Statistics and Probability for Engineering

Mohammad Saifuddin, Assistant Professor

Table of contents

Preface

This is an eBook developed for the undergrad students of engineering faculty to provide knowledge of statistics, probability, probability distributions, statistical inference, correlation and regression analysis,stochastic process and design and analysis of experiment with applications in the engineering field.

The whole book is developed using R Programming Language (R Core Team 2024).

1 Introduction to statistics and Data analysis

The field of **statistics** deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes. In simple terms, **statistics is the science of data**.

Because many aspects of engineering practice involve working with data, obviously knowledge of statistics is just as important to an engineer as are the other engineering sciences.

Specifically, statistical techniques can be powerful aids in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

Statistical methods are used to help us describe and understand **variability**. By variability, we mean that successive observations of a system or phenomenon do not produce the same result. We all encounter variability in our everyday lives, and **statistical thinking** can give us a useful way to incorporate this variability into our decision-making processes (Montgomery and Runger 2014a).

There are two types of data sets you will use when studying statistics. These data sets are called **populations** and **samples**.

A **population** is the collection of all outcomes, responses, measurements, or counts that are of interest.
A **sample** is a subset, or part, of a population.

Two important terms that are used throughout this course are ***parameter*** and ***statistic***.

A ***parameter*** is a numerical description of a *population characteristic.*
A ***statistic*** is a numerical description of a *sample characteristic.*

Data refers to a collection of facts, measurements, observations, or information that is gathered to be analyzed and used for decision-making, understanding patterns, or testing hypotheses.

These facts can represent various forms of information, such as numbers, words, measurements, or even categories.

Types of data

Quantitative Data (Numerical Data)

- Data that represents numerical values.
- Example: Heights of people, temperatures, test scores.

- Subtypes:
  - **Discrete Data**: Countable values (e.g., number of students in a class).
  - **Continuous Data**: Measurable values that can take any value within a range (e.g., weight, time).

Qualitative Data (Categorical Data)

- Data that represents categories or labels.
- Example: Colors of cars, types of animals, survey responses (e.g., yes/no).
- Subtypes:
  - **Nominal Data**: Categories without a natural order (e.g., gender, blood type).
  - **Ordinal Data**: Categories with a meaningful order (e.g., rankings, education levels).

Collecting Engineering Data

In the previous subsection, we illustrated some simple methods for summarizing data. Sometimes the data are all of the observations in the population. This results in a **census**. However, in the engineering environment, the data are almost always a **sample** that has been selected from the population. Three basic methods of collecting data are

(i) A **retrospective study** using historical data (ii) An **observational study** (iii)A **designed experiment**

An effective data-collection procedure can greatly simplify the analysis and lead to improved understanding of the population or process that is being studied. We now consider some examples of these data-collection methods. (For detail, *see* Montgomery and Runger (2014a) **, Chapter 1**)

The study of statistics is divided into two parts- **Descriptive statistics** and **Inferential statistics**.

Descriptive statistics is the branch of statistics that involves the organization, summarization, and display of data.
Inferential statistics is the branch of statistics that involves using a sample to draw conclusions about a population.

A basic tool in the study of inferential statistics is **probability** and **probability distribution**s.

The goal of inferential statistics is to draw conclusions from a sample and generalize them to the population. The most common methodologies used are hypothesis tests, Analysis of variance, etc.

We will begin with **descriptive statistics.**

1.1 Descriptive statistics

An important aspect of dealing with data is organizing and summarizing the data in ways that facilitate its interpretation and subsequent analysis. This aspect of statistics is called **descriptive statistics**. Usually, data are summarized both **numerically** and **graphically**. In this module, our focus will be concentrated on the following topics:

(i) Numerical Summaries of Data (ii) Graphical summaries (Histogram and Box Plot)

Numerical summaries of the data

Often to summarize data we use some numerical measures like measures of central tendency, measures of location and measures of variation. The common numerical summaries are:

1) Measures of central tendency	2) Measures of relative standing: Quantile	3) Measures of variability
Mean/ Arithmetic mean/average, Median, Mode etc	a)Percentiles, Quartiles etc.	a) Range, Inter-quartile range (IQR), Variance and standard deviation, Coefficient of variation (CV)% etc.

While studying these numerical summaries, we consider **sample data**.

1.2 Measures of central tendency and quantile

1.2.1 Mean

- **Sample mean:** Suppose n observation of a variable X is drawn from a population. Then the sample mean is denoted by $\bar{x}$ and

  $$\bar{x} = \frac{\sum x}{n}$$

  The sample mean $\bar{x}$ is a sample statistic.

- **Population mean:** Suppose in a population there are N values of variable X. Then the population mean is denoted by μ and

$$\mu = \frac{\sum x}{N}$$

The $\bar{x}$ is a point estimator of the population mean μ.

Example 1.1: Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds): 12 .6, 12. 9, 13. 4, 12. 3, 13. 6, 13 .5, 12. 6, 13. 1.

Suppose,X =pull-off forces (in pounds). So,

$X = \{x_1, x_2, ..., x_8\} = \{12.6, 12.9, ..., 13.1\}$

The sample mean is, pounds

$$\bar{x} = \frac{\sum x}{n} = \frac{12.6 + 12.9 + ... + 13.1}{8} = \frac{104}{8} = 13.0 \quad pounds$$

1.2.2 Median

The **median** is another measure of central location. The median is the value in the middle when the data are arranged in ascending order (smallest value to largest value).

- For an *odd* number of observations, median is the **middle value**
- For an *even* number of observations, median is the **average of the two middle values**

Example 1.2: CPU time of 9 jobs (in seconds):
Data: 59, 139, 46, 37, 42, 30, 55, 56, 82
Arranged/sorted data: 30, 37, 42, 46, 55, 56, 59, 82, 139
Median=55; whereas the **Mean=60.67**

Example 1.3 : CPU time of 10 jobs (in seconds): 59, 139, 46, 37, 42, 30, 55, 56, 36, 82
Arranged/sorted data: 30, 36, 37, 42, 46, 55, 56, 59, 82, 139.

Median = (46+55)/2=50.5; whereas **Mean**=58.2.

ℹ Note

Notice that the median is unaffected by the size of the largest CPU time. It impiles that, mean is affected by extreme value or outlier but median is not.

1.2.3 Mode

is the most frequently occurring data value.

1.2.4 Percentiles

Percentiles divide the whole data set into approximately 100 equal parts. So, there are 99 percentiles -$P_1, P_2, ..., P_{99}$. In this lecture j^{th} percentile will be denoted by P_j .

We can compute the j^{th} percentile as follows:

$$P_j = (j * \frac{n+1}{100})^{th} \ \ value; \ \ j = 1, 2, ..., 99.$$

Example 1.4: Compute the 25th and 60th percentile from the following data:

CPU time of 9 jobs (in seconds): 59, 46, 37, 42, 30, 55, 56, 36, and 82.

Solution: Here, $n = 9$.

Sorted data: 30, 36, 37, 42, 46, 55, 56, 59, and 82.

So, $P_{25} = (25 * \frac{9+1}{100})^{th} \ \ value = 2.5^{th} \ \ value$

$= 2^{nd} \ \ value + 0.5(3^{rd} - 2^{nd}) = 36 + 0.5(37 - 36) = 36.5.$

Interpretation: $P_{25} = 36.5$ implies that approximately 25% of the total observations lie below or equal to 36.5.

Similarly, $P_{60} = (60 * \frac{9+1}{100})^{th} \ \ value = 6^{th} \ \ value = 55.$

Interpretation: $P_{60} = 55$ implies that approximately 60% of the total observations lie below or equal to 55.

1.2.5 Quartiles

It is often desirable to divide data into four parts, with each part containing approximately one-fourth, or 25% of the observations. The division points are referred to as the quartiles and are defined as

Q_1 = first quartile, or 25th percentile

Q_2 = second quartile, or 50th percentile (also the median)

Q_3 = third quartile, or 75th percentile.

Example 1.5: Here is the monthly starting salary ($) of 12 graduates:

3450, 3550, 3650, 3480, 3355, 3310, 3490, 3730, 3540, 3925, 3520, 3480

Compute Q1 and Q3 of the above data (*Will be solved in class*).

1.3 Measures of variability

Variability in data means lack of uniformity. It is also referred to as spread, scatter, or dispersion. We turn now to a discussion of some commonly used measures of variability.

1.3.1 Range

$$Range = Largest \;\; value - Smallest \;\; value$$

- Heavily influenced by extreme values

1.3.2 Inter-quartile-range (IQR)

$$IQR = Q_3 - Q_1$$

- It exhibits the variability in the middle 50% of the observations.
- The interquartile range is less sensitive to the extreme values in the sample than is the ordinary sample range.

Detection of outliers using 1.5(IQR) rule

In this method, we discuss fences.

- The lower fence of distribution is , $lf = Q_1 - 1.5(IQR)$
- The upper fence of distribution is , $uf = Q_3 + 1.5(IQR)$
- If any value lies outside the interval $[lf, uf]$; then it will be considered an outlier.

Example 1.6: The following data set represents the number of new computer accounts registered during ten consecutive days. 43, 37, 50, 51, 58, 105, 52, 45, 45, 10. Check for outliers using the 1.5(IQR) rule.

1.3.3 Variance and standard deviation

The variability or scatter in the data around mean may be described by the variance and standard deviation.

- **Population variance:** If $x_1, x_2, ..., x_N$ is a *population* of N observations , the population variance is

$$\sigma^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+...+(x_N-\mu)^2}{N}$$

$$= \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}x_i^2}{N} - \mu^2$$

- **Sample variance:** If $x_1, x_2, ..., x_n$ is a *sample* of n observations , the sample variance is

$$s^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+...+(x_n-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

- An alternative formula for the computation of the sample variance is:

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

where, $\sum x_i^2 = x_1^2 + x_2^2 + ... + x_n^2$

Example 1.7: Eight prototype units are produced and their pull-off forces measured, resulting in the following data (in pounds): 12 .6, 12. 9, 13. 4, 12. 3, 13. 6, 13 .5, 12. 6, 13. **Compute** sample variance.

Solution:

Table 1.2: Computation of the sample variance for the pull-off force data

Pull-off force(x_i)	Sample mean, $\bar{x}$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
12.6	13	-0.4	0.16
12.9	13	-0.1	0.01
13.4	13	0.4	0.16
12.3	13	-0.7	0.49
13.6	13	0.6	0.36
13.5	13	0.5	0.25
12.6	13	-0.4	0.16
13.1	13	0.1	0.01

Pull-off force(x_i)	Sample mean, $\bar{x}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
		$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 1.6$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{1.6}{8-1} = 0.2286 \quad (pounds)^2$$

Alternative: Here $\sum x^2 = 12.6^2 + 12.9^2 + \cdots + 13.1^2 = 1353.6$

$\bar{x} = 13$

So, $s^2 = \frac{\sum x^2 - n \times \bar{x}^2}{n-1} = \frac{1353.6 - 8 \times (13^2)}{8-1} = 0.2285714 \approx 0.2286 \quad (pounds)^2$

Standard deviation

The **standard deviation** is defined to be the positive square root of the variance

- **Sample standard deviation=**$s = \sqrt{s^2}$
- **Population standard deviation=**$\sigma = \sqrt{\sigma^2}$

The sample standard deviation s is the estimator of population standard deviation σ.

Example 1.8: The standard deviation of the previous example is :

$$s = \sqrt{0.2286} \approx 0.48 \quad pounds$$

Note:

The standard deviation is easier to interpret than the variance because the standard deviation is measured in the same units as the data.

For example, the sample variance for the pull-off force data of prototype units is $s^2 = 0.2286 \quad (pounds)^2$.

Because the standard deviation is the square root of the variance, the units of the variance, pounds squared, are converted to pound in the standard deviation.

Thus, the standard deviation of the pull-off force data is 0.48 pounds. In other words, the standard deviation is measured in the same units as the original data. For this reason the standard deviation is more easily compared to the mean and other statistics that are measured in the same units as the original data.

1.3.4 Coefficient of variation

In some situations we may be interested in a descriptive statistic that indicates how large the standard deviation is relative to the mean. This measure is called the **coefficient of variation** and is usually expressed as a percentage.

Coefficient variation,

$$CV = \frac{Standard\ \ deviation}{Mean}$$

- The coefficient of variation is a relative measure of variability; it measures the standard deviation relative to the mean.
- In general, the coefficient of variation is a useful statistic for comparing the variability of variables that have different standard deviations and different means.

Example 1.9: The table at the left shows the population heights (in inches) and weights (in pounds) of the members of a basketball team. Find the coefficient of variation for the heights and the weights. Then compare the results.

1.4 Data

Heights (inches)	Weights (pounds)
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

1.5 Coefficient of variation

The mean height $\mu \approx 72.8 \ \ inches$ with a standard deviation $\sigma = 3.3 \ \ inches.$

The coefficient of variation for the heights is

$CV_{height} = \frac{\sigma}{\mu}.100\% = \frac{3.3}{72.8}.100\% \approx 4.5\%.$

The mean weight $\mu \approx 187.8 \ \ pounds$ with a standard deviation $\sigma = 17.7 \ \ pounds.$

The coefficient of variation for the weights is

$CV_{weight} = \frac{\sigma}{\mu}.100\% = \frac{17.7}{187.8}.100\% \approx 9.4\%$

Interpretation The weights (9.4%) are more variable than the heights (4.5%).

1.6 Graphical summary/visualization

Before diving into advanced analysis we should have a look at the data. Because " A picture is worth a thousand words". Often a summary of a collection of data via a graphical display can provide insight regarding the system from which the data were taken.

1.6.1 Frequency Distributions and Histograms

To construct a frequency distribution, we must divide the range of the data into intervals, which are usually called **class intervals, cells**, or **bins,** and count how many observations fall into each **bin**.

The **histogram** is a visual display of the frequency distribution.

- A **frequency histogram** consists of columns, one for each **bin**, whose height is determined by the **number** of observations in the **bin**. [Frequency distribution]
- A **relative frequency histogram** has the same shape but a different vertical scale. Its column heights represent the **proportion** of all data that appeared in each bin. [Relative frequency distribution]

What is the appropriate size of **bins?**

There are several rules, but one of them is, the number of bins, suppose $k = \sqrt{n}$. We will discuss how to construct a frequency distribution, relative frequency distribution, and cumulative frequency distribution in the following example.

Example 1.10: The following data are the joint temperatures of the O-rings (°F) for each test firing or actual launch of the space shuttle rocket motor (from *Presidential Commission on the Space Shuttle Challenger Accident*, Vol. 1, pp. 129–131):

> 67, 40, 58, 76, 58, 70, 72, 67, 75, 70, 57, 83, 53, 45, 70, 81, 78, 76, 67, 73, 61, 52,
> 31, 67, 79, 75, 69, 84, 68, 80

To construct frequency distribution and others follow the steps:

Step-1: Find the **number bins**, $k = \sqrt{n} = \sqrt{3}0 = 5.48 \approx 6$ (round to nearest integer).

Step-2: Find the range of the data, $R = Maximum - Minimum = 84 - 31 = 53$

Step-3: Determine **bin width**, $w = \frac{R}{k} = \frac{53}{6} = 8.83 \approx 10$

Step-4: Define the **bins** in ***exclusive*** method, starting from a suitable data value close to lowest value, for example **[30, 40), [40, 50), and so on** until we have the highest data value.

Step-5: Now count the observation fall in each bin, using tally.

Table 1.4: Frequency, Relative frequency, Cumulative frequency Distribution for the O-rings temperature data (n=30)

Temperatures (in ^{o} F)	Tally	Frequency (f)	Relative frequency (rf)	Cumulative frequency (cf)
[30,40)	|	1	0.03	1
[40,50)	||	2	0.07	3
[50,60)	卌	5	0.17	8
[60,70)	卌 ||	7	0.23	15
[70,80)	卌 卌 |	11	0.37	26
[80,90)	||||	4	0.13	30
Total		n=30	1.00	

1.6.2 Histogram

The histogram provides a visual impression of the **shape** of the distribution of the measurements and information about the **central tendency** and scatter or **dispersion** in the data.

The histogram from the previous example is shown in Figure 1.1:

Histogram and shape of the distribution

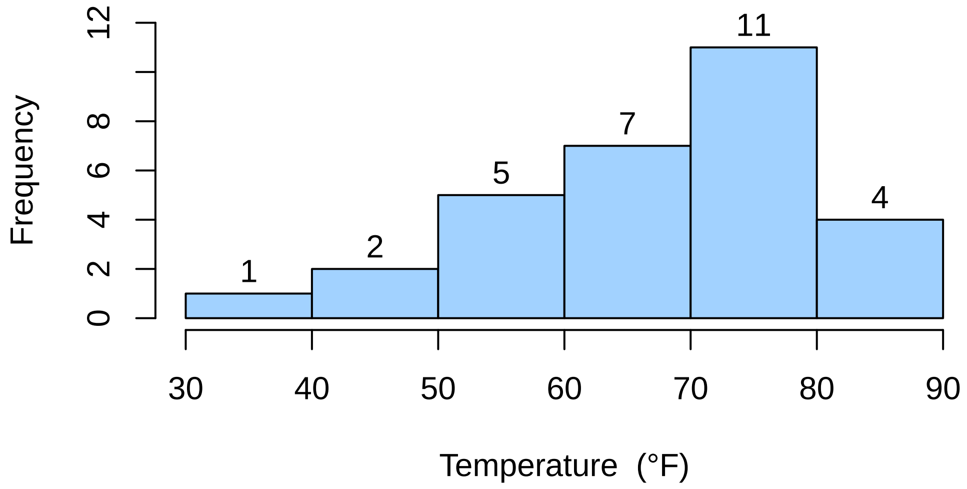


Figure 1.1: Frequency histogram of temperature (°F)

When the sample size is large, the histogram can provide a reasonably reliable indicator of the general **shape** of the distribution or population of measurements from which the sample was drawn. (for detail *see* Montgomery and Runger (2014a)).

1.6.3 Box-plot

The **box plot** is a graphical display that simultaneously describes several important features of a data set, such as *center*, *spread*, departure from *symmetry*, and identification of unusual observations or *outliers*.

A box plot, sometimes called *box-and-whisker plots*, displays the **three quartiles**, **the minimum/lower fence**, and **the maximum/upper fence** of the data on a rectangular box, aligned either horizontally or vertically.

If any value falls outside the **fences** then it will shown as a **circle** in the box-plot.

