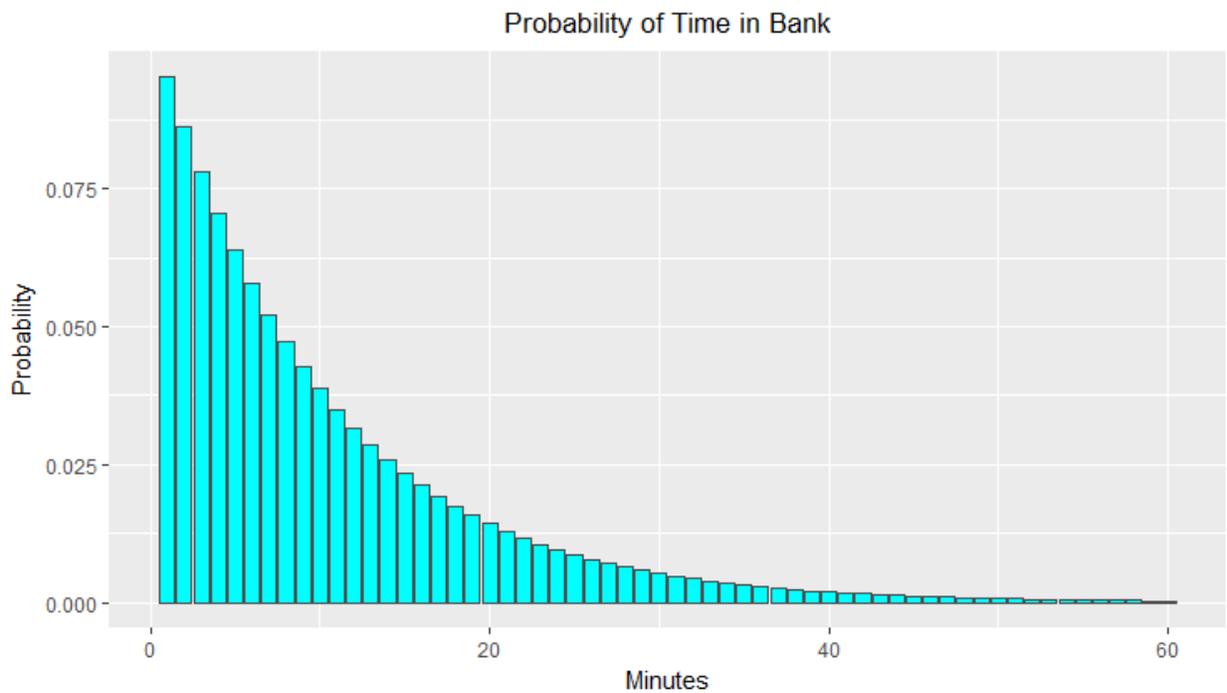


Figure 10: Probability histogram of time spent visiting a bank in minutes

While the population distribution is clearly skewed, the central limit theorem says the distribution of the sample mean will be approximately normally distributed if a sufficiently large sample is gathered. This can be tested by gathering different size samples and looking at this resulting distributions of means. Figure 11 shows the distribution for the sample mean when repeated samples of size 5 are independently taken from the population. The distribution still has a long tail, but it is not as heavily skewed as the population distribution (Figure 10).