Example 1.11

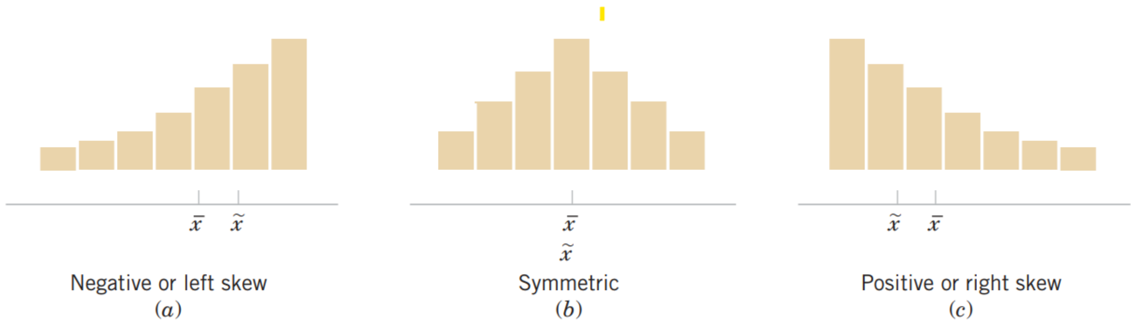
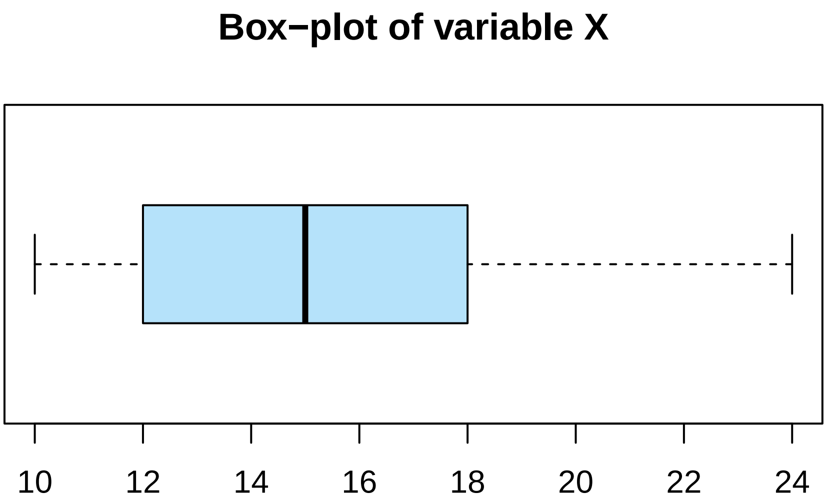


Figure 1.2: Histograms for symmetric and skewed distributions

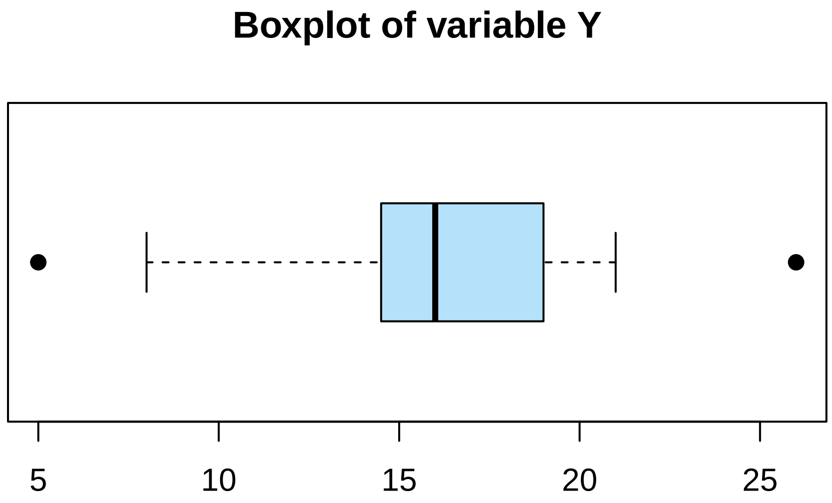
1.7 Variable-I

Suppose, X={10,15,14,18,17,12,16,15,19,21,32,12,58}



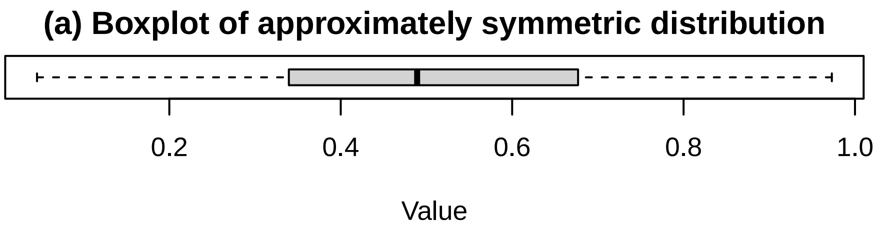
1.8 Variable-II

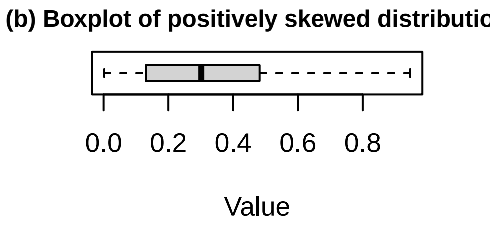
Suppose, Y={5,8,14,15,18,17,16,14,15,21,21,35,17,16,20}

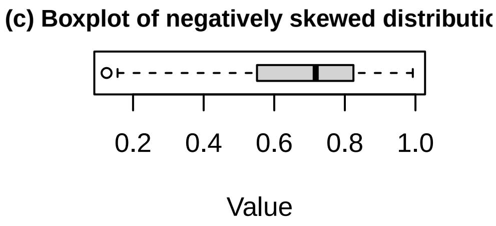


1.8.1 Boxplot and skewness of the data

When we discuss the frequency histogram we also learned about shape of the distribution. By visual inspection of boxplot we can also tell about the distribution shape of a variable. The following boxplots are the typycal examples of skewness of the data.



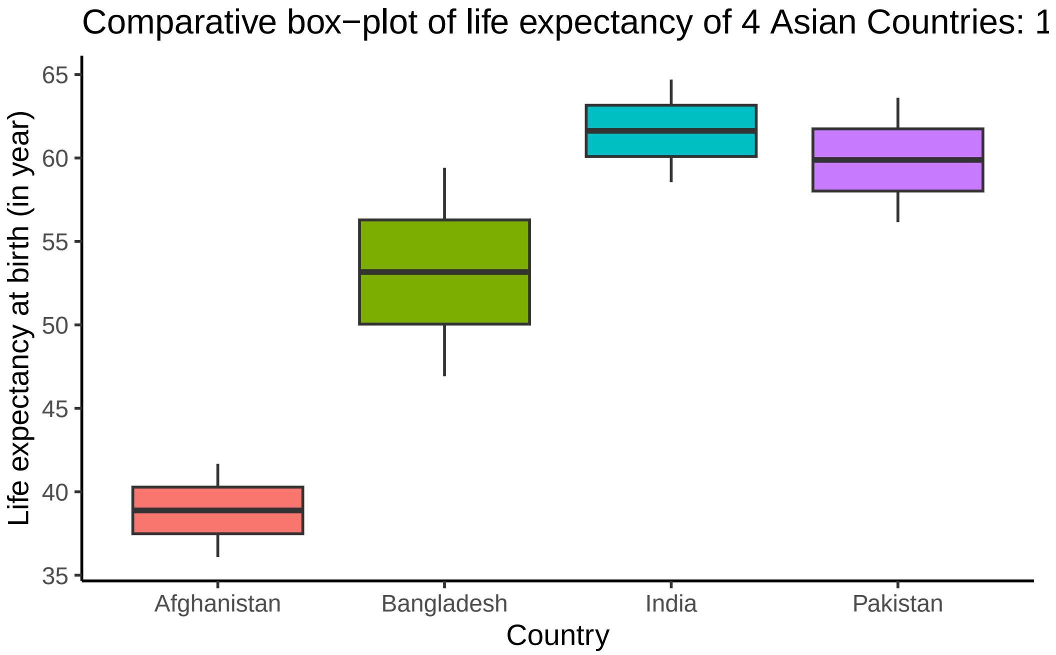




Comparative/Parallel box-plots

Box plots are very useful in graphical comparisons among data sets because they have high visual impact and are easy to understand.

- For instance, here the **Life expectancy at birth (in year)** of Afghanistan, Bangladesh, India and Pakistan is compared using a **comparative box-plot** from 1971 to 2007.



Question: What is/are your observation(s) from this above comparative box-plot?

1.9 Exercises (Practice as more as you can)

1.1 Define statistics, population, sample, descriptive and inferential statistics.

1.2 Define data. Discuss the types of data with example.

1.3 What are the common measures of location? When median is preferable to mean?

1.4 What are the common measures of variation?

1.5 Define five-number summary. How to detect outliers using quartiles?

1.6 The data contains the overall gallons per kilometer (GpK) of a medium-sized mobile home unit.

35, 30, 37, 35, 34, 35, 35, 42,40, 37, 42, 38, 35, 34, 35, 34, 34.

Calculate and **explain** Median. **Find** the value above which 15% GpK values lie.

1.7 In automobile mileage and gasoline-consumption testing, 13 automobiles were road tested for 300 miles in city driving conditions. The following data were recorded for miles-per-gallon (mpg) performance.

13.2, 14.4, 15.2, 15.3, 15.3, 15.3, 15.9, 16.0, 16.1, 16.2, 16.2, 16.7, 16.8

i) **Construct** a simple boxplot (in horizontal direction) of mpg. Are most of the automobiles' mpg relatively low?

ii) Suppose you want to buy a new car and you don't afford enough money, so the car's mileage must be in *bottom* 10%. What should be the mileage of your car based on this data?

iii) Suppose you want to buy a new car which mileage must be in *top* 10%. What should be the mileage of your car based on this data?

1.8 The data set lists the prices (in dollars) of 20 portable global positioning system (GPS) navigators.

128, 100, 180, 150, 200, 90, 340, 105, 85, 270, 200, 65, 230, 150, 150, 120, 130, 80, 230, 200

i) **Construct** a frequency distribution and percent frequency distribution of prices.

ii) **Draw** a frequency histogram. What is your observation about price data?

1.9 The lengths of power failures, in minutes, are recorded in the following table.

18, 135, 15, 90, 78, 69, 98, 102, 83, 55, 28

i) **Compute** sample mean and standard deviation

ii) **Compute** sample median and IQR

iii) If you want to be in the bottom 10% of power failures in your residence, then what should be the cutoff value of power failure lengths in minutes?

1.10 Sample annual salaries (in thousands of dollars) for entry-level electrical engineers in Boston and Chicago are listed.

Boston 70.4, 84.2, 58.5, 64.5, 71.6, 79.9, 88.3, 80.1, 69.9

Chicago 69.4, 71.5, 65.4, 59.9, 70.9, 68.5, 62.9, 70.1, 60.9

Find the coefficient of variation for each of the two data sets. Then **compare** the results.

1.11 The shear strengths of 20 spot welds in a titanium alloy are as follows:

5408 5431 5475 5442 5376 5388 5459 5422 5416 5435

5420 5429 5401 5446 5487 5416 5382 5357 5388 5457

Construct a frequency histogram of shear strength data. **Conclude** whether the shear strength is approximately symmetric or not.

1.12 The "cold start ignition time" of an automobile engine is being investigated by a gasoline manufacturer. The following times (in seconds) were obtained for a test vehicle:

1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91.

Check for outliers using the 1.5*IQR rule of cold start ignition time.

1.13 The cylindrical compressive strength (in MPa) was measured for 11 beams. The results were:

38.43, 38.43, 38.39, 38.83, 38.45, 38.35, 38.43, 38.31, 38.32, 38.48, 38.50.

i) Construct a simple boxplot and comment about the shape of the distribution of compressive strength.

ii) Compute sample mean, standard deviation and coefficient of variation (CV) compressive strength.

1.14 The shear strengths (in N/m2) of 10 spot welds in a titanium alloy are:

5408, 5431, 5475, 5442, 5376, 5388, 5459, 5422, 5416, 5435.

Compute sample mean, standard deviation and coefficient of variation of shear strength.

1.15 An article describes an experiment to test the yield strength of circular tubes with caps welded to the ends. The first 18 yields (in kN) are:

96, 96, 102, 102, 102, 104, 104, 108, 126,126, 128, 128, 140, 156, 160, 160, 164,170.

Construct a simple box-plot of yield strength. **Find** the value above which top 10% yield strength lie.

1.16 The cylindrical compressive strength (in MPa) was measured for 11 beams. The results were:

38.43, 38.43, 38.39, 38.83, 38.45, 38.35, 38.43, 38.31, 38.32, 38.48, 38.50.

i) **Compute** 1st quartile, 3rd quartile and IQR.

ii) **Check** for outlier (s).

iii) **Construct** a box-plot, **show** the outliers and **comment** about the symmetry of the distribution.

1.17 The following sample data presents the thickness (Å) of a metal layer on 20 silicon wafers resulting from a chemical vapor deposition (CVD) process. Scientists believe that the thickness is usually normally distributed having a bell-shaped distribution for a smooth process. **Construct** a frequency histogram of metal thickness data. **Conclude** whether the process is smooth or not.

468, 459, 450, 453, 473, 454, 458, 438, 447, 463,

445, 466, 456, 434, 471, 437, 459, 445, 454, 423

2 Probability

2.1 Introduction

A probability is the chance, or likelihood, that a particular event will occur. These are examples of events representing typical probability-type questions:

- How many customers will arrive in a super shop in next 30 minutes?
- What is probability that a stock price will rise or fall?

To answer these kind of questions in the face of uncertainty we need to study probability. To answer these type of questions which are raised in real life; at first we have to learn some basic concepts of probability.

2.2 Random experiment

A **random experiment** is a process leading to two or more possible outcomes, without knowing exactly which outcome will occur.

Example: Tossing a coin, throwing a dice, change in the stock prices etc.

2.3 Sample space

A **sample space** is the collection of all outcomes of a random experiment. The sample space is usually denoted by S or Greek letter Ω (omega).

Example 2.1:

- If we toss a coin then the sample space is: $S = \{H, T\}$
- If we toss 2 coins then the sample space is: $S = \{HH, HT, TH, TT\}$

2.4 Event

An **event** is a *subset* of a *sample space*.

For example suppose, $S = \{HH, HT, TH, TT\}$ and $A = \{one \quad head \quad occurs\}$. So, $A = \{HT, TH\}$.

Since A is a subset of sample space S, so A is an *event*.

2.5 Complement of an event

The complement of an event A with respect to Ω is the subset of all elements of Ω that are not in A. We denote the complement of A by the symbol A^C.

Example 2.2: Consider the sample space:

$\Omega = \{1, 2, 3, 4, 5, 6\}$

Let, $A = \{1, 3, 5\}$. Then the complement of A is $A^C = \Omega - A = \{2, 4, 6\}$

2.6 Mutually exclusive events

The occurrence of one event means that none of the other events can occur at the same time.

Example

- The variable "Employment status" presents mutually exclusive outcomes, *employed* and *unemployed*. An employee selected at random is either male or female but cannot be both.
- A manufactured part is acceptable or unacceptable. The part cannot be both acceptable and unacceptable at the same time.

2.7 Axiomatic definition of Probability

The **probability** of an event A is the sum of the weights of all sample points in A. Therefore

(a) $0 \leq P(A) \leq 1$; $P(\phi) = 0$ and $P(\Omega) = 1$.

(b) If $A_1, A_2, A_3, ...$ is a sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup A_3 \cup ...). = P(A_1) + P(A_2) + P(A_3) + ...$$

2.8 Probability of an event (Classical approach)

Suppose an event A is defined in the sample space Ω. Then the probability of event A is defined as :

$$P(A) = \frac{n(A)}{n(\Omega)};$$

Here,

$n(A) =$ number of outcomes favorable to event A;

$n(\Omega) =$ total number of outcomes in the sample space Ω.

2.9 Probability of an event (Empirical approach)

Empirical Probability is a type of probability that is calculated based on actual observations, experiments, or historical data rather than theoretical assumptions. It measures the likelihood of an event occurring by analyzing past occurrences or experimental results.

Formula for Empirical Probability:

$$P(E) = \frac{Number\ \ of\ \ times\ \ the\ \ event\ \ occurs}{Total\ \ number\ \ of\ \ trials}$$

Where:

- $P(E)$ is the probability of the event E,
- The numerator is the count of occurrences of the event, and
- The denominator is the total number of trials or observations.

Example 2.3: Suppose in a class there are 30 students; 20 are male and 10 are females. If a student is selected at random what is the probability that he is a male?

Solution: Let, $A_1 =$ set of male students and $A_2 =$ set of female students. And, $\Omega =$ set of all students

So, probability that a male student is selected is:

$$P(A_1) = \frac{n(A_1)}{n(\Omega)} = \frac{20}{30} = 0.66667 \approx 0.67$$

Interpretation There is almost 67% chance that the selected student will be male.

Example 2.4: If 3 books are picked at random from a shelf containing 5 novels, 3 books of poems, and a dictionary, what is the probability that

(a) the dictionary is selected?

(b) 2 novels and 1 book of poems are selected?

2.10 Properties of Probability Laws

Probability laws have a number of properties, which can be deduced from the axioms. Some of them are summarized below.

a) $P(A^C) = 1 - P(A)$ [**complement rule**]

b) $P(A \cap B^C) = P(A) - P(A \cap B)$ [**only A happens**]

c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ [**additive rule**]

d) $P(A^C \cap B^C) = P(A \cup B)^C = 1 - P(A \cup B)$. [**neither A NOR B happens**]

Example 2.5: In a class 65% students prefer tea and 35% students prefer coffee. While 15% students prefer both tea and coffee. If a student is selected at random from the class **find** the probability that

i) he/she prefers only coffee

ii) he/she prefers tea or coffee

iii) he/she prefers none (neither tea nor coffee)

Example All Seasons Plumbing has two service trucks that frequently need repair. If the probability the first truck is available is 0.80, the probability the second truck is available is 0.60, and the probability that both trucks are available is .30, what is the probability neither truck is available?

2.11 Conditional Probability

The conditional probability of an event A, *given* an event B with $P(B) > 0$, is defined by,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example 2.6 (Walpole et al. 2017a , 9th ed., page 63): Suppose that our sample space Ω is the population of adults in a small town who have completed the requirements for a college degree. We shall categorize them according to gender and employment status. The data are given in Table 2.1.

Table 2.1: Categorization of the Adults in a Small Town

	Employed	Unployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

A person is selected at random. What is the probability that the selected person is:

(i) a Male

(ii) a Female and employed

(iii) a Male or unemployed

(iv) a Male given that he is employed

2.12 The Multiplication Rule

If in an experiment the events A and B can both occur, then

$$P(A \cap B) = P(A)P(B|A); \quad \textit{provided} \;\; P(A) > 0$$

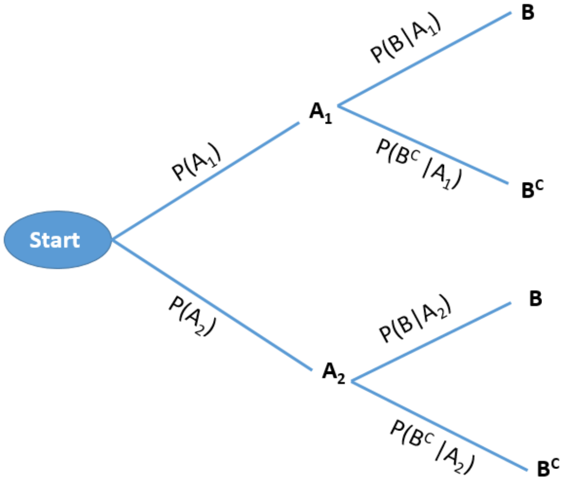
In general, assuming that all of the conditioning events (let, 3 events) have positive probability, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

Example 2.7 [Walpole et al. (2017a) , Example 2.36]**:** Suppose that we have a fuse box containing 20 fuses, of which 5 are defective. If 2 fuses are selected at random and removed from the box in succession without replacing the first, what is the probability that both fuses are defective?

2.13 Probability trees

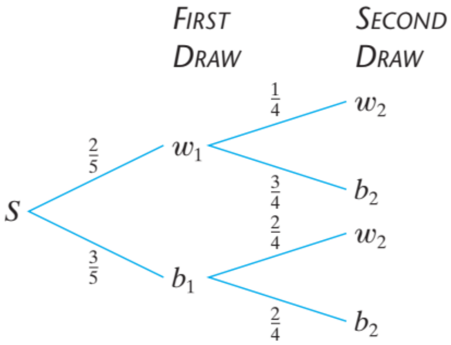
Consider a sequential experiment where in the **first stage** either A_1 or A_2 can be happened with some probabilities . And in the **second stage** event B can be happened. If B^C is the complement of B then this experiment can be shown in the following **tree diagram.**



Example 2.8: Two balls are drawn in succession, without replacement, from a box containing 3 blue and 2 white balls .

i) What is the probability that both balls will be white?

Solution: Here, two balls are drawn in succession (one by one) without replacement. This experiment can be shown in the following tree:



The probability of drawing a white ball on the first draw and a white ball on the second draw (both are white) is:

$P(w_1 \cap w_2) = P(w_1)P(w_2|w_1) = (\frac{2}{5})(\frac{1}{4}) = \frac{1}{10}$

ii) What is the probability that the second ball is white?

Solution:

$P(w_2) = P(w_1 \cap w_2) + P(b_1 \cap w_2)$

$= P(w_1)P(w_2|w_1) + P(b_1)P(w_2|b_1)$

$= (\frac{2}{5})(\frac{1}{4}) + (\frac{3}{5})(\frac{2}{4}) = \frac{1}{10} + \frac{3}{10} = \frac{4}{10} = \frac{2}{5}$

***Example 2.9 [Walpole et al. (2017a),** Example 2.37]**:** One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black? ***(Hints:** Apply probability tree)*

Example 2.10-Radar Detection (Bertsekas and Tsitsiklis 2008) **:** If an aircraft is present in a certain area, radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of no aircraft presence and a false alarm? What is the probability of aircraft presence and no detection?

2.14 Independent events

If two events A and B are independent, the probability that both of them occur is equal to the product of their individual probabilities i.e.

$$P(A \cap B) = P(A)P(B)$$

- **Corollary:** If A and B are independent events then their complement events also be independent that is,

$$P(A^C \cap B^C) = P(A^C)P(B^C)$$

- **Independence Rule for Multiple events:**

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Example 2.11 (Walpole et al. 2017b, Exercise 2.89) A town has two fire engines operating independently. The probability that a specific engine is available when needed is 0.96.

(a) What is the probability that neither is available when needed?

(b) What is the probability that a fire engine is available when needed?

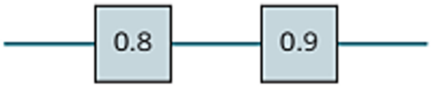
Example (Lind, Marchal, and Wathen 2012, 182) You take a trip by air that involves three independent flights. If there is an 80 percent chance each specific leg of the trip is done on time, what is the probability all three flights arrive on time?

Example (Lind, Marchal, and Wathen 2012, 182) The probability a HP network server is down is .05. If you have three independent servers, what is the probability that at least one of them is operational?

Example (Lind, Marchal, and Wathen 2012, 182) Twenty-two percent of all liquid crystal displays (LCDs) are manufactured by Samsung. What is the probability that in a collection of three independent LCD purchases, at least one is a Samsung?

2.15 System Reliability (Montgomery and Runger 2014b, 38)

- **Series circuit:** Suppose component L and R are connected in series from left to right . Also assume L and R operate (or fail) independently.

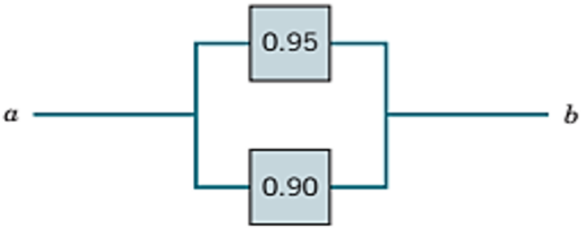


The probability that the circuit operates is

$$P(L \ \ and \ \ R) = P(L \cap R) = P(L)P(R) = 0.8 * 0.9 = 0.72$$

Practical interpretation: Notice that the probability that the circuit operates degrades to approximately 0.7 when all devices are required to be functional. The probability that each device is functional needs to be large for a circuit to operate when many devices are connected in series.

- **Parallel circuit:** The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



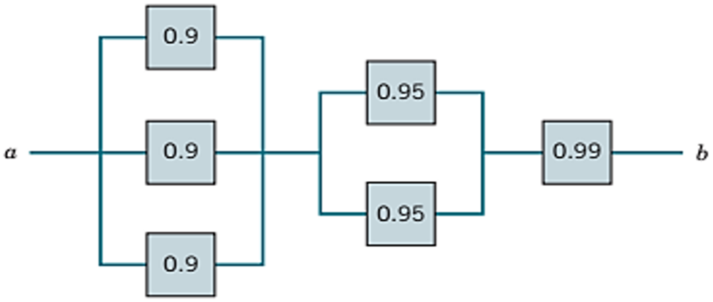
Let T and B denote the events that the top and bottom devices operate, respectively. There is a path if at least one device operates. The probability that the circuit operates is

$$P(T \quad or \quad B) = P(T \cup B) = 1 - P(T^C \cap B^C)$$

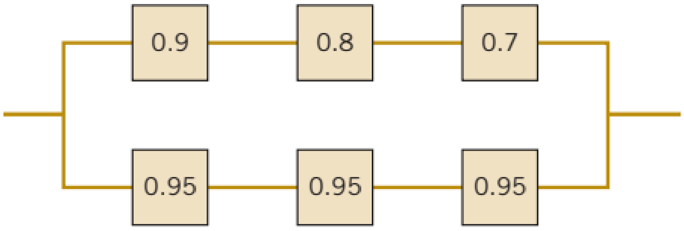
$$= 1 - P(T^C)P(B^C) = 1 - (0.05)(0.10) = 1 - 0.005 = 0.995$$

Practical Interpretation: Notice that the probability that the circuit operates is larger than the probability that either device is functional. This is an advantage of a parallel architecture. A disadvantage is that multiple devices are needed.

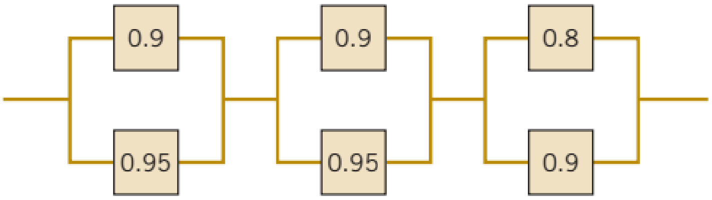
Advance circuit The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates? (**Ans.:** 0.9865)



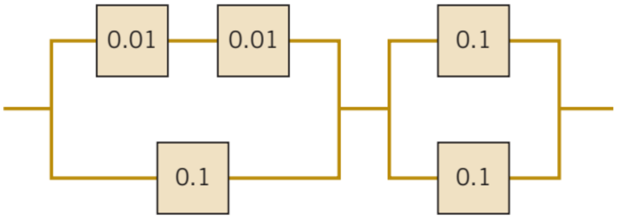
Example 2.12: The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates? (**Ans.:** 0.9293)



Example 2.13: The following circuit operates if and only if there is a path of functional devices from left to right. The probability that each device functions is as shown. Assume that the probability that a device is functional does not depend on whether or not other devices are functional. What is the probability that the circuit operates? (**Ans.:** 0.9702)



***Example 2.14:** The following circuit operates if and only if there is a path of functional devices from left to right. Assume devices fail independently and that the probability of *failure* of each device is as shown. What is the probability that the circuit operates?

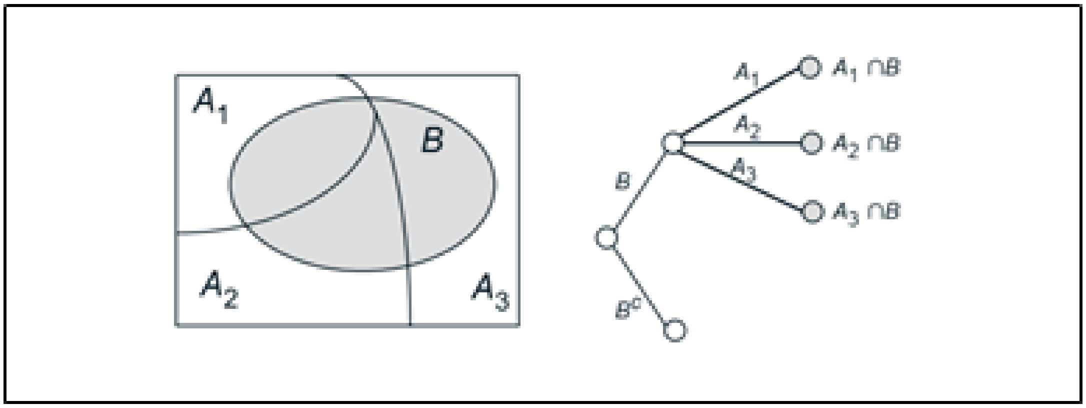


2.16 Total Probability Theorem and Bayes' Rule

In this section, we explore some applications of conditional probability. We start with the following theorem, which is often useful for computing the probabilities of various events, using a "divide-and-conquer" approach.

2.16.1 Total Probability Theorem

Let $A_1, ..., A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, ..., A_n$) and assume that $P(A_i) > 0$, for all $i = 1, ..., n$. Then, for any event B, we have



$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + ... + P(A_n \cap B)$$

$$= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ... + P(A_n)P(B|A_n)$$

Example 2.15: Suppose that A_1, A_2, A_3, and B are events where A_1, A_2, and A_3 are mutually exclusive and $P(A_1) = 0.2, P(A_2) = 0.5, P(A_3) = 0.3$. Also given $P(B|A_1) = 0.02, P(B|A_2) = 0.05, P(B|A_3) = 0.04$. Find $P(B)$.

2.16.2 Bayes' Rule/Theorem

Let $A_1, A_2, ..., A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i. Then, for any event B such that $P(B) > 0$, we have

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + ... + P(A_n)P(B|A_n)}$$

Example 2.16 (Walpole, 9th ed, Example 2.41,page 74)**:** In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Example 2.17 (Walpole, 9th ed, Example 2.42)**:** With reference to Example 2.41, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B3?

Example 2.18 (Walpole, 9th ed, Exercise 2.101) A paint-store chain produces and sells latex and semigloss paint. Based on long-range sales, the probability that a customer will purchase latex paint is 0.75. Of those that purchase latex paint, 60% also purchase rollers. But only 30% of semigloss paint buyers purchase rollers. A randomly selected buyer purchases a roller and a can of paint. What is the probability that the paint is latex?

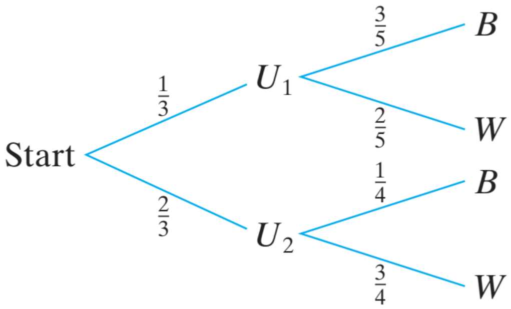
Example 2.19 (Radar Detection revisited): If an aircraft is present in a certain area, a radar detects it and generates an alarm signal with probability 0.99. If an aircraft is not present. the radar generates a (false) alarm, with probability 0.10. We assume that an aircraft is present with probability 0.05. What is the probability of no aircraft presence and a false alarm?

i) What is the probability that the radar generates alarm?
ii) If the radar generates alarm, what is the probability that there was an aircraft?
iii) If the radar does not generate alarm, what is the probability that there was not any aircraft?

Example 2.20(Montgomery, 6th ed., Exercise 2-179)**:** An e-mail filter is planned to separate valid e-mails from spam. The word free occurs in 60% of the spam messages and only 4% of the valid messages. Also, 20% of the messages are spam. Determine the following probabilities:

(a) The message contains free.
(b) The message is spam given that it contains free.
(c) The message is valid given that it does not contain free.

Example 2.21: One urn has 3 blue and 2 white balls; a second urn has 1 blue and 3 white balls. A single fair die is rolled and if 1 or 2 comes up, a ball is drawn out of the first urn; otherwise, a ball is drawn out of the second urn. If the drawn ball is blue, what is the probability that it came out of the first urn? Out of the second urn?



***Example 2.22:** A binary communication channel carries data as one of two sets of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1, and a transmitted 1 is sometimes received as a 0. For a given channel, it can be assumed that a transmitted 0 is correctly received with probability 0.95 and a transmitted 1 is correctly received with probability 0.75. Also, 60% of all messages are transmitted as a 0. If a signal is sent, determine the probability that:

(a) a 1 was received;

(b) a 0 was received;

(c) an error occurred;

(d) a 1 was transmitted given that a 1 was received ;

(e) a 0 was transmitted given that a 0 was received.

3 Random variable: Discrete

In many situations, it is desirable to assign a numerical value to each outcome of a random experiment. Such an assignment is called a **random variable.**

Mathematically, a random variable is a **real-valued function** of the experimental outcome.

Definition: A random variable is a function that associates a real number with each element in the sample space.

ℹ Note

We shall use a capital letter, say X, to denote a random variable and its corresponding small letter, x in this case, for one of its values.

Illustration: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls.

- Let, Y is the number of red balls.
- The small letter y is the numerical value for each possible outcomes.

The possible outcomes and the values of y of the random variable Y are:

Sample space	y
RR	2
RB	1
BR	1
BB	0

Since the values of Y is determined by a random experiment so, Y is a random variable (discrete).

3.1 Types of random variable

There are two important types of random variables, **discrete** and **continuous**.

A **discrete random variable** is one whose possible values form a discrete set; in other words, the values can be ordered, and there are gaps between adjacent values. The random variable Y, just described, is discrete.

In contrast, the possible values of a **continuous random variable** always contain an interval, that is, all the points between some two numbers.

3.2 Discrete random variable and Probability mass function

Suppose X is a discrete random variable. The **probability mass function (PMF)** of X can be denoted as $f(x)$ where

$$f(x) = P(X = x)$$

For each possible outcome x ; $f(x)$ must satisfies:

1. $$f(x) \geq 0$$
2. $$\sum_x f(x) = 1$$

3.2.1 The Cumulative Distribution Function of a Discrete Random Variable

A function called the **cumulative distribution function (CDF)** specifies the probability that a random variable is less than or equal to a given value. The cumulative distribution function of the random variable X is the function

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

Example 3.1: Calculating probabilities A certain industrial process is brought down for recalibration whenever the quality of the items produced falls below specifications. Let X represent the number of times the process is recalibrated during a week, and assume that X has the following probability mass function.

x	0	1	2	3	4
$f(x)$	0.35	0.25	0.20	0.15	0.05

Compute the following :

i) $P(X = 2)$;

ii) $P(X < 3)$ and $P(X > 2)$;

iii) $F(2)$

Example (Walpole et al. 2017a, Example 3.8)A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, **find** the probability distribution for *the number of defectives*.

Solution: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and 2. Now ,

$P(X = 0) = f(0) = P(N, N) = \frac{\binom{17}{2}\binom{3}{0}}{\binom{20}{2}} = \frac{136}{190}$

$P(X = 1) = f(1) = P(D, N) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$

$P(X = 2) = f(2) = P(D, D) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{136}{190}$	$\frac{51}{190}$	$\frac{3}{190}$

Example (Baron 2019, Exercise 3.1 (a)) A computer virus is trying to corrupt two files. The first file will be corrupted with probability 0.4. Independently of it, the second file will be corrupted with probability 0.3.

Compute the probability mass function (**PMF**) of X, *the number of corrupted files*.

Solution: (Will be discussed in class)

3.2.2 Expectation (Population Mean) of discrete random variable

Let X be a discrete random variable with probability mass function $f(x) = P(X = x)$.

The mean of X is given by

$$\mu = \sum_x x.f(x)$$

The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$ or by μ.

Example 3.2 : Let X represent the number of times the process is recalibrated during a week, and assume that X has the following probability mass function.

x	0	1	2	3	4
$f(x)$	0.35	0.25	0.20	0.15	0.05

Compute the **mean** or **expected value** of X.

Solution:

The **expected value** of X is:

$$\mu = E[X] = \sum_{x=0}^{4} x.f(x)$$

$$= 0(0.35) + 1(0.25) + 2(0.20) + 3(0.15) + 4(0.05) = 1.30$$

3.2.3 Variance (population) of discrete random variable

Let X be a discrete random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$var(X) = \sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2$$

Where,

$$E(X^2) = \sum_{x} x^2.f(x)$$

Example 3.2 : Let X represent the number of times the process is recalibrated during a week, and assume that X has the following probability mass function.

Table 3.5: Compute the **expected value** and **variance** of X.

x	0	1	2	3	4
$f(x)$	0.35	0.25	0.20	0.15	0.05

Solution: From Example 3.1, we have

Expected value of X is $\mu = E(X) = 1.30$

Now,

$$E(X^2) = \sum_{x=0}^{4} x^2.f(x)$$

$= 0^2(0.35) + 1^2(0.25) + 2^2(0.20) + 3^2(0.15) + 4^2(0.05)$

$= 3.20$

Hence, $var(X) = \sigma^2 = E(X^2) - \mu^2 = 3.20 - (1.30)^2 = 1.51$

- The standard deviation is the square root of the variance:

  $\sigma = \sqrt{var(X)}$

ℹ Properties of E(.) and var(.)

If a and b are constants, then
a) $E(b) = b$
b) $E(aX + b) = aE(X) + b$
c) $var(b) = 0$
d) $var(aX + b) = a^2 \quad var(X)$

Example Computer chips often contain surface imperfections. For a certain type of computer chip, the probability mass function of the number of defects X is presented in the following table.

x	**0**	**1**	**2**	**3**	**4**
$f(x)$	0.4	0.3	0.15	0.1	0.05

a. Find $P(X \leq 2)$.

b. Find $P(X > 1)$.

c. Find expected value/average of X.

d. Find variance and standard deviation of X.

e. Also find $E(2X + 5)$ and $var(2X + 5)$.

4 Some special discrete random variables

4.1 Bernoulli r.v

Bernoulli r.v comes from **Bernoulli trial-**a trial which has **TWO** possible outcomes (*success* or *failure*).

Consider the toss of a **biased coin**, which comes up a head with probability p, and a tail with probability $1-p$. The **Bernoulli random variable** takes the two values 1 and 0, depending on whether the outcome is a head or a tail:

$$X = 1; if \quad a \quad head, X = 0; if \quad a \quad tail.$$

Then the PMF of X is:

- **PMF**:

$$P(X = x) = f(x) = p^x(1-p)^{1-x}; \quad x = 0, 1 \tag{4.1}$$

- **Mean**: $\mu = E(X) = p$
- **Variance**: $\sigma^2 = E(X-\mu)^2 = E(X^2) - \mu^2 = p(1-p)$

For all its simplicity, the Bernoulli random variable is very important. In practice, it is used to model generic probabilistic situations with just two outcomes, such as:

(a) The state of a telephone at a given time that can be either free or busy.

(b) A person who can be either healthy or sick with a certain disease.

(c) The preference of a person who can be either for or against a certain political candidate.

Furthermore, by combining multiple Bernoulli random variables, one can construct more complicated random variables.

ℹ Derivation of Mean and Variance of Bernoulli r.v

Mean:

$$E(X) = \sum_{x=0}^{1} x \cdot f(x) = (0)f(0) + (1)f(1) = 0 + 1 \cdot p = p$$

Variance:
$Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$

4.2 Binomial r.v

In a Binomial experiment , the **Bernoulli trial** is repeated n times with the following conditions:

a) The trials are independent

b) In each trial $P(success) = p$ remains constant

Suppose $X = number\ \ of\ \ successs\ \ in\ \ n\ \ trials$. Then X is called a **Binomial r.v** or follows **Binomial distribution.**

PMF:

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}; x = 0, 1, 2, ..., n \tag{4.2}$$

CDF: $P(X \leq x) = F(x) = f(0) + f(1) + ... + f(x)$

Properties

a) $\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$

b) Mean: $E(X) = np$

c) Variance: $Var(X) = np(1-p)$

We write $X \sim Bin(n, p)$

Illustration

- Consider an experiment of tossing a biased coin 3(number of trials, n) times.
- Tosses are *independent*, each toss has only **TWO** Outcomes-*Head (Success)* and *Tail (Failure)*

This type of trial is called the *Bernoulli Trial*

- Suppose, $P(H) = p$ and remain constant in each toss, consequently, $P(T) = 1 - p = q$ (let).

Suppose, $X = \#\ \ of\ \ head\ \ (successes)\ \ in\ \ 3\ \ tosses$

Now, what is the probability that, we will have **exactly 2 heads (success) in 3 tosses?**

That is, $P(X = 2) = ?$

Now, this can happen in the following ways:

$$P(X = 2) = P(HHT) + P(HTH) + P(THH)$$
$$= P(H)P(H)P(T) + P(H)P(T)P(H) + P(T)P(H)P(H)$$

$[Since\ tosses\ are\ independent]$

$$= p.p.q + p.q.p + q.p.p$$
$$= p^2q + p^2q + p^2q = 3p^2q$$
$$\therefore P(X = 2) = \binom{3}{2} p^2 q^{3-2}$$

If, $p = 0.6$ is given, then we can easily compute $P(X = 2) = f(2)$. Now, if we repeat the toss 10 times ($n = 10$), with $P(H) = p$, what is the value of $P(X = 3) = f(3)$?