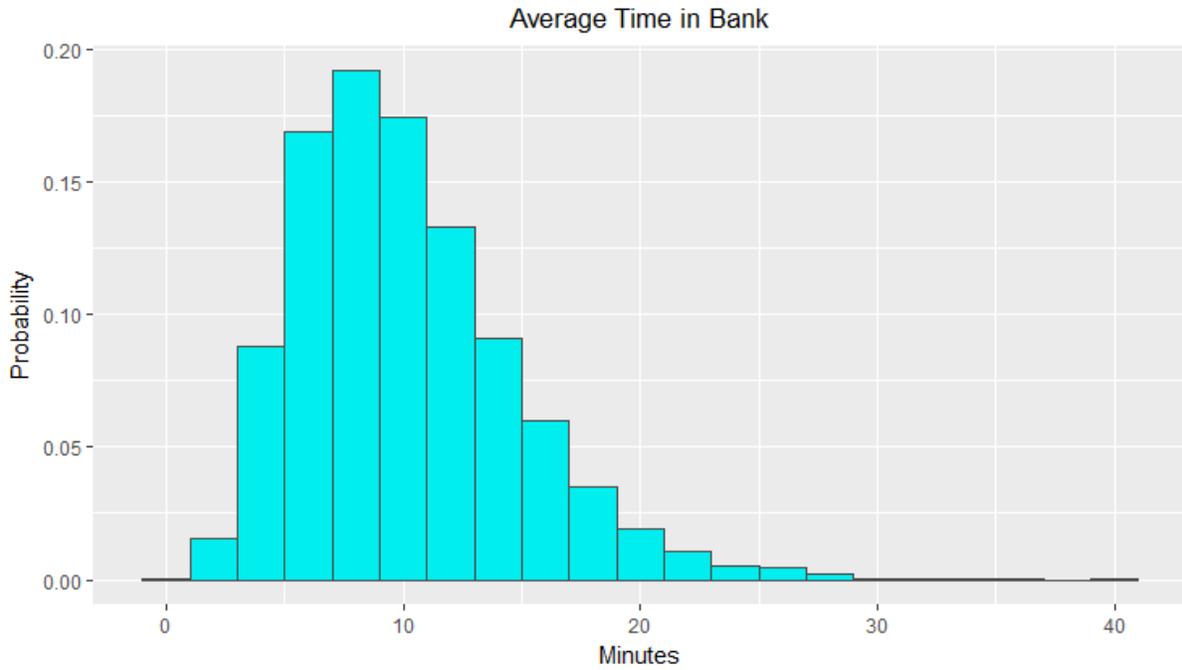


Figure 11: Sampling distribution of the sample mean with a sample size 5

Figure 12 shows the sampling distribution for the sample mean when repeated samples of size 10 are instead taken from the population; this distribution is much closer to being normally distributed, and note that the spread of the distribution is smaller than in Figure 11.

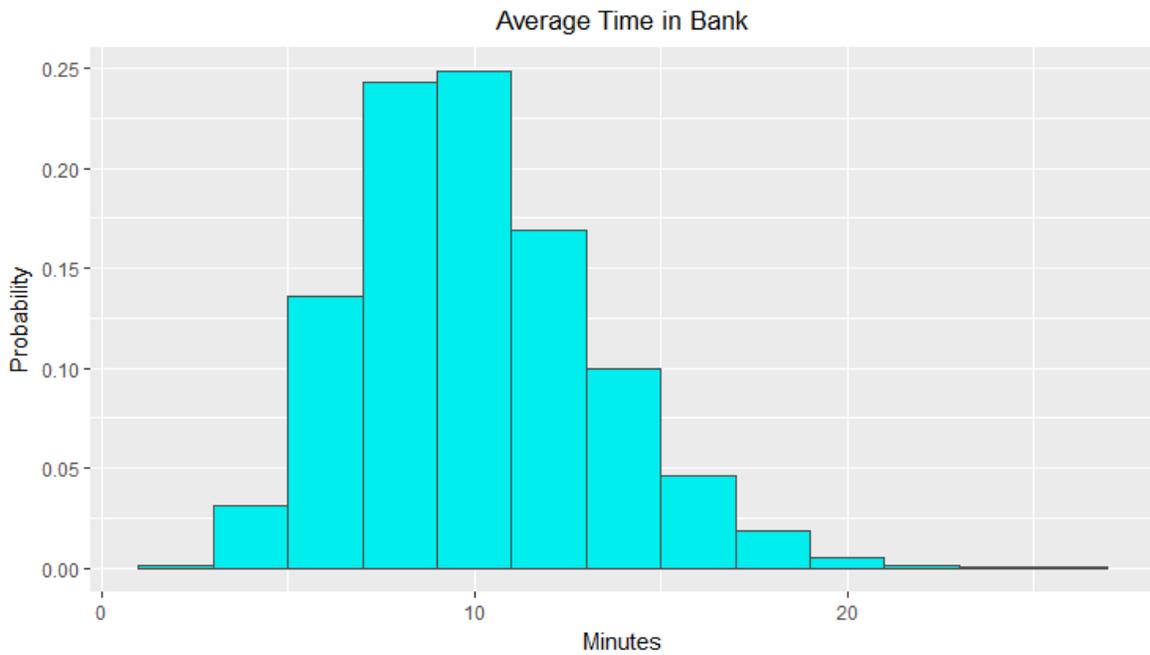


Figure 12: Sampling distribution of the sample mean with a sample size 10

Finally, Figure 13 shows the distribution when samples of size 30 are repeatedly taken from the population. This distribution of the sample mean very closely resembles a normal distribution. Also notice that the standard deviation for this is smaller than that seen in the population.