ℹ Note

- n and p are said to be the *parameters* of the Binomial distribution.
- $f(x) = F(x) - F(x-1)$ i.e $f(3) = F(3) - F(2)$
- If $Y = number\ \ of\ \ failures\ \ in\ \ n\ \ trials$ then $Y \sim Bin(n, 1-p)$

ℹ Relation between Bernoulli r.v and Binomial r.v

A **Binomial Random Variable** Is a **Sum of Bernoulli Random Variables**
Let, Y_i is a Bernoulli r.v appeared in i^{th} Bernoulli trial. If we conduct n independent Bernoulli trials then we have n independent Bernoulli r.vs such as $Y_1, Y_2, ..., Y_n$. Each Y_i has values of either 1 or 0.
Now if X is a Binomial r.v then,

$$X = Y_1 + Y_2 + ... + Y_n = \sum_{i=1}^{n} Y_i$$

Derivation of Mean and Variance of Binomial r.v

From previous note, we know if Y_i is a Bernoulli r.v then
$E(Y_i) = p$ and $Var(Y_i) = p(1-p)$
So, the mean of Binomial r.v that is

$$E(X) = E(Y_1 + Y_2 + ... + Y_n)$$

$$= E(Y_1) + E(Y_2) + ... + E(Y_n)$$

$$= p + p + ... + p = np$$

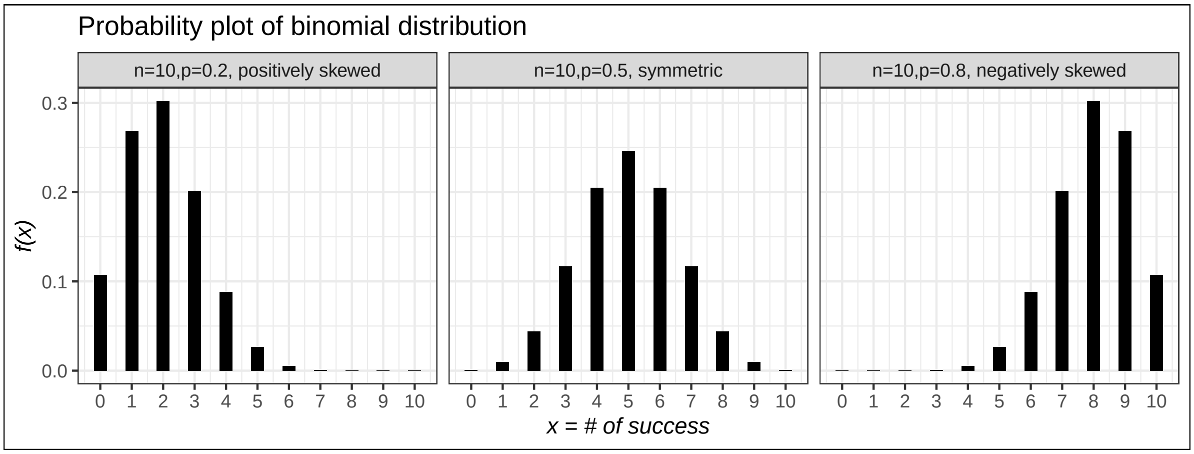
Now, the variance of X is:

$$Var(X) = Var(Y_1 + Y_2 + ... + Y_n)$$

$$= Var(Y_1) + Var(Y_2) + ... + Var(Y_n)$$

$$= p(1-p) + p(1-p) + ... + p(1-p) = np(1-p)$$

Probability plot of binomial r.v for different values of p and shape characteristics



Finding Binomial probability manually

Suppose, $X \sim Binom(n, p)$; where $n = 5$ and $p = 0.6$. Find, *(i)* $P(X = 2)$ *(ii)* $P(X \leq 2)$ *(iii)* $P(X \geq 3)$.

Solution:

PMF of X: $P(X = x) = f(x) = \binom{5}{x} 0.6^x (0.4)^{5-x} \quad ; x = 0, 1, 2, ..., 5$

(i) $P(X=2)=f(2)=\binom{5}{2}0.6^2(0.4)^{5-2}=0.2304$

(ii) $P(X\leq 2)=F(2)=f(0)+f(1)+f(2)=0.0102+0.0768+0.2304=0.3174$

(iii) $P(X\geq 3)=f(3)+f(4)+f(5)=0.6826$

Alternative:(iii)

$P(X\geq 3)=1-P(X<3)=1-P(X\leq 2)=1-F(2)=1-0.3174=0.6826$

Finding Binomial probability using Binomial Table

In the end of any Statistics book there are some **Probability Distribution Table**. We can use these table to compute the required probability for *specific values of the parameters* of certain probability distribution. Here I share the 1st page of **Binomial distribution table** (Baron 2019).

Table A2. Binomial distribution

$$F(x) = \boldsymbol{P}\{X \le x\} = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

n	x	p																		
		.050	.100	.150	.200	.250	.300	.350	.400	.450	.500	.550	.600	.650	.700	.750	.800	.850	.900	.950
1	0	.950	.900	.850	.800	.750	.700	.650	.600	.550	.500	.450	.400	.350	.300	.250	.200	.150	.100	.050
2	0	.903	.810	.723	.640	.563	.490	.423	.360	.303	.250	.203	.160	.123	.090	.063	.040	.023	.010	.003
	1	.998	.990	.978	.960	.938	.910	.878	.840	.798	.750	.698	.640	.578	.510	.438	.360	.278	.190	.098
3	0	.857	.729	.614	.512	.422	.343	.275	.216	.166	.125	.091	.064	.043	.027	.016	.008	.003	.001	.000
	1	.993	.972	.939	.896	.844	.784	.718	.648	.575	.500	.425	.352	.282	.216	.156	.104	.061	.028	.007
	2	1.0	.999	.997	.992	.984	.973	.957	.936	.909	.875	.834	.784	.725	.657	.578	.488	.386	.271	.143
4	0	.815	.656	.522	.410	.316	.240	.179	.130	.092	.063	.041	.026	.015	.008	.004	.002	.001	.000	.000
	1	.986	.948	.890	.819	.738	.652	.563	.475	.391	.313	.241	.179	.126	.084	.051	.027	.012	.004	.000
	2	1.0	.996	.988	.973	.949	.916	.874	.821	.759	.688	.609	.525	.437	.348	.262	.181	.110	.052	.014
	3	1.0	1.0	.999	.998	.996	.992	.985	.974	.959	.938	.908	.870	.821	.760	.684	.590	.478	.344	.185
5	0	.774	.590	.444	.328	.237	.168	.116	.078	.050	.031	.018	.010	.005	.002	.001	.000	.000	.000	.000
	1	.977	.919	.835	.737	.633	.528	.428	.337	.256	.188	.131	.087	.054	.031	.016	.007	.002	.000	.000
	2	.999	.991	.973	.942	.896	.837	.765	.683	.593	.500	.407	.317	.235	.163	.104	.058	.027	.009	.001
	3	1.0	1.0	.998	.993	.984	.969	.946	.913	.869	.813	.744	.663	.572	.472	.367	.263	.165	.081	.023
	4	1.0	1.0	1.0	1.0	.999	.998	.995	.990	.982	.969	.950	.922	.884	.832	.763	.672	.556	.410	.226
6	0	.735	.531	.377	.262	.178	.118	.075	.047	.028	.016	.008	.004	.002	.001	.000	.000	.000	.000	.000
	1	.967	.886	.776	.655	.534	.420	.319	.233	.164	.109	.069	.041	.022	.011	.005	.002	.000	.000	.000
	2	.998	.984	.953	.901	.831	.744	.647	.544	.442	.344	.255	.179	.117	.070	.038	.017	.006	.001	.000
	3	1.0	.999	.994	.983	.962	.930	.883	.821	.745	.656	.558	.456	.353	.256	.169	.099	.047	.016	.002
	4	1.0	1.0	1.0	.998	.995	.989	.978	.959	.931	.891	.836	.767	.681	.580	.466	.345	.224	.114	.033
	5	1.0	1.0	1.0	1.0	1.0	.999	.998	.996	.992	.984	.972	.953	.925	.882	.822	.738	.623	.469	.265
7	0	.698	.478	.321	.210	.133	.082	.049	.028	.015	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000
	1	.956	.850	.717	.577	.445	.329	.234	.159	.102	.063	.036	.019	.009	.004	.001	.000	.000	.000	.000
	2	.996	.974	.926	.852	.756	.647	.532	.420	.316	.227	.153	.096	.056	.029	.013	.005	.001	.000	.000
	3	1.0	.997	.988	.967	.929	.874	.800	.710	.608	.500	.392	.290	.200	.126	.071	.033	.012	.003	.000
	4	1.0	1.0	.999	.995	.987	.971	.944	.904	.847	.773	.684	.580	.468	.353	.244	.148	.074	.026	.004
	5	1.0	1.0	1.0	1.0	.999	.996	.991	.981	.964	.938	.898	.841	.766	.671	.555	.423	.283	.150	.044
	6	1.0	1.0	1.0	1.0	1.0	1.0	.999	.998	.996	.992	.985	.972	.951	.918	.867	.790	.679	.522	.302
8	0	.663	.430	.272	.168	.100	.058	.032	.017	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000	.000
	1	.943	.813	.657	.503	.367	.255	.169	.106	.063	.035	.018	.009	.004	.001	.000	.000	.000	.000	.000
	2	.994	.962	.895	.797	.679	.552	.428	.315	.220	.145	.088	.050	.025	.011	.004	.001	.000	.000	.000
	3	1.0	.995	.979	.944	.886	.806	.706	.594	.477	.363	.260	.174	.106	.058	.027	.010	.003	.000	.000
	4	1.0	1.0	.997	.990	.973	.942	.894	.826	.740	.637	.523	.406	.294	.194	.114	.056	.021	.005	.000
	5	1.0	1.0	1.0	.999	.996	.989	.975	.950	.912	.855	.780	.685	.572	.448	.321	.203	.105	.038	.006
	6	1.0	1.0	1.0	1.0	1.0	.999	.996	.991	.982	.965	.937	.894	.831	.745	.633	.497	.343	.187	.057
	7	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.999	.998	.996	.992	.983	.968	.942	.900	.832	.728	.570	.337
9	0	.630	.387	.232	.134	.075	.040	.021	.010	.005	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
	1	.929	.775	.599	.436	.300	.196	.121	.071	.039	.020	.009	.004	.001	.000	.000	.000	.000	.000	.000
	2	.992	.947	.859	.738	.601	.463	.337	.232	.150	.090	.050	.025	.011	.004	.001	.000	.000	.000	.000
	3	.999	.992	.966	.914	.834	.730	.609	.483	.361	.254	.166	.099	.054	.025	.010	.003	.001	.000	.000
	4	1.0	.999	.994	.980	.951	.901	.828	.733	.621	.500	.379	.267	.172	.099	.049	.020	.006	.001	.000
	5	1.0	1.0	.999	.997	.990	.975	.946	.901	.834	.746	.639	.517	.391	.270	.166	.086	.034	.008	.001
	6	1.0	1.0	1.0	1.0	.999	.996	.989	.975	.950	.910	.850	.768	.663	.537	.399	.262	.141	.053	.008
	7	1.0	1.0	1.0	1.0	1.0	1.0	.999	.996	.991	.980	.961	.929	.879	.804	.700	.564	.401	.225	.071
	8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.999	.998	.995	.990	.979	.960	.925	.866	.768	.613	.370
10	0	.599	.349	.197	.107	.056	.028	.013	.006	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.914	.736	.544	.376	.244	.149	.086	.046	.023	.011	.005	.002	.001	.000	.000	.000	.000	.000	.000
	2	.988	.930	.820	.678	.526	.383	.262	.167	.100	.055	.027	.012	.005	.002	.000	.000	.000	.000	.000
	3	.999	.987	.950	.879	.776	.650	.514	.382	.266	.172	.102	.055	.026	.011	.004	.001	.000	.000	.000
	4	1.0	.998	.990	.967	.922	.850	.751	.633	.504	.377	.262	.166	.095	.047	.020	.006	.001	.000	.000
	5	1.0	1.0	.999	.994	.980	.953	.905	.834	.738	.623	.496	.367	.249	.150	.078	.033	.010	.002	.000
	6	1.0	1.0	1.0	.999	.996	.989	.974	.945	.898	.828	.734	.618	.486	.350	.224	.121	.050	.013	.001
	7	1.0	1.0	1.0	1.0	1.0	.998	.995	.988	.973	.945	.900	.833	.738	.617	.474	.322	.180	.070	.012
	8	1.0	1.0	1.0	1.0	1.0	1.0	.999	.998	.995	.989	.977	.954	.914	.851	.756	.624	.456	.264	.086
	9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.999	.997	.994	.987	.972	.944	.893	.803	.651	.401

Suppose, $X \sim Binom(n,p)$; where $n = 5$ and $p = 0.6$. Find, *(i)* $P(X = 2)$ *(ii)* $P(X \leq 2)$ *(iii)* $P(X \geq 3)$ using **Table**.

Solution:

(i) $P(X = 2) = f(2) = F(2) - F(1) = 0.3174 - 0.0870 = 0.2304$.

(ii) $P(X \leq 2 = F(2) = 0.3174$

(iii) $P(X \geq 3) = 1 - P(X < 3) = 1 - F(2) = 1 - 0.3174 = 0.6826$

Exercise(Walpole et al. 2017a)

5.9 In testing a certain kind of truck tire over rugged terrain, it is found that 25% of the trucks fail to complete the test run without a blowout. Of the next 15 trucks tested, find the probability that: (a) from 3 to 6 have blowouts; (b) fewer than 4 have blowouts; (c) more than 5 have blowouts.

Solution:

Let, X= *number of trucks that have blowouts*

Given, $n = 15$; $p = Pr(blowout) = 0.25$; $q = 1 - p = 0.75$. Hence, $X \sim Binom(n = 15, p = 0.25)$, that is:

$$P(X = x) = f(x) = \binom{15}{x}(0.25)^x(0.75)^{15-x}; x = 0, 1, 2, ..., 15.$$

Now,

(a) $P(3 \leq X \leq 6) = f(3) + f(4) + f(5) + f(6)$=0.225+0.225+0.165+0.092=0.707.

Alternative: $P(3 \leq X \leq 6) = F(6) - F(2) = 0.943 - 0.236 = 0.707$ [from **Table**]

(b) $P(X < 4) = f(0) + f(1) + f(2) + f(3)$=0.013+0.067+0.156+0.225=0.461.

Alternative: $P(X < 4) = F(3) = 0.461$ [from **Table A2**]

(c) $P(X > 5) = 1 - P(X \leq 5) = 1 - F(5)$=0.148

5.12 A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

5.16 Suppose that airplane engines operate independently and fail with probability equal to 0.4. Assuming that a plane makes a safe flight if at least one-half of its engines run, determine whether a 4-engine plane or a 2-engine plane has the higher probability for a successful flight.

5.25 Suppose that for a very large shipment of integrated-circuit chips, the probability of failure for any one chip is 0.10. Assuming that the assumptions underlying the binomial distributions are met, find the probability that at most 3 chips fail in a random sample of 20.

Exercise(Montgomery and Runger 2014c)

3-93 Let X be a binomial random variable with $p = 0.1$. and $n = 10$. Calculate the following probabilities from the binomial probability mass function and from the binomial table in Appendix A and compare results. (a) $P(X \leq 2)$ (b) P(X>8) (c) P(X = 4) (d) $P(5 \leq X \leq 7)$

3-115 The probability that a visitor to a Web site provides contact data for additional information is 0.01. Assume that 1000 visitors to the site behave independently. Determine the following probabilities: (a) No visitor provides contact data. (b) Exactly 10 visitors provide contact data. (c) More than 3 visitors provide contact data

Exercise(Baron 2019)

3.21. A lab network consisting of 20 computers was attacked by a computer virus. This virus enters each computer with probability 0.4, independently of other computers. Find the probability that it entered at least 10 computers.

3.22. Five percent of computer parts produced by a certain supplier are defective. What is the probability that a sample of 16 parts contains more than 3 defective ones?

And so on….

4.3 Geometric r.v

The number of Bernoulli trials needed to get the first success has Geometric distribution.

Let X be the number of trials needed to get the *first* success.

PMF:

$P(X = x) = P(the \;\; 1^{st} \;\; success \;\; occurs \;\; in \;\; the \;\; x^{th \;\; trial})$

$$f(x) = (1-p)^{x-1}p \;\; ; x = 1, 2, ..., \infty \tag{4.3}$$

Properties

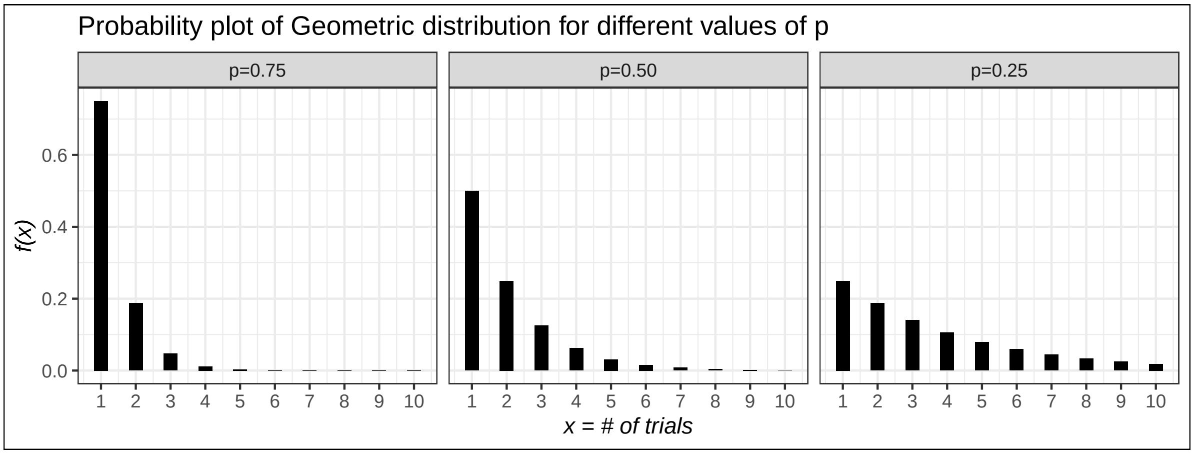
a) $\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} (1-p)^{x-1}p = 1$

b) $E(X) = \frac{1}{p}$

c) $Var(X) = \frac{1-p}{p^2}$

We write, $X \sim Geom(p)$

Probability plot for different values of p



The probability plot illustrates that as the value of (p) increases, the likelihood of achieving the first success at an earlier trial also increases.

Memorylessness property of Geometric distribution

Let X be the number of trials needed to get the *first* success and $X \sim Geom(p)$ .Suppose that we have been watching the process for n trials and no success has been recorded. What is the probability that in additional k trials we will observe the *first* success? Mathematically,

$$P(X = n + k|X > n) = ?$$

It can be shown that:

$$P(X = n + k|X > n) = P(X = k)$$

It implies that the probability of needing k more trials, given that we've already had n failures, is the same as the original probability of needing k trials to get the *first* success. This is known as **memorylessness property.**

Real life examples

- A search engine goes through a list of sites looking for a given key phrase. Suppose the search terminates as soon as the key phrase is found. The number of sites visited is Geometric.
- A hiring manager interviews candidates, one by one, to fill a vacancy. The number of candidates interviewed until one candidate receives an offer has Geometric distribution.

Example 5.15 (Walpole): For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Example 5.16 (Walpole): At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. *We are interested in knowing the probability that 5 attempts are necessary for* ***a successful call.***

Quite often, in applications dealing with the geometric distribution, the mean and variance are important. For example, in **Example 5.16,** the *expected* number of calls necessary to make a connection is quite important. So, we can easily compute mean and variance of a Geometric r.v.

Exercise 5.55 (Walpole) The probability that a student pilot passes the written test for a private pilot's license is 0.7. Find the probability that a given student will pass the test

(a) on the third try;

(b) before the fourth try

Exercise 3.24 (Michael Baron). An internet search engine looks for a certain keyword in a sequence of independent web sites. It is believed that 20% of the sites contain this keyword. Compute the probability that the search engine had to visit at least 5 sites in order to find the first occurrence of a keyword.

A memoryless Customer Service Call center

A customer calls a tech support center, which takes independent attempts to resolve an issue. On each call, there is a 20% chance that the issue is resolved.

(a) What is the probability that the issue will be resolved after the 4th call?
(b) Suppose the customer has already called 4times and the issue is still unresolved. What is the probability that it will still take more than 3 additional calls to get the issue resolved?
(c) Explain why the answer in (b) makes sense in terms of the memoryless property.

ℹ Derivation and proofs related to Geometric distribution

1) Memorylessness property
Proof:
PMF of Geometric r.v is : $f(x) = q^{x-1}p \;\; ; x = 1, 2, ..., \infty$ with $q = 1 - p$.
Now,
$P(X = n + k|X > n) = \frac{P(\{X=n+k\}\cap\{X>n\})}{P(X>n)} = \frac{P(X=n+k)}{P(X>n)}$
$= \frac{q^{n+k-1}}{q^n(1-q)^{-1}} = \frac{q^{k-1}}{p^{-1}} = q^{k-1}p = P(X = k)$

$\therefore P(X = n + k|X > n) = P(X = k)$

2) Mean of geometric distribution
Remember the geometric series :
$1 + q + q^2 + q^3 + ... = \sum_{x=0}^{\infty} q^x = \frac{1}{1-q} = (1 - q)^{-1}$
Now,

$$\frac{d}{dq}\sum_{x=0}^{\infty} q^x = \frac{d}{dq}(1 - q)^{-1}$$

$$Or, \sum_{x=1}^{\infty} xq^{x-1} = (1 - q)^{-2} \quad ...(A)$$

So, $E(X) = \sum_{x=1}^{\infty} x \cdot q^{x-1}p = p\sum_{x=1}^{\infty} x \cdot q^{x-1} = p(1 - q)^{-2}$
$= pp^{-2} = p^{-1} = \frac{1}{p}$
3) Variance
$Var(X) = E(X^2) - [E(X)]^2 = E[X(X - 1)] + E(X) - [E(X)]^2 \quad ...(B)$
To derive the variance of geometric distribution, differentiate (A) w.r.t q again. After some manipulation and using (B) we will have

$$Var(X) = \frac{1 - p}{p^2}$$

4.4 Poisson r.v

The number of events occur randomly in an interval or in a region usually follows Poisson distribution. A famous French mathematician *SimB4eon-Denis Poisson* (1781–1840) first introduced this distribution.

Example

The Poisson distribution may be useful to model variables like:

- The *no. of calls* arrive at a customer care in *15 minites*
- The *no. of arrivals* at a car wash in *one hour*
- The *no. of repairs* needed in *10 miles* of highway
- The *no. of leaks* in *100 miles* of pipeline etc.

Usually Poisson distribution is used to evaluate probability of “*Rare*” event.

The probability mass function of the Poisson random variable X, representing the *number of outcomes* occurring in a given time interval denoted by t, is:

- **PMF**:

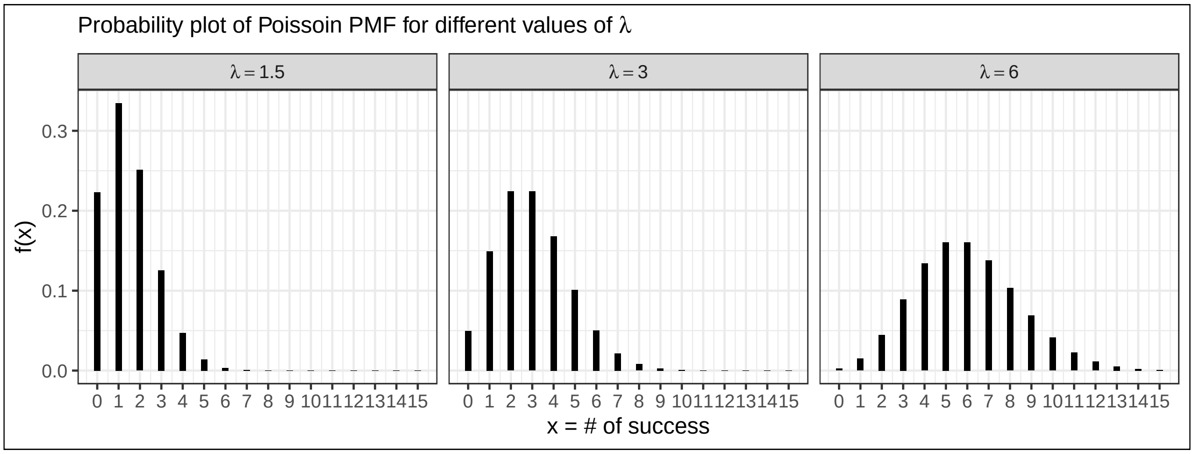
$$P(X=x)=f(x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!};\quad x=0,1,2,...,\infty. \tag{4.4}$$

Here, λ is called *arrival rate* or *average number of occurrences* in long-run. And only *parameter* of Poisson distribution.

- **Mean**: $\mu=E(X)=\lambda t$
- **Variance**: $\sigma^2=\lambda t$
- **We write:** $X\sim Pois(\lambda t)$

N.B: The mean and variance of Poisson random random variable are *identical.* This is the *unique property* of Poisson r.v.

Probability of plot of poisson r.v for different values of λ (for a fixed interval $t=1$)



We can see that, for small λ the distribution of Poisson r.v is positively skewed and as the value of λ increases the distribution tends to symmetry.

Finding Poisson probability

Consider a discrete r.v say $X\sim Pois(\lambda t)$. Suppose, $\lambda=1.5$ and $t=2$. Find, (i) P(X=4) (ii)$P(X\leq 2)$ (iii) $P(X\geq 3)$.

Solution:

PMF of X: $P(X=x)=f(x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!};x=0,1,...,\infty.$

(i) For $t=2$, $\mu=\lambda t=1.5*2=3.$

So, $P(X=4)=f(4)=\frac{e^{-3}(3)^4}{4!}=0.1680$

(ii) $P(X \leq 2) = \sum_{x=0}^{2} f(x) = \sum_{x=0}^{2} \frac{e^{-3}(3)^x}{x!} = e^{-3}[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}] = 0.4232$

(iii) $P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - 0.423 = 0.5768$

Finding Poisson probability using Table

We can use Poisson distribution table to compute Poisson probabilities. Here I share the 1st page of **Poisson distribution table** (Baron 2019).

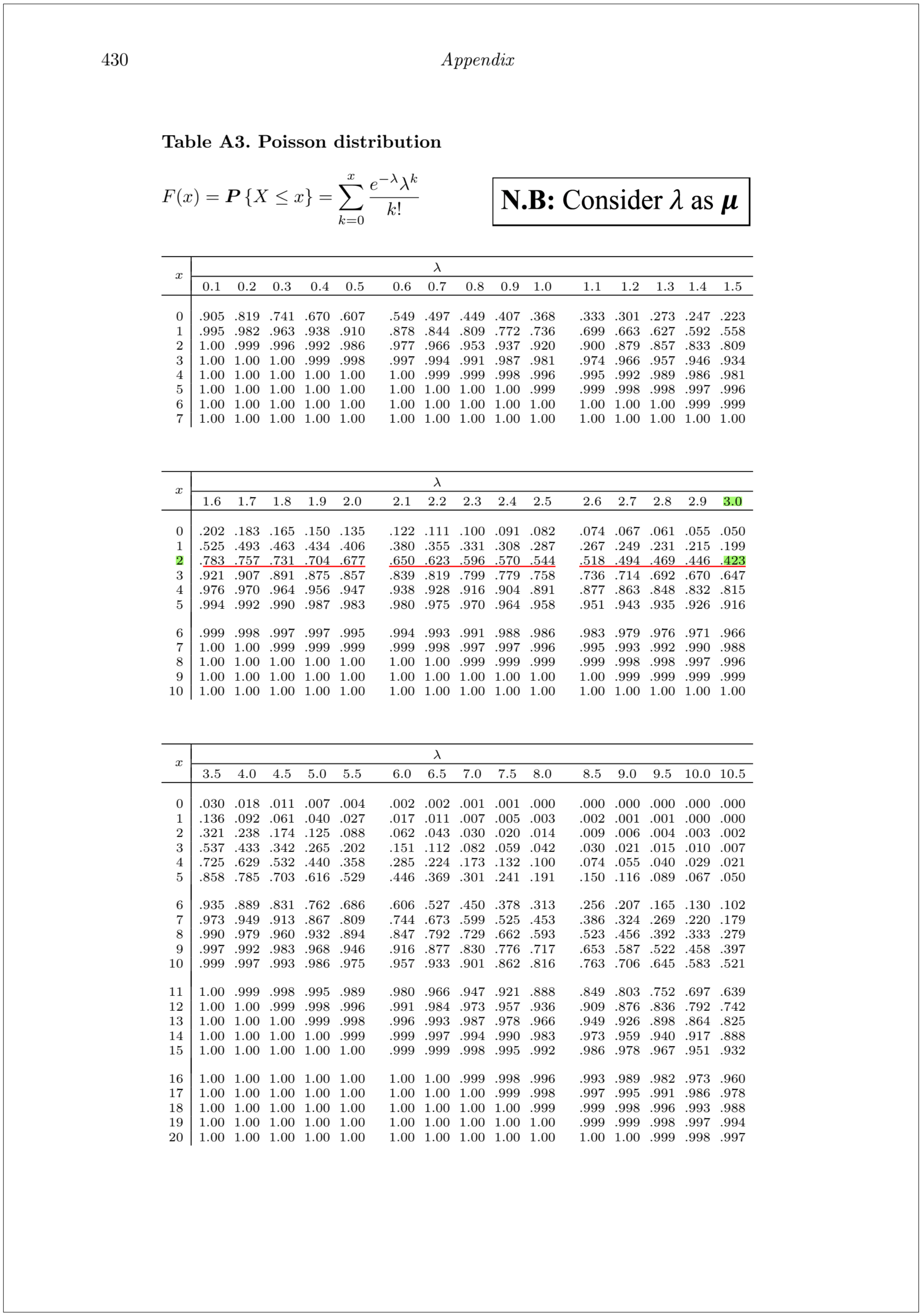


Table A3. Poisson distribution

$$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

N.B: Consider λ as $\boldsymbol{\mu}$

x	λ														
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
0	.905	.819	.741	.670	.607	.549	.497	.449	.407	.368	.333	.301	.273	.247	.223
1	.995	.982	.963	.938	.910	.878	.844	.809	.772	.736	.699	.663	.627	.592	.558
2	1.00	.999	.996	.992	.986	.977	.966	.953	.937	.920	.900	.879	.857	.833	.809
3	1.00	1.00	1.00	.999	.998	.997	.994	.991	.987	.981	.974	.966	.957	.946	.934
4	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.995	.992	.989	.986	.981
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.998	.997	.996
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

x	λ														
	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.202	.183	.165	.150	.135	.122	.111	.100	.091	.082	.074	.067	.061	.055	.050
1	.525	.493	.463	.434	.406	.380	.355	.331	.308	.287	.267	.249	.231	.215	.199
2	.783	.757	.731	.704	.677	.650	.623	.596	.570	.544	.518	.494	.469	.446	.423
3	.921	.907	.891	.875	.857	.839	.819	.799	.779	.758	.736	.714	.692	.670	.647
4	.976	.970	.964	.956	.947	.938	.928	.916	.904	.891	.877	.863	.848	.832	.815
5	.994	.992	.990	.987	.983	.980	.975	.970	.964	.958	.951	.943	.935	.926	.916
6	.999	.998	.997	.997	.995	.994	.993	.991	.988	.986	.983	.979	.976	.971	.966
7	1.00	1.00	.999	.999	.999	.999	.998	.997	.997	.996	.995	.993	.992	.990	.988
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999	.998	.998	.997	.996
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.999	.999
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

x	λ														
	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5
0	.030	.018	.011	.007	.004	.002	.002	.001	.001	.000	.000	.000	.000	.000	.000
1	.136	.092	.061	.040	.027	.017	.011	.007	.005	.003	.002	.001	.001	.000	.000
2	.321	.238	.174	.125	.088	.062	.043	.030	.020	.014	.009	.006	.004	.003	.002
3	.537	.433	.342	.265	.202	.151	.112	.082	.059	.042	.030	.021	.015	.010	.007
4	.725	.629	.532	.440	.358	.285	.224	.173	.132	.100	.074	.055	.040	.029	.021
5	.858	.785	.703	.616	.529	.446	.369	.301	.241	.191	.150	.116	.089	.067	.050
6	.935	.889	.831	.762	.686	.606	.527	.450	.378	.313	.256	.207	.165	.130	.102
7	.973	.949	.913	.867	.809	.744	.673	.599	.525	.453	.386	.324	.269	.220	.179
8	.990	.979	.960	.932	.894	.847	.792	.729	.662	.593	.523	.456	.392	.333	.279
9	.997	.992	.983	.968	.946	.916	.877	.830	.776	.717	.653	.587	.522	.458	.397
10	.999	.997	.993	.986	.975	.957	.933	.901	.862	.816	.763	.706	.645	.583	.521
11	1.00	.999	.998	.995	.989	.980	.966	.947	.921	.888	.849	.803	.752	.697	.639
12	1.00	1.00	.999	.998	.996	.991	.984	.973	.957	.936	.909	.876	.836	.792	.742
13	1.00	1.00	1.00	.999	.998	.996	.993	.987	.978	.966	.949	.926	.898	.864	.825
14	1.00	1.00	1.00	1.00	.999	.999	.997	.994	.990	.983	.973	.959	.940	.917	.888
15	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.995	.992	.986	.978	.967	.951	.932
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.993	.989	.982	.973	.960
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.997	.995	.991	.986	.978
18	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.996	.993	.988
19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.999	.998	.997	.994
20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.997

Consider a discrete r.v say $X \sim Pois(\lambda t)$. Suppose, $\lambda = 1.5$ and $t = 2$. **Find**, (i) $P(X \leq 2)$ (ii)P(X=4)

Solution by using Table:

For $t = 2$, $\mu = \lambda t = 1.5 * 2 = 3$.

(i) $P(X \leq 2) = F(2) = 0.423$

[For x=2 and $\mu \;\; or \; \lambda = 3$; corresponding probability in **Table A3** is 0.423]

(ii) $P(X = 4) = f(4) = F(4) - F(3) = 0.815 - 0.647 = 0.168$

Example 5.17:(Walpole et al. 2017a) During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Example 5.18:(Walpole et al. 2017a) Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

Example 3.8:[(Pishro-Nik 2014)] The number of emails that I get in a weekday can be modeled by a Poisson distribution with an average of 0.2 emails per minute.

1. What is the probability that I get no emails in an interval of length 5 minutes?
2. What is the probability that I get more than 3 emails in an interval of length 10 minutes?

Solution

Let, X=number of emails that I get in a given interval.

Given, $\lambda = 0.2 \;\; min^{-1}$.

X will follow $Pois(\lambda t)$

1. In this case $\mu = \lambda t = 0.2 * 5 = 1$. **So**, $P(X = 0) = f(0) = e^{-\mu} = e^{-1} = 0.3679$.

2. In this case $\mu = \lambda t = 0.2 * 10 = 2$. **So**,$P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = 1 - 0.857 = 0.143$. [From **Table A3**]

Approximation of Binomial Distribution to Poisson

When,

- $p \rightarrow 0$ (*Success rate is very low*);
- $n \rightarrow \infty$ (*Number of trials is very large*);

Then **Binomial distribution** can be *approximated* by **Poisson distribution**.

- Mathematically, $Binom(x; n, p) \approx Pois(\lambda)$; where $\lambda = np$.

N.B: In practical situation if $n \geq 30$ and $p \leq 0.05$;hence $q \geq 0.95$,then the approximation is close enough to use the Poisson distribution for binomial problems(Baron 2019).

Example 5.20:(Walpole et al. 2017a) In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, 1 in every 1000 of these items produced has one or more bubbles. What is the probability that a random sample of 8000 will yield fewer than 7 items possessing bubbles?

Solution:

Let,$X = number \ \ of \ \ glasses \ \ possesing \ \ bubbles$

Given, $Pr(buuble \ \ occurs) = p = 1/1000 = 0.001$ which is less than 0.05, and $n = 8000$ which is greater than 30. So, the PMF of X can be approximated by Poisson distribution with

$$\lambda = np = 8000 * 0.001 = 8$$

that is $X \sim Pois(\lambda = 8)$

According to question,

$P(X < 7) = f(0) + f(1) + ... + f(6) = F(6) = 0.313$ (*Ans.*)

[By using **Table A3**]

Exercise 5.87:(Walpole et al. 2017a) Imperfections in computer circuit boards and computer chips lend themselves to statistical treatment. For a particular type of board, the probability of a diode failure is 0.03 and the board contains 200 diodes.

(a) What is the mean number of failures among the diodes? (***Ans:*** $\mu = np = 200 * 0.03 = 6$)

(b) What is the variance?(***Ans:*** $\sigma^2 = np(1-p) = 200 * 0.03 * (1 - 0.03) = 5.82$)

(c) The board will work if there are no defective diodes. What is the probability that a board will work? ***Ans (c):*** The board will work if there are no defective diodes.

So, P(The board will work)=$P(X = 0) = f(0) = e^{-\mu} = e^{-6} = 0.0025$.

5 Continuous r.v and probability density function

5.1 Definition

A continuous r.v X must have a probability density function (PDF) $f(x)$ such that

1) $f(x) \geq 0$ [**Non-negativity**]

2) $\int_{x\in\mathbb{R}} f(x)dx = 1$ [**Total AREA under the curve** $f(x)$ always 1]

3) $P(a < X < b) = \int_a^b f(x)dx$

5.2 Illustration with an example

Given $f(x) = \frac{1}{2}x \;\; ; 0 \leq x \leq 2$

a) Show/plot the graph of $f(x)$.

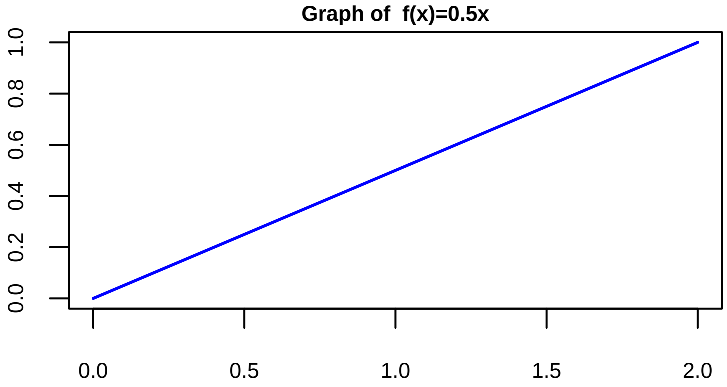
b) Is $f(x)$ a PDF?

c) Find $P(X < 1.0)$.

d) Find $P(X = 1.0)$

Solution:

(a)

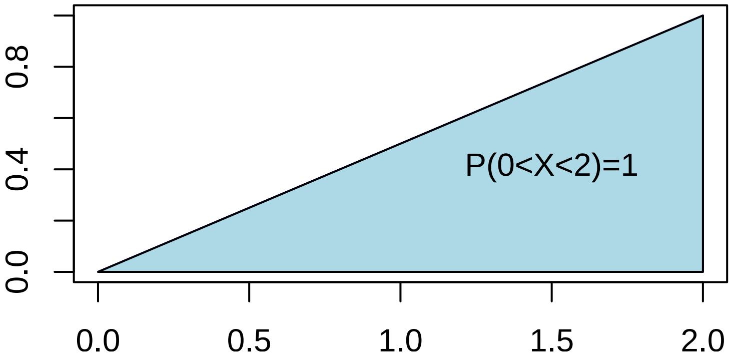


b) Here, $f(x) \geq 0$ for all values of x in the interval $0 \leq x \leq 2$.

Now, **total area under curve** $f(x)$ from $x = 0$ to $x = 2$ is

$\int_0^2 f(x)dx$

$= AREA\ \ of\ \ the\ \ SHADED\ \ Triangle$



$$= \frac{1}{2} \times base \times height$$

$$= \frac{1}{2} \times 2 \times 1 = 1$$

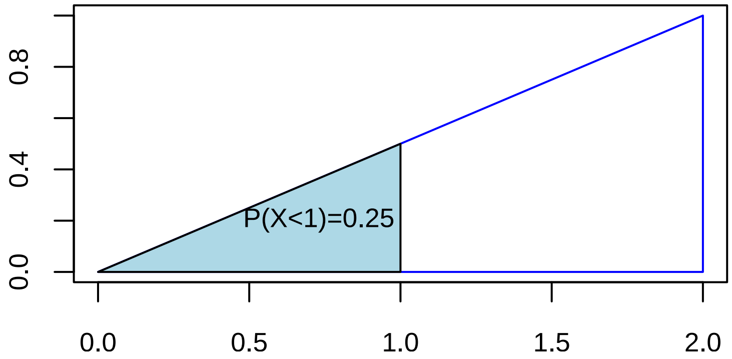
So, total area under curve $f(x)$ is 1 that is $\int_0^2 f(x)dx = 1$.

Hence, $f(x)$ is a PDF.

c) Here,

$$P(X < 1) = Area \ \ under \ \ the \ \ curve \ \ from \ \ x = 0 \ \ to \ \ x = 1$$

$$= Area \ \ of \ \ the \ \ SHADED \ \ Triangle$$



$$= \frac{1}{2} \times 1 \times f(1) = \frac{1}{2} \times 1 \times 0.5 = 0.25$$

Therefore $P(X < 1) = 0.25$

d) $P(X = 1.0) = 0$ [Because there is no area at $x = 1.0$]

ℹ Note

We always remember that **Probability in an interval of X is actually the *AREA* under the pdf $f(x)$.**

Problem 6.2.1 A random variable has the following density function.

$$f(x) = 1 - 0.5x \ \ ; \ \ 0 < x < 2$$

a) Graph the density function.

b) Verify that $f(x)$ is a density function.

c) Fond $P(X > 1)$.

d) Find $P(X < 0.5)$.

e) Find $P(X = 1.5)$.

N.B: $P(X = a) = 0$ as well as $P(X = b) = 0$. So, $P(X \leq a)$ is same as $P(X < a)$.

5.3 CDF of continuous r.v X

By definition, CDF,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Therefore,

- $f(x) = \frac{d}{dx}F(x)$.
- $P(a < X < b) = F(b) - F(a)$.