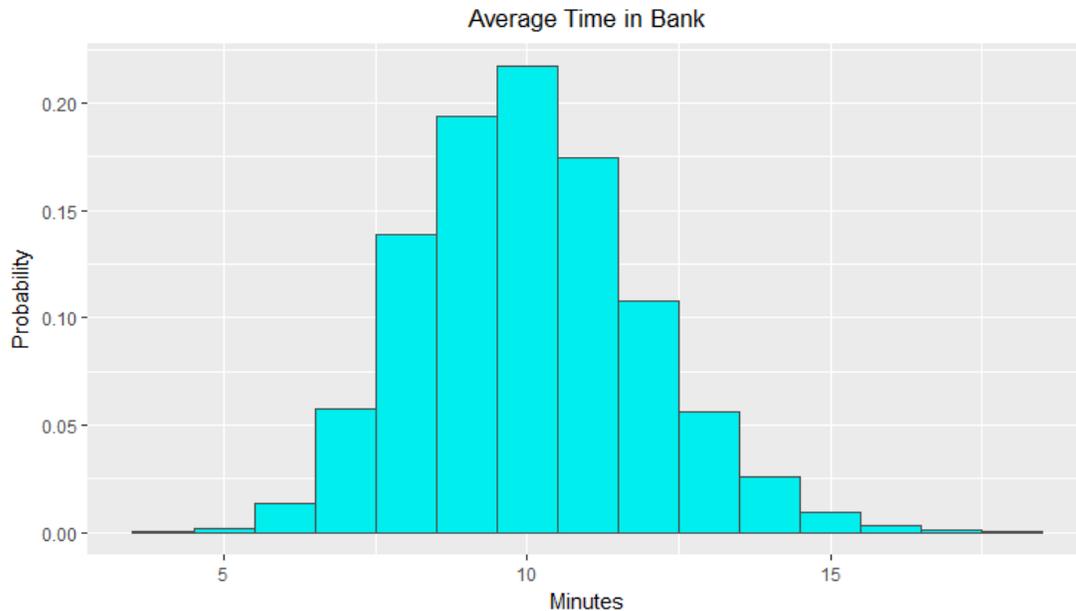


Figure 13: Sampling distribution of the sample mean with a sample size 30

Conclusion

Probability is a tool that can help inform relationships between the sample and population information. Very commonly, probability is referred to in terms of a probability distribution. For discrete data, distributions such as the binomial, Poisson, and hypergeometric are seen. For continuous data, distributions like the normal, student's t , chi-squared, F , exponential, and Weibull are instead seen. These are only a subset of all the possible distributions. However, these tend to be some of the most commonly used distributions in engineering and reliability examples. Additional distributions and more detailed information regarding probability can be found in introductory level statistics textbooks.

References

Casella, George, and Roger L Berger. *Statistical Inference*. 2nd ed., Duxbury, 2002.

Mendenhall, William, et al. *Introduction to Probability and Statistics*. 14th ed., Cengage Learning, 2013.

Ott, Lyman, and Michael Longnecker. *An Introduction to Statistical Methods & Data Analysis*. 7th ed., Cengage Learning, 2016.

Ross, Sheldon M. *A First Course in Probability*. 8th ed., Pearson Prentice Hall, 2010.

Appendix A: Distribution Formulas

Binomial probability mass function where n is the number of trials, p is the probability of success, and y is the number of successes:

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y = 0, 1, \dots, n$$

Poisson probability mass function where y is the number of events in an interval and λ is the average number of events that occur in the interval:

$$p(y) = e^{-\lambda} \frac{\lambda^y}{y!} \quad y = 0, 1, 2, 3, \dots$$

Hypergeometric probability mass function where y is the number of successes, N is the total number in the population, n is the total number chosen from the population, and k is the number of possible successes:

$$p(y) = \frac{\binom{k}{y} \binom{N-k}{n-y}}{\binom{N}{n}} \quad y = 0, 1, \dots, k$$

Normal probability density function where x is the observed value, μ is the distribution mean, and σ is the distribution standard deviation:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$$

Exponential probability density function where x is the observed value and λ is the rate parameter:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Weibull probability density function where x is the observed value, α is the scale parameter, β is the shape parameter, and v is the location parameter:

$$f(x) = \begin{cases} 0 & \text{if } x \leq v \\ \frac{\beta}{\alpha} \left(\frac{x-v}{\alpha}\right)^{\beta-1} e^{-((x-v)/\alpha)^\beta} & \text{if } x > v \end{cases}$$

The student's t , chi-squared, and F probability density functions are omitted due to their complicated natures. However, details about these distributions including all means and variances can be found in the Table of Common Distributions in the back of *Statistical Inference* by Casella and Berger. Wikipedia may also be used for more details about any of the above mentioned distributions.