5.4 Expectation and variance of continuous r.v

If X is a continuous r.v with PDF $f(x)$ then

Expected value of X is

$$\mu = E(X) = \int_{x\in\mathbb{R}} x \cdot f(x)dx$$

Variance of X is

$$Var(X) = E(X^2) - \mu^2 = \int_{x\in\mathbb{R}} x^2 \cdot f(x)dx - \mu^2$$

Example 3.11(Walpole et al. 2017a) Suppose that the error in the reaction temperature, in ^0C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \frac{x^2}{3}; -1 < x < 2.$$

(a) Verify that $f(x)$ is a density function.
(b) Find $P(0 < X \leq 1)$.

Example 3.12(Walpole et al. 2017a) Find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

H.W: Find $E(X)$ and $Var(X)$ where,$f(x) = \frac{x^2}{3}; -1 < x < 2$.

Exercise 3.29(Walpole et al. 2017a) An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = 3x^{-4}; x > 1$$

(a) Verify that this is a valid density function.
(b) Evaluate $F(x)$.
(c) What is the probability that a random particle from the manufactured fuel exceeds 4 micrometers?

Exercise 3.69(Walpole et al. 2017a) The life span in hours of an electrical component is a random variable with cumulative distribution function

$$F(x) = 1 - e^{-\frac{x}{50}}; x > 0$$

(a) Determine its probability density function (PDF).

(b) Determine the probability that the life span of such a component will exceed 70 hours.

Exercise 3.36 (Walpole et al. 2017a) On a laboratory assignment, if the equipment is working, the density function of the observed outcome, X, is

$$f(x) = 2(1 - x) \quad ; 0 < x < 1$$

a. Calculate $P(X \leq 1)$

b. What is the probability that X will exceed 0.5?

c. Given that $X \geq 0.5$, what is the probability that X will be less than 0.75?

Hints: We can solve this exercise by either using $F(x)$ or simply drawing function $f(x)$.

6 Some special continuous random variables

6.1 Uniform distribution/r.v

A continuous r.v X is said to be uniform r.v ranges between a to b if it has the following PDF

$$f(x) = \frac{1}{b-a} \quad ; \quad a < x < b \tag{6.1}$$

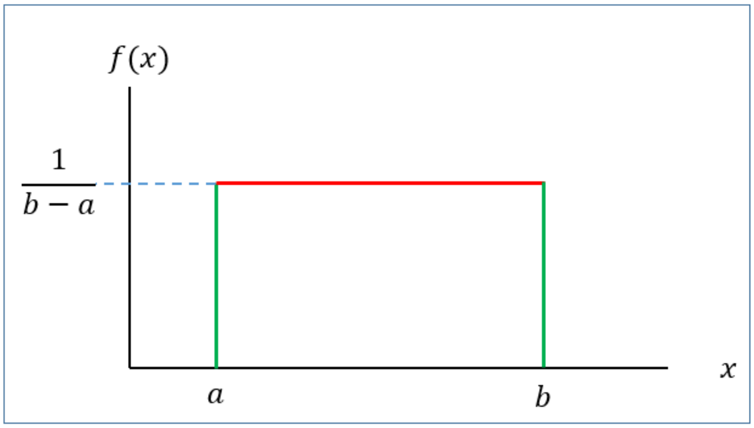


Figure 6.1: Graph of f(x)

with

Mean: $\mu = E(X) = \frac{a+b}{2}$

Variance: $\sigma^2 = \frac{(b-a)^2}{12}$

We write, $X \sim U(a, b)$

6.1.1 Finding probability for uniform r.v

If $X \sim U(a, b)$ then the $P(x_1 < X < x_2)$ is actually the **area of the shaded rectangle.**

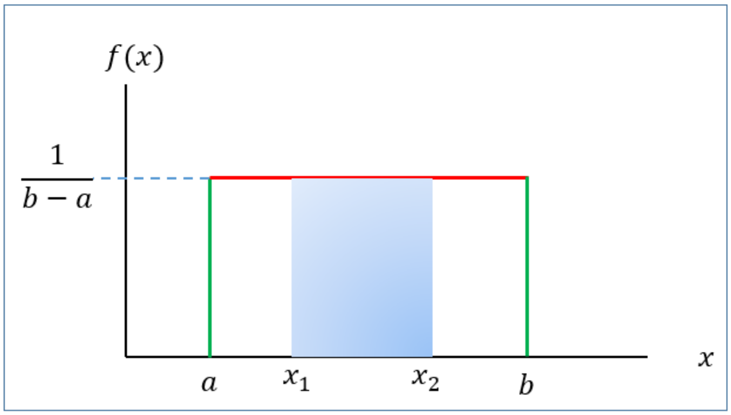


Figure 6.2: Computing area for an interval of Uniform distribution

That is,

$$P(x_1 < X < x_2) = Base \times Height = (x_2 - x_1) \times \frac{1}{b-a}$$

Problem 1 The phase angle, Θ, of the signal at the input to a modem is uniformly distributed between 0 and 2π radians.

a) What are the PDF, expected value, and variance of Θ?

b) Find the probability that phase angle exceeds $\frac{3\pi}{2}$?

Problem 2 In a radio communications system, the phase difference X between the transmitter and receiver is modeled as having a uniform density in $[—\pi, +\pi]$. Find $P(X < 0)$ and $P(X < \pi/2)$.

Problem 3(Navidi 2011, 278) Resistors are labeled 100 Ω. In fact, the actual resistances are uniformly distributed on the interval $(95, 103)$.

a. Find the mean resistance.

b. Find the standard deviation of the resistances.

c. Find the probability that the resistance is between 98 and 102 Ω.

d. Suppose that resistances of different resistors are independent. What is the probability that three out of six resistors have resistances greater than 100 Ω?

6.1.2 Generating Uniform random number in R

```r
par(mar=c(4,4,1,1)) # Adjust graph margin
set.seed(911)
u=runif(1000,10,30) # Generating 1000 random numbers from U(10,30)
hist(u,main = " ",col="steelblue",ylim = c(0,120)) # Frequency histogram of U
```

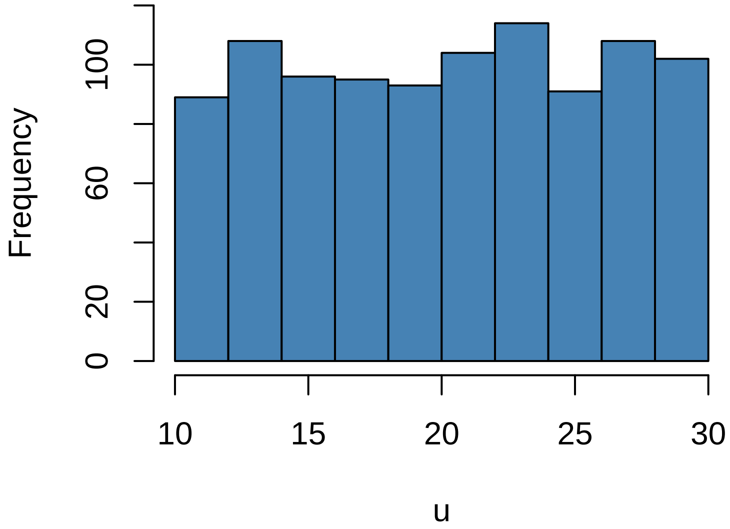


Figure 6.3: Histogram of Uniform random numbers from U(10,30), $n = 1000$

6.2 Normal or Gaussian r.v

The most important probability distribution for describing a continuous random variable is the **normal probability distribution**. The normal distribution has been used in a wide variety of practical applications in which the random variables are heights and weights of people, test scores, scientific measurements, amounts of rainfall, and other similar values.

6.2.1 Definition

A continuous r.v X is said to be normal r.v if it has the following **PDF:**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; -\infty < x < \infty \tag{6.2}$$

The graph of $f(x)$ is called **normal curve** (Figure 6.4).

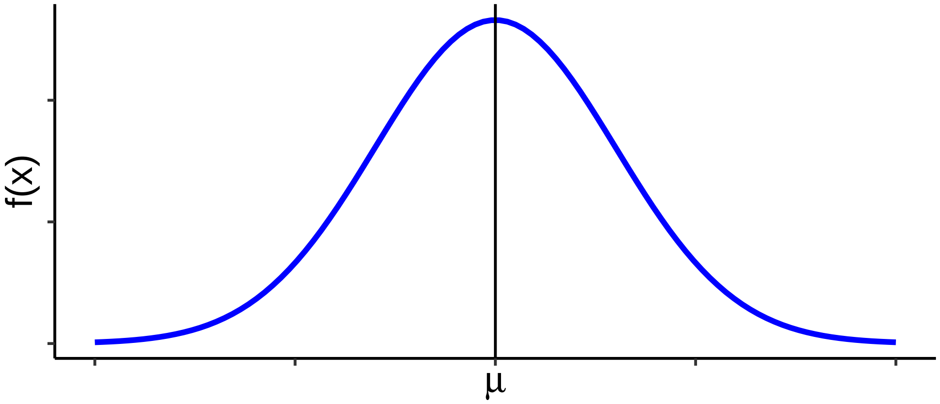


Figure 6.4: Normal Curve

Mean: $E(X) = \mu$

Variance: $Var(X) = \sigma^2$

We write: $X \sim N(\mu, \sigma^2)$

Properties of normal distribution

- The **total area** under the normal curve $f(x)$ is 1 that is

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- Normal distribution is symmetric about mean, μ
- Mean, median and mode is identical in normal distribution that is $Mean = Median = Mode = \mu$
- Almost 99% observations of X lie within **3 standard deviation of mean** that is

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.99$$

- Almost 95% observations of X lie within **2 standard deviation of mean** that is

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$$

- Almost 68% observations of X lie within **1 standard deviation of mean** that is

$$P(\mu - \sigma < X < \mu + \sigma) \approx 0.68$$

6.2.2 Standard normal r.v

Suppose $X \sim N(\mu, \sigma^2)$. Then the variable $Z = \frac{X-\mu}{\sigma}$ is said to be **standard normal variable** with **PDF**

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2} \quad ; -\infty < z < \infty \tag{6.3}$$

Mean: $E(Z) = 0$

Variance: $Var(Z) = 1$

We write: $Z \sim N(0, 1)$

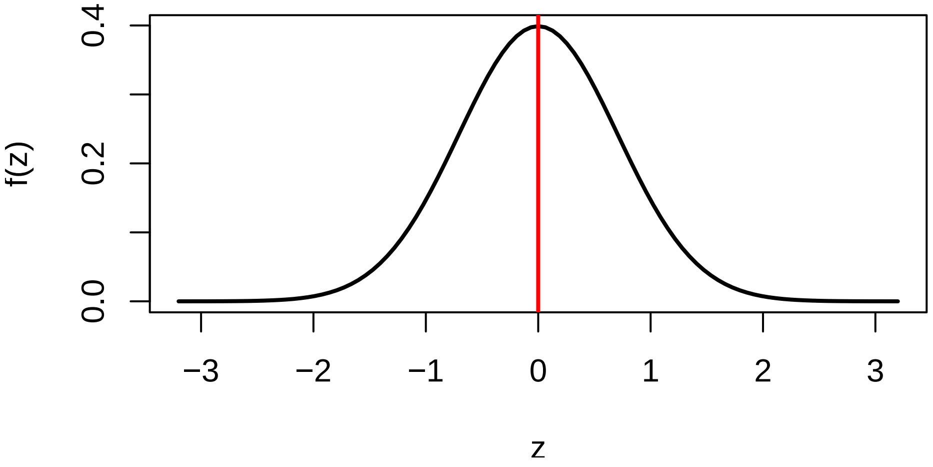


Figure 6.5: Standard normal curve

6.2.3 Computing probability(area) under standard normal curve

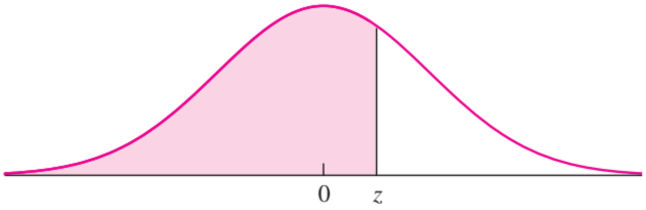
To compute area (probability) under the standard normal curve for a given interval of z we use **standard Normal Distribution table** which provides cumulative probabilities.

RULE-I: Suppose we want to find $P(Z < 1.25)$.

From **TABLE A.2** in Navidi (2011) we have

$$P(Z < 1.25) = 0.8944$$

TABLE A.2 Cumulative normal distribution (continued)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

The probability $P(Z < 1.25)$ is shown in Figure 6.6.

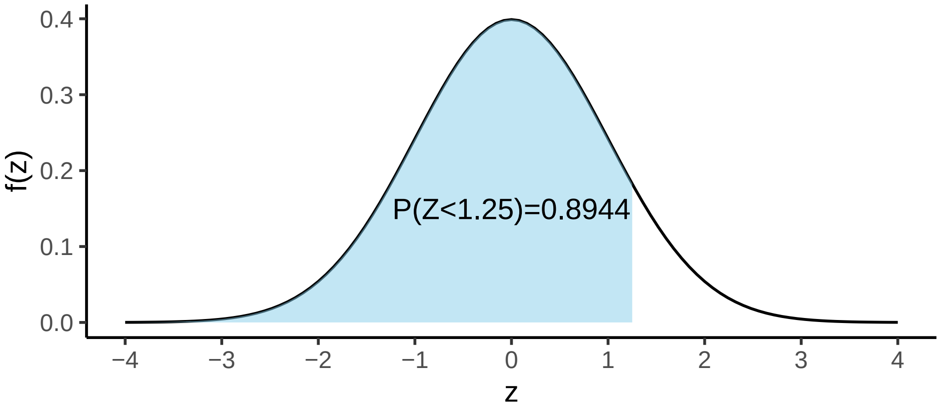


Figure 6.6: Area under standard normal curve for Z<1.25

RULE-II: Now we find $P(Z > 1.36)$

So, due to symmetry we can write $P(Z > 1.36) = P(Z < -1.36) = 0.0869$

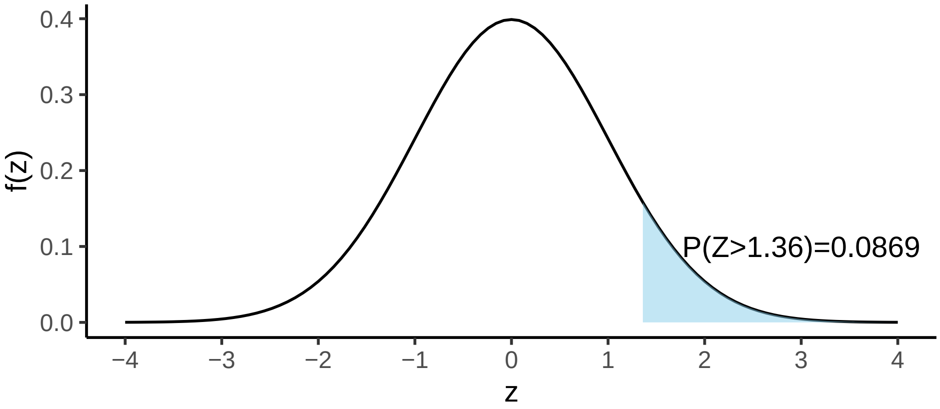


Figure 6.7: Area under standard normal curve for Z>1.36

RULE-III: Let us evaluate $P(-1.96 < Z < 2.58)$.

We can write

$$= P(-1.96 < Z < 2.58)$$

$$= P(Z < 2.58) - P(Z < -1.96)$$

$$= 0.9951 - 0.0250 = 0.9701$$

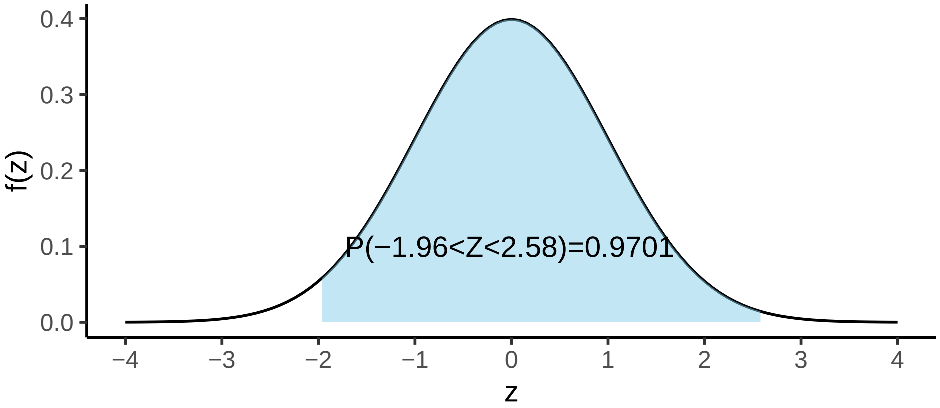


Figure 6.8: Area under standard normal curve for -1.96<Z<2.58

6.2.4 Finding quantiles (percentiles, quartiles, deciles etc.) of Z

What is the 90^{th} percentile of Z? To answer this question, let k is the 90^{th} percentile of Z. So we can write

$$P(Z < k) = 0.90 \qquad \cdots (1)$$

From **TABLE A.2** in Navidi (2011) we have

$$P(Z < 1.28) = 0.90 \qquad \cdots (2)$$

Comparing eq.(1) with eq.(2) we have $k = 1.28$. So the 90^{th} percentile of Z is 1.28.

Problem 1 Find c such that $P(Z > c) = 0.05$.

Problem 2 Find c such that $P(-c < Z < c) = 0.95$.

6.2.5 Computing probability(area) under normal curve:

Suppose $X \sim N(30, 5^2)$. Then find the following:

a) $P(X < 22)$

b) $P(X > 44)$

c) $P(20 < X < 35)$

d) If $P(X < x) = 0.25$ then find the value of x.

Solution:

Here, $\mu = 30$ and $\sigma = 5$

---

a) $P(X < 22) = P(\frac{X-\mu}{\sigma} < \frac{22-30}{5}) = P(Z < -1.60) = 0.0548.$

---

b) $P(X > 44) = P(\frac{X-\mu}{\sigma} > \frac{44-30}{5})$

$= P(Z > 2.80) = P(Z < -2.80) = 0.0026$

---

c) $P(20 < X < 35) = P(\frac{20-30}{5} < \frac{X-\mu}{\sigma} < \frac{35-30}{5})$

$= P(-2 < Z < 1) = P(Z < 1) - P(Z < -2)$

$= 0.8413 - 0.0228 = 0.8185$

---

d) To find the value of x we proceed this way.

$$P(X < x) = 0.25$$

$$\implies P(\frac{X-\mu}{\sigma} < \frac{x-30}{5}) = 0.25$$

$$\implies P(Z < \frac{x-30}{5}) = 0.25 \qquad \cdots (1)$$

From TABLE (Appendix B) we have

$$P(Z < -0.67) = 0.25 \qquad \cdots (2)$$

Comparing (1) with (2) we can write

$$\frac{x-30}{5} = -0.67$$
$$\implies x = 30 + (-0.67) \times 5$$

$$\therefore x = 26.65$$

Note

If $P(X < x) = p$ and
$P(Z < z) = p$ then

$$x = \mu + z\sigma$$

6.2.6 Applications of the Normal Distribution (Walpole et al. 2017a)

Example 6.7: A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

Example 6.8 : An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

Example 6.9 : In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped?

Example 6.10 : Gauges are used to reject all components for which a certain dimension is not within the specification $1.50 \pm d$. It is known that this measurement is normally distributed with a mean of 1.50 and a standard deviation of 0.2. Determine the value d such that the specifications "cover" 95% of the measurements.

Solution:

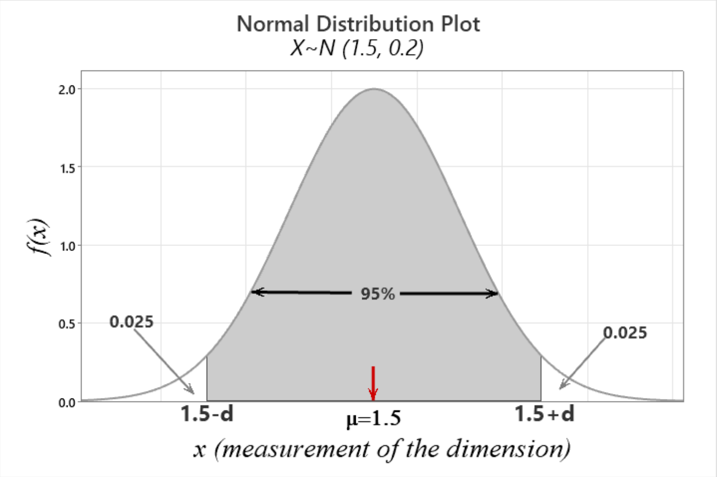
Let, $X = measurement \;\; of \;\; certain \;\; dimension$

Given, $X \sim N(1.5, 0.2)$

According to question,

$P(1.5 - d < X < 1.5 + d) = 0.95$

So, $P(X < 1.5 - d) + P(X > 1.5 + d) = 0.05$



So, $P(X < 1.5 - d) = 0.025 \quad \cdots (1)$

$\implies P(\frac{X-\mu}{\sigma} < \frac{1.5-d-1.5}{0.2}) = 0.025$

$\implies P(Z < \frac{-d}{0.2}) = 0.025 \quad \cdots (1)$

From TABLE A.2 we have

$P(Z < -1.96) = 0.025 \quad \cdots (2)$

Comparing eq.(1) with eq. (2) we have

$\frac{-d}{0.2} = -1.96$

$\implies -d = -0.392$

$\therefore d = 0.392$

Alternative:

$P(X < 1.5 - d) = 0.025 \quad \cdots (1)$

But, $P(Z < -1.96) = 0.025 \quad \cdots (2)$

Hence, $1.5 - d = \mu + z\sigma$; where $z = -1.96$

$\implies 1.5 - d = 1.5 - 1.96 \times 0.2$

$\therefore d = 0.392$

Exercise 6.11 : A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

(a) what fraction of the cups will contain more than 224 milliliters?
(b) what is the probability that a cup contains between 191 and 209 milliliters?
(c) how many cups will probably overflow if 230-milliliter cups are used for the next 1000 drinks?
(d) below what value do we get the smallest 25% of the drinks?

Exercise 6.14 The finished inside diameter of a piston ring is normally distributed with a mean of 10 centimeters and a standard deviation of 0.03 centimeter.

(a) What proportion of rings will have inside diameters exceeding 10.075 centimeters?
(b) What is the probability that a piston ring will have an inside diameter between 9.97 and 10.03 centimeters?
(c) Below what value of inside diameter will 15% of the piston rings fall?

Exercise 6.17 : The average life of a certain type of small motor is 10 years with a standard deviation of 2 years. The manufacturer replaces free all motors that fail while under guarantee. If she is willing to replace only 3% of the motors that fail, how long a guarantee should be offered? Assume that the lifetime of a motor follows a normal distribution.

6.2.7 Normal approximation to Binomial distribution

Binom tends to normal when $n \to \infty$ and $p \to 0.5$

From Figure 6.9, Figure 6.10 and Figure 6.11 we observe that as n getting large with $p = 0.2$ (which is not extremely close to 0) the binomial distribution can be approximated to normal distribution.

So, when n become large and p is not extremely small (close to 0) or extremely large (close to 1) the normal approximation is most useful in calculating binomial sums.

! Normal Approximation to the Binomial Distribution

- Let X be a binomial random variable with parameters n and p. For large n, X has approximately a normal distribution with $\mu = np$ and $\sigma^2 = np(1-p)$.

- The approximation will be good if np and $n(1-p)$ are greater than or equal to 5 (Walpole et al. 2017a).

Continuity correction

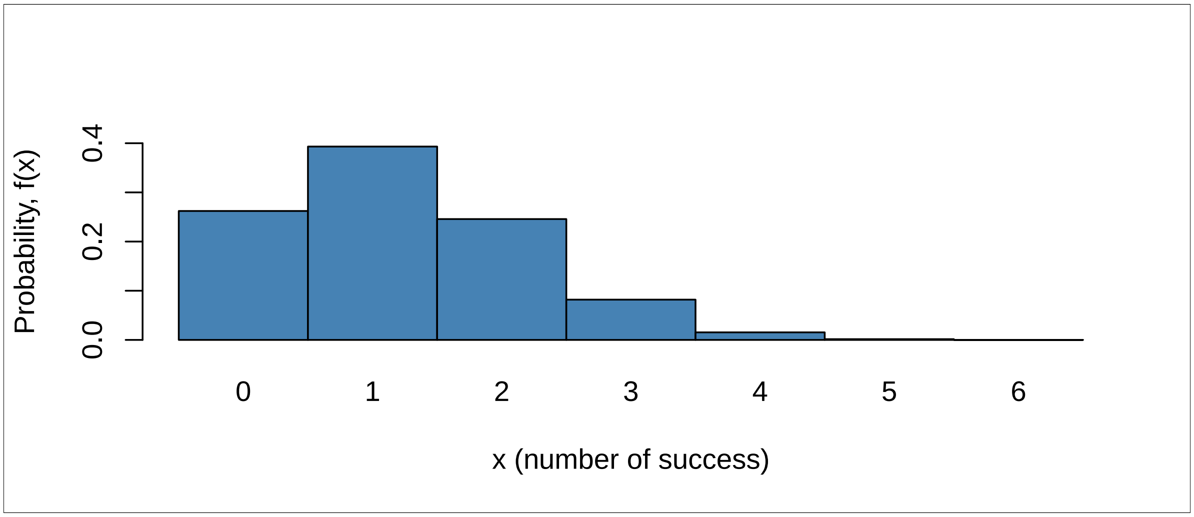


Figure 6.9: Histogram of $Bin(x; n = 6, p = 0.2)$

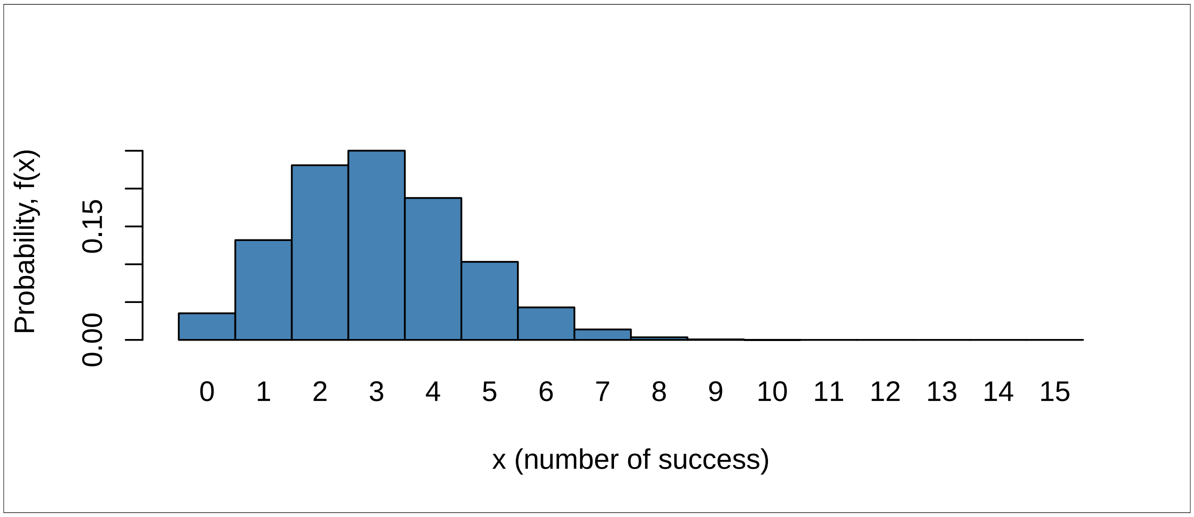


Figure 6.10: Histogram of $Bin(x; n = 15, p = 0.2)$

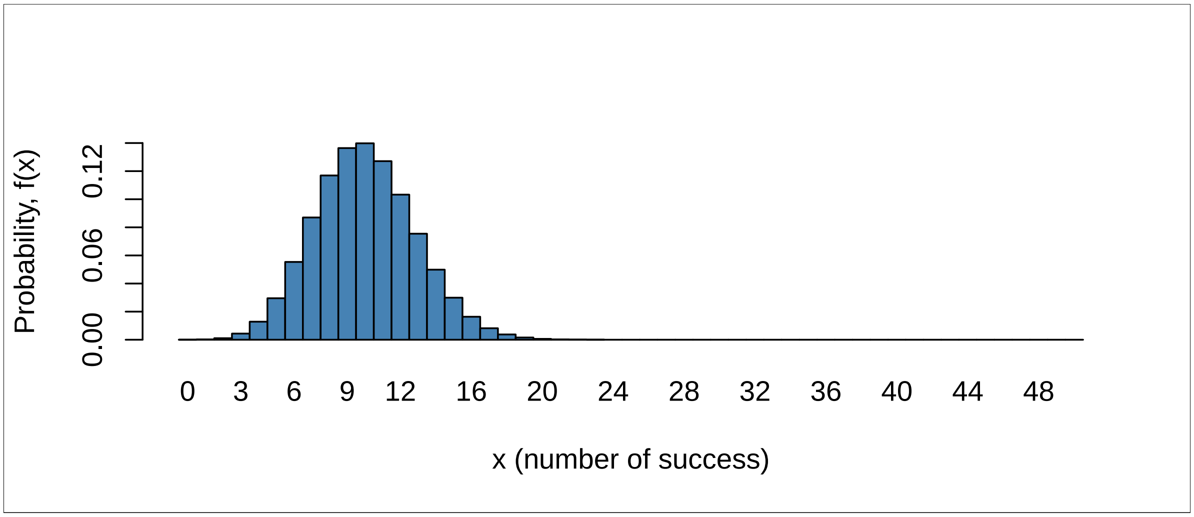


Figure 6.11: Histogram of $Bin(x; n = 50, p = 0.2)$

This correction is needed when we approximate a discrete distribution (Binomial in this case) by a continuous distribution (Normal). Recall that the probability $P(X = x)$ may be positive if X is discrete, whereas it is always 0 for continuous X. This is resolved by introducing a continuity correction. Expand the interval by 0.5 units in each direction, then use the Normal approximation (Baron 2019).

Table 6.1: Continuity correction for binomial probabilities

Binomial Probability	With continuity correction
$P(X = x)$	$P(x - 0.5 < X < x + 0.5)$
$P(X \leq x)$	$P(X \leq x + 0.5)$

Example 6.15 (Walpole et al. 2017a, 191): The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that fewer than 30 survive?

6.3 Exponential r.v

In many situations, such as when modeling waiting times, inter-arrival times, the lifespan of hardware, breakdown times, and the intervals between phone calls, the exponential distribution is utilized. The time (suppose T) between rare events in Poisson process with arrival rate λ (*number of arrival per unit time*) can be treated as exponential r.v.

The exponential r.v T has the following PDF

$$f(t) = \lambda e^{-\lambda t} \quad ; \quad t > 0 \tag{6.4}$$

- **CDF:** $F(t) = P(T \leq t) = P(T < t) = 1 - e^{-\lambda t}; t > 0$

  Hence, $P(T > t) = 1 - P(T \leq t) = 1 - F(t) = e^{-\lambda t}$
- **Mean:** $E(T) = \frac{1}{\lambda}$
- **Variance:** $Var(T) = \frac{1}{\lambda^2}$
- **We write,** $T \sim Exp(\lambda)$

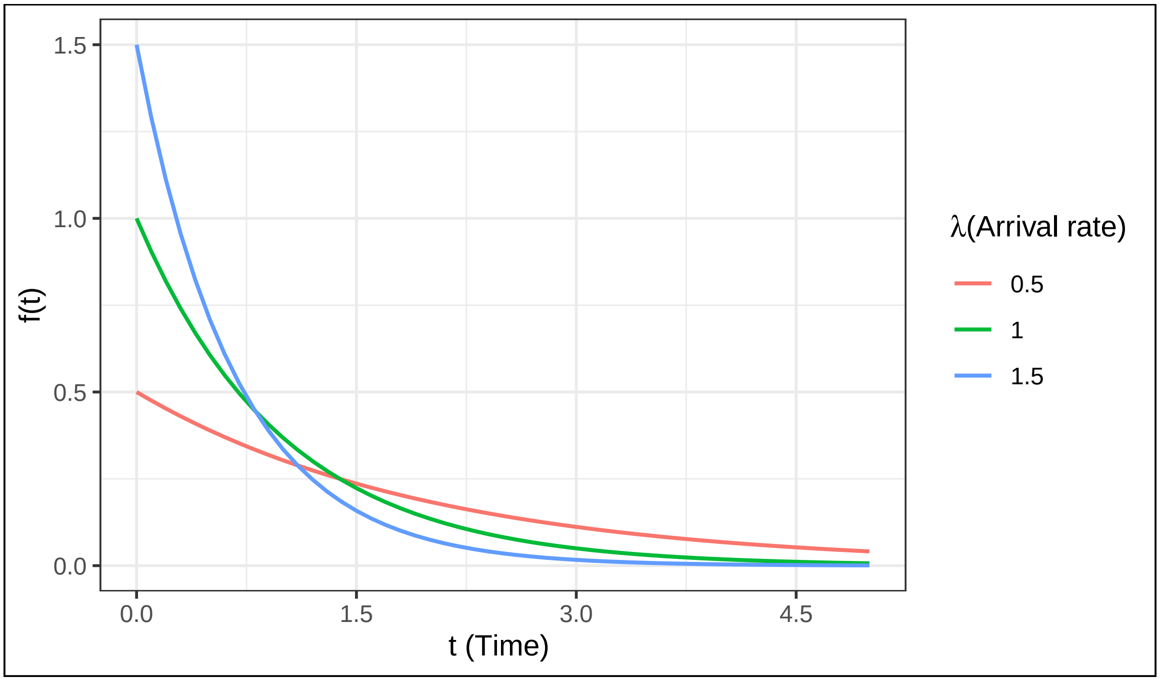


Figure 6.12: Exponential probability density functions for selected values of λ

The quantity λ is a parameter of Exponential distribution, and its meaning is clear from $E(T) = \frac{1}{\lambda}$. If T is time, measured in minutes, then λ is a frequency, measured in min^{-1}. For example, if arrivals occur every half a minute, on the average, then $E(T) = 0.5$min and $\lambda = 2$, saying that they occur with a frequency (arrival rate) of 2 arrivals per minute. This λ has the same meaning as the parameter of Poisson distribution(Baron 2019).

Example 4.5(Baron 2019) Jobs are sent to a printer at an average rate of 3 jobs per hour.

(a) What is the expected time between jobs?

(b) What is the probability that the next job is sent within 5 minutes?

Solution: Given, number of jobs per hour, $\lambda = 3 \ \ hr^{-1}$ per hour. Let, T=time elapsed between jobs (hour).

So, $T \sim Exp(\lambda)$

(a) $E(T) = \frac{1}{\lambda}hr = \frac{1}{3}hr = 20 \ \ mins$;

(b) Here, $5 \ \ mins = \frac{5}{60}hr = \frac{1}{12}hr$ We know, $F(t) = 1 - e^{-\lambda t}; t > 0$

So, $P(T < 5 \ \ mins) = P(T < \frac{1}{12}) = F(\frac{1}{12}) = 1 - e^{-3*\frac{1}{12}} = 0.22$

Example 4.58 (Navidi 2011) A radioactive mass emits particles according to a Poisson process at a mean rate of 15 particles per minute. At some point, a clock is started. What is the probability that more than 5 seconds will elapse before the next emission? What is the mean waiting time until the next particle is emitted?

Solution

Let, $T = elapsed \ \ time \ \ before \ \ the \ \ next \ \ emission(in \ \ second)$

Given, $\lambda = 15min^{-1} = \frac{15}{60}s^{-1} = 0.25s^{-1}$ and

$T \sim Exp(\lambda)$;

$P(T \leq t) = F(t) = 1 - e^{-\lambda t}$

P(more than 5 seconds will elapse before the next emission)=$P(T > 5) = e^{-\lambda * 5} = e^{-0.25*5} = 0.2865$

Mean waiting time, $E(T) = \frac{1}{\lambda}s = \frac{1}{0.25}s = 4s$

6.3.1 Lack of Memory Property

If $T \sim Exp(\lambda)$, and t and s are positive numbers, then

$$P(T > t + s | T > s) = P(T > t)$$

We can also express the *Lack of memory property* in this way:

$$P(T < t + s | T > s) = P(T < t)$$

The probability that we must wait additional t units, given that we have already waited s units, is the same as the probability that we must wait t units from the start. The exponential distribution does not "*remember*" how long we have been waiting.

- In particular, if the lifetime of a component follows the exponential distribution, then the probability that a component that is s time units old will last an additional t time units is the same as the probability that a new component will last t time units.
- In other words, a component whose lifetime follows an exponential distribution does not show any effects of age or wear(Navidi 2011).
- But if the failure of the component is a result of gradual or slow wear (as in mechanical wear), then the exponential does not apply and either the **gamma** or the **Weibull distribution** (see (Walpole et al. 2017a), Section 6.10) may be more appropriate.

Example 4.59(Navidi 2011) The lifetime of a particular integrated circuit has an exponential distribution with mean 2 years. Find the probability that the circuit lasts longer than three years.

Example 4.60(Navidi 2011) Refer to Example 4.59. Assume the circuit is now four years old and is still functioning. Find the probability that it functions for more than three additional years (Hints: Apply Lack of Memory Property).

Exercises for Section 4.7(Navidi 2011)

1.Let $T \sim Exp(0.45)$. Find $\mu_T, \sigma_T^2, P(T > 3)$ and the median of T.

2.The time between requests to a web server is exponentially distributed with mean 0.5 seconds.

a. What is the value of the parameter λ?
b. What is the median time between requests?
c. What is the standard deviation?
d. What is the 80th percentile?
e. Find the probability that more than one second elapses between requests.
f. If there have been no requests for the past two seconds, what is the probability that there more than one additional second will elapse before the next request?

Solution

Let, $T = time \;\; between \;\; requests \;\; in \;\; second$

If, $T \sim Exp(\lambda)$; then it is given that

$E(T) = 0.5 \;\; second$

$\implies \frac{1}{\lambda} = 0.5 \;\; second$

a. $\therefore \lambda = 2s^{-1}$

b. If M is the median time between request then,

$P(T \leq M) = 0.5$

$\implies F(M) = 0.5$

$\implies 1 - e^{-\lambda * M} = 0.5$

$\implies e^{-\lambda * M} = 0.5$

$\implies -\lambda * M = ln(0.5)$

$\implies M = \frac{ln(0.5)}{-\lambda} = \frac{ln(0.5)}{-2} = 0.3466 \approx 0.35$

So, median time between request is $0.35s$

c. Standard deviation of T, $\sigma_T = \frac{1}{\lambda}s = 1/2 = 0.5s$

d. Let, P_{80} denotes 80th percentile.

So, solve the following equation for P_{80}

$P(T \leq P_{80}) = 0.80$

$\implies P(T > P_{80}) = 0.20$

$\implies e^{-\lambda \times P_{80}} = 0.20$

$\implies P_{80} = \frac{ln(0.20)}{-\lambda} = 0.805$

$\therefore P_{80} = 0.805s$

e. $P(T > 1) = e^{-\lambda * 1} = 0.1353$

f. If there have been no requests for the past two seconds, the probability that there more than ***one* additional second** will elapse before the next request is:

$P(T > 1 + 2/T > 2) = P(T > 1) = e^{-\lambda * 1} = 0.1353$ [by using *Lack of memory property*]

8.A radioactive mass emits particles according to a Poisson process at a mean rate of 2 per second. Let T be the waiting time, in seconds, between emissions.

a. What is the mean waiting time?
b. What is the median waiting time?
c. Find $P(T > 2)$. Hint:$P(T > 2) = e^{-\lambda * 2}$
d. Find $P(T < 0.1)$.Hint:$P(T < 0.1) = F(0.1)$
e. Find $P(0.3 < T < 1.5)$. Hint:$P(0.3 < T < 1.5) = F(1.5) - F(0.3)$
f. If 3 seconds have elapsed with no emission, what is the probability that there will be an emission within the next second? (Use Lack of Memory Property)

Solution of f.

Since T is exponentially distributed and hold lack of memory property, so it dose not matter what was happened in past 3 seconds. We have to just compute that *there will be an emission (event will occur) within the next second.* So,

$P(T < 1) = F(1) = 1 - e^{-\lambda * 1}$ (*do yourself*)

6.3.2 Generating exponential random numbers in R

```r
par(mar=c(4,4,1,1)) # Adjust graph margin
set.seed(911)
t=rexp(1000,rate = 2.5)
hist(t, main = "",col = "steelblue")
```

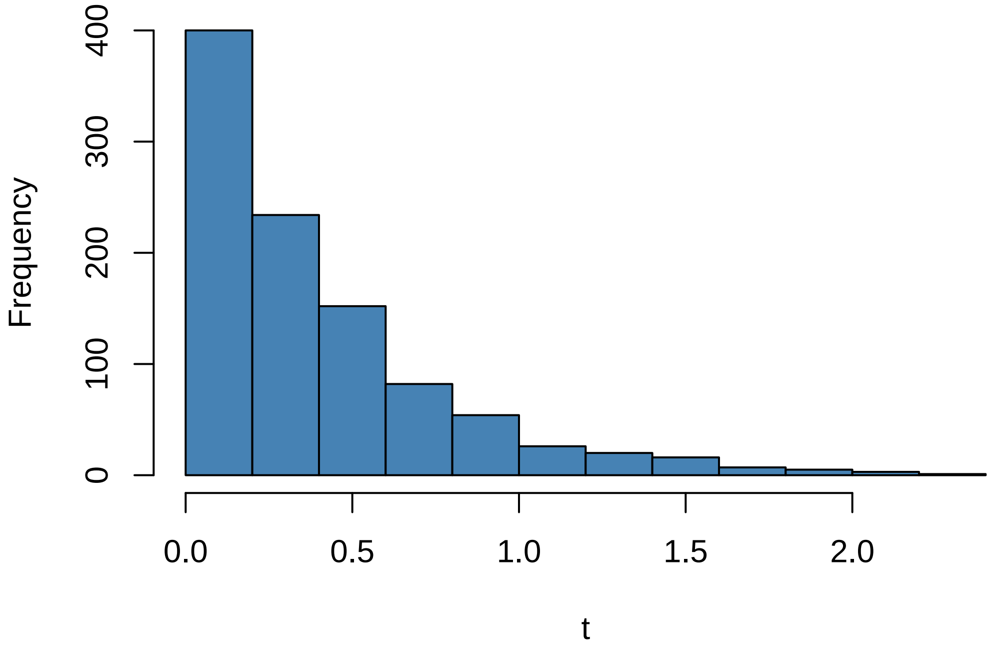


Figure 6.13: Histogram of exponential random variable from Exp($\lambda = 2.5$), $n = 1000$

7 Statistical inference

Statistical inference refers to the branch of statistics that focuses on making decisions and drawing conclusions about populations based on sample data. These methods use information collected from a sample to infer characteristics of the larger population.

Statistical inference may be divided into two major areas:

- **Parameter estimation** and
- **Hypothesis testing**

Parameter estimation involves estimation of population **parameter** from **sample data**.

Hypothesis testing involves test any **claim** about **population parameter** using sample data.

7.1 Parameter estimation

7.1.1 Point estimation

Point Estimation is a type of statistical inference where we use sample data to calculate a **single number** (a point) that serves as the best guess for an unknown population parameter.

Note

Random sample
Statistic
Point estimator
Point estimate

Table 7.1: Some common population parameters and their point estimators

Population parameter	Symbol	Point estimator
Population mean	μ	Sample mean, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

Population parameter	Symbol	Point estimator
Population standard deviation	σ	Sample standard deviation, $S=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{\sum X^2-n\cdot\bar{X}^2}{n-1}}$
Population proportion	p	Sample proportion, $\hat{P}=\frac{\#\ \ of\ \ outcomes\ \ of\ \ interest}{n}$

7.1.2 Properties of Point Estimators

Suppose

θ be the population parameter of interest

$\hat{\theta}$ be the sample statistic or point estimator of θ

A "good" estimator has some desirable properties.

Unbiasedness

A sample statistic $\hat{\theta}$ is said to be unbiased estimator of the population parameter θ if

$$E(\hat{\theta})=\theta$$

Efficiency

Consistency

7.1.3 Sampling Distributions and the Central Limit Theorem

The probability distribution of a **sample statistic** is called a **sampling distribution.**

For example, due to sampling variability the **sample mean** $\bar{X}$ has a sampling distribution.

Sampling distribution of $\bar{X}$

Suppose that a random sample of size n is taken from a normal population with mean μ and variance σ^2. Now each observation in this sample, say, $X_1, X_2, ..., X_n$, is a normally and independently distributed random variable with mean μ and variance σ^2. Then because linear functions of independent, normally distributed random variables are also normally distributed (Chapter ?), we conclude that the sample mean

$$\bar{X}=\frac{X_1+X_2+....+X_n}{n}$$

has a normal distribution with mean $\mu_{\bar{X}}=\mu$ and variance $\sigma^2_{\bar{X}}=\frac{\sigma^2}{n}$.

Symbolically, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

But what if we sampling from a non-normal population?

The sampling distribution of the sample mean will still be approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ if the sample size n is large. This is one of the most useful theorems in statistics, called the central limit theorem. The statement is as follows:

ℹ Central limit theorem

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2 and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

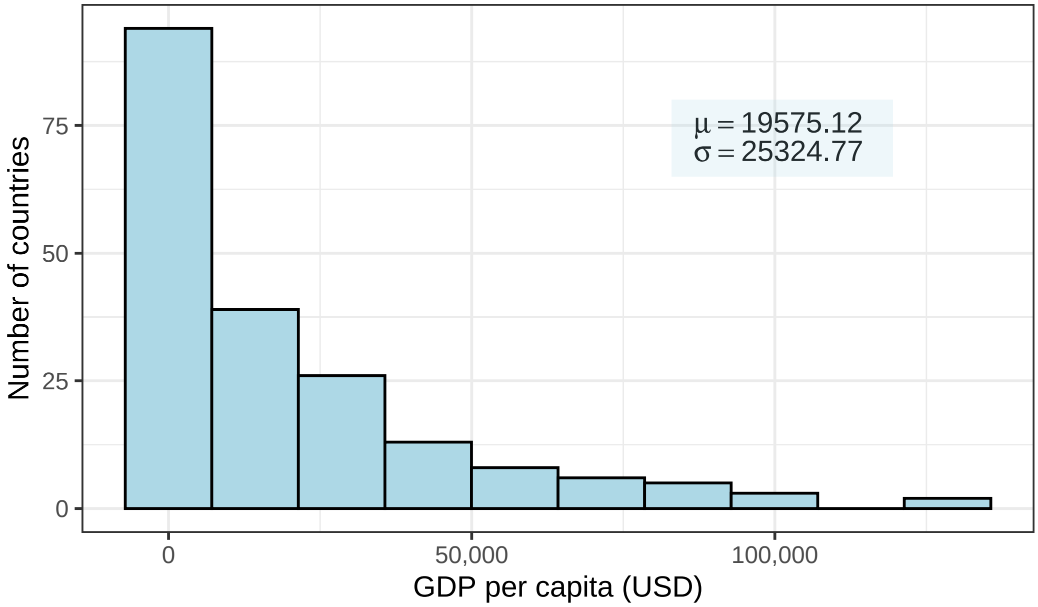
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

as $n \to \infty$, is the standard normal distribution that is $Z \sim N(0, 1)$

The definition of "sufficiently large" depends on the extent of non-normality of X . Some authors consider a sample will be sufficiently large if $n \geq 30$ (Walpole et al. 2017a).

! Central Limit Theorem through simulation

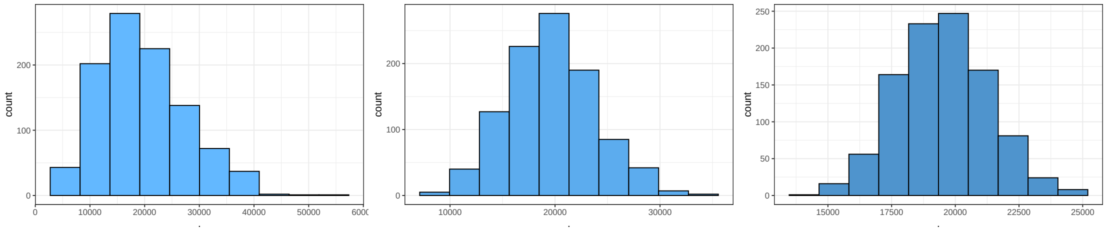
In this section we illustrates how sampling distributions of sample means approximate to normal or bell shaped distribution as we increase the sample size .
At first, we consider a population data regarding `gdp per capita (USD),2023` of 218 countries. We can see that the distribution of `gdp per capita` is highly skewed to the right (see Figure 7.1).



Source: World Bank, 2023

Figure 7.1: Frequency histogram of GDP percapita of N=218 countries

Now we draw 1000 random samples (without replacement) of different sample sizes and then plot the histogram of samples means.



(a) Sampling distribution of sample mean for sample size n=10

(b) Sampling distribution of sample mean for sample size n=30

(c) Sampling distribution of sample mean for sample size n=100

Figure 7.2: Demonstration of Central Limit Theorem through simulation

From Figure 7.2 we can see that as the sample size increases, the sampling distribution of **sample mean** tends to bell-shaped or normal though the population data was very skewed to the right. This simulation clearly demonstrate the fact of Central Limit Theorem (CLT).
For more interactive simulation of **CLT** please click here to visit the ShinyApp for Central

Limit Theorem Simulation.

Problem 7.1 An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. **Find** the probability that a random sample of n = 25 resistors will have an average resistance of fewer than 95 ohms.

Problem 7.2 Resistors are labeled 100 Ω. In fact, the actual resistances are uniformly distributed on the interval $(95, 103)$. Suppose 40 resistors are randomly selected. **Determine** the probability that the sample mean of 40 resistors will be less than 100 Ω?

Solution: Let X be the resistance in ohm. Given $X \sim U(95, 103)$.

Hence, $\mu = E(X) = \frac{a+b}{2} = \frac{95+103}{2} = 99$ and

$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(103-95)^2}{12} = 5.33.$

Since $n = 40$ is sufficiently large so according to **CLT** $\bar{X} \sim N(\mu, \sigma^2_{\bar{X}})$ approximately.

Here, $\sigma^2_{\bar{X}} = \sigma^2/n = 5.33/40 = 0.13325$ and $\sigma_{\bar{X}} = \sqrt{0}.13325 = 0.3650$

$$\therefore P(\bar{X} < 100) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} < \frac{100-99}{0.3650}\right) = P(Z < 2.74) = 0.9969$$

.

Problem 7.3 A synthetic fiber used in manufacturing carpet has tensile strength that is normally distributed with mean 75.5 psi and standard deviation 3.5 psi. **Find** the probability that a random sample of n = 6 fiber specimens will have sample mean tensile strength that exceeds 75.75 psi.

7.1.4 Methods of point estimation

Two popular methods of estimations are (a) the **method of moments** and (b)the **method of maximum likelihood** .

7.1.4.1 Method of moments

Method of moments uses the relationship between population and sample moments to estimate parameters of interest.

ℹ Moments

The k^{th} population moment is

$$\mu'_k = E(X^k), \quad k = 1, 2, ...$$

The corresponding k^{th} sample moment is

$$m'_k = \frac{\sum_{i=1}^{n} X_i}{n}, \quad k = 1, 2, ...$$

7.1.4.2 Moment Estimators

To estimate k parameters, equate the *first* k population moments to the first k sample moments and solving the resulting equations for the unknown parameters.

For instance, to estimate **two parameter** we can write

$$\mu'_1 = m'_1$$

and

$$\mu'_2 = m'_2$$

For details see (Montgomery and Runger 2014c, 256) and (Baron 2019, 245).

7.1.4.3 Method of maximum likelihood

The maximum likelihood technique is among the most effective ways to get a point estimator of a parameter. The renowned British statistician Sir R. A. Fisher created this method in the 1920s. As the name suggests, the value of the parameter that maximizes the **likelihood function** will serve as the estimator.

ℹ Maximum Likelihood Estimator

Suppose that X is a random variable with probability distribution $f(x;\theta)$ where θ is a single unknown parameter. Let $x_1, x_2, ..., x_n$ be the observed values in a random sample of size n. Then the likelihood function of the sample is

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta)....f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

The **maximum likelihood estimator (MLE)** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

For details see (Montgomery and Runger 2014c, 258) and (Baron 2019, 248).

7.2 Interval estimation

Instead of estimating a population parameter by a single value (point estimator) it is more reasonable to estimate with an **interval** with some confidence (probability) that our **parameter** value will be in the **interval.**

ℹ Interval Estimator

An **interval estimator** is a rule for determining (based on sample information) an interval that is likely to include the parameter. The general form of an interval estimate is as follows:

$$Point \quad estimate \pm margin \quad of \quad error$$

Due to sampling variability, **interval estimator** is also random.

Developing (1-α)% CI for μ

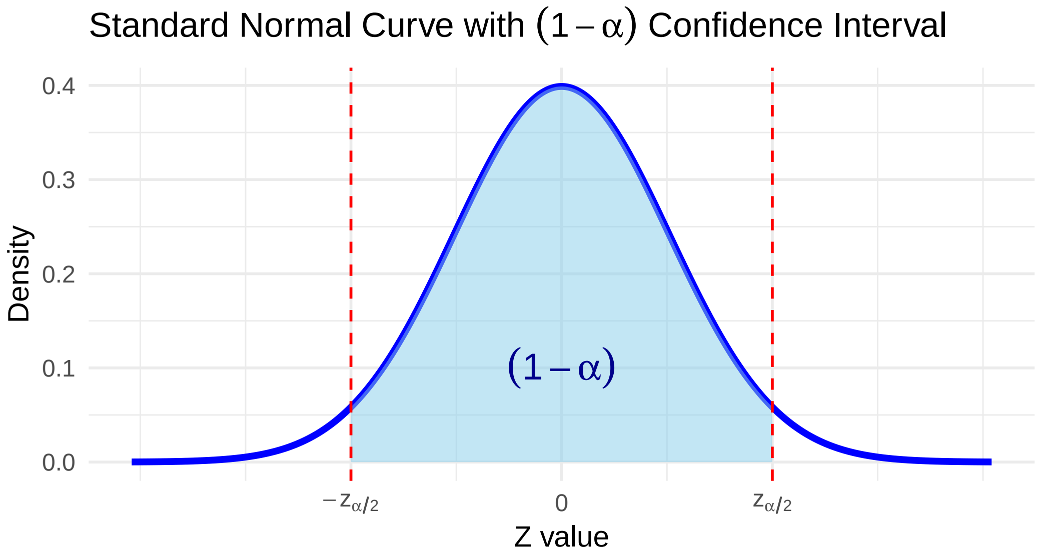


Figure 7.3: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

7.2.1 Interval estimate of a population mean: σ known

The $(1-\alpha)100\%$ confidence interval for μ is :

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{7.1}$$

Or,

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

We can express this confidence interval in a probabilistic way:

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

NOTE:

1) Here, $z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal distribution that is $P(Z > z_{\alpha/2}) = \alpha/2$.

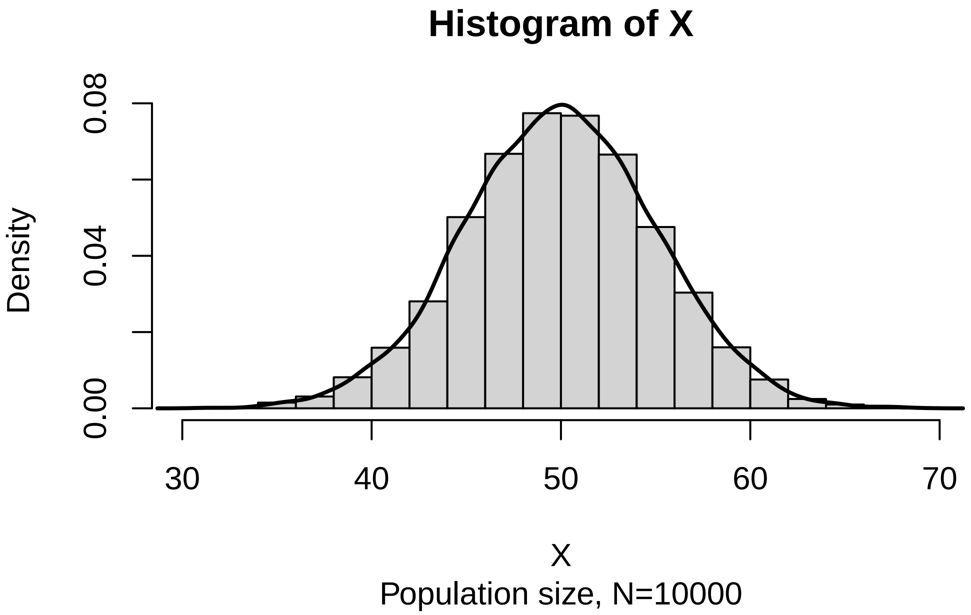
2) $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ is often called **margin of error (ME)**.

7.2.2 Interpretation of confidence interval

The probabilistic equation of confidence interval says that, if we repeatedly construct confidence intervals in this manner, we will expect $(1-\alpha)100\%$ of them contain μ.

7.2.3 Understanding confidence interval through Simulation

Suppose $X \sim N(50, 5^2)$. Now consider a population data of size $N = 10000$ and the histogram of X is:



Now we draw a random sample of size $n = 50$ from this population and construct a 95% confidence interval (CI) for μ. The CI may or may not include the $\mu = 50$!!!

```
Sample data : 52.60842 55.16664 59.23435 44.2092 50.94234 43.34063 44.65922 53.81687 47.60075
```

```
Sample mean: 49.3
```

```
95%  CI:
 [Lower ,Upper]
 [ 47.91 , 50.68 ]
```

Luckily our 95% CI contains the true population mean $\mu = 50$.

Lets simulate 100 samples each of size $n = 50$ and construct all 95% CIs.

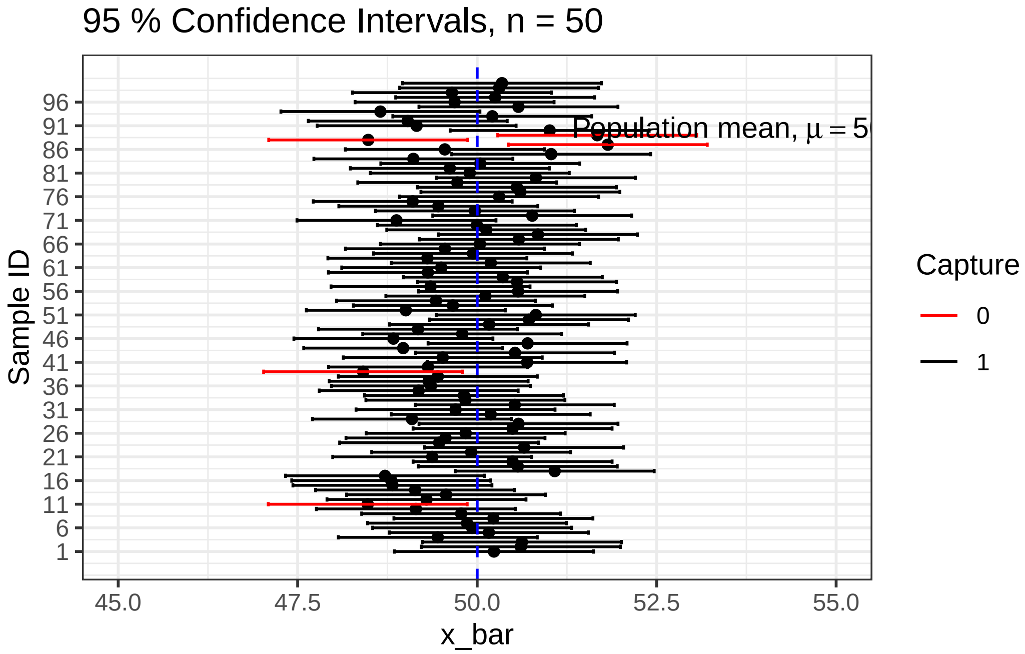


Figure 7.4: Simulation of 95% confidence intervals for μ

We can see that out of 100 CIs , 95 of them contain true population mean $\mu = 50$ and the rest 5 do not.

Table 7.2: Four Commonly Used Confidence Levels and $z_{\alpha/2}$

$1 - \alpha$	α	$z_{\alpha/2}$
0.90	0.10	1.645
0.95	0.05	1.96
0.98	0.02	2.33
0.99	0.01	2.575

7.2.4 Interval estimate of a population mean: σ unknown

The $(1-\alpha)100\%$ confidence interval for μ is :

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} \tag{7.2}$$

Or,

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

We can express this confidence interval in a probabilistic way:

$$P\left(\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

Here, $t_{\alpha/2}$ is the t value providing an area of $\alpha/2$ in the upper tail of the t distribution with $(n-1)$ degrees of freedom that is $P(T > t_{\alpha/2,n-1}) = \alpha/2$.

t-Distribution

Let $Z \sim N(0,1)$ and $V \sim \chi^2_\nu$. If Z and V are independent then the random variable

$$T = \frac{Z}{\sqrt{V/\nu}}$$

said to have a *Student-t distribution with ν degrees of freedom.* The PDF of T is

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\ \ \Gamma(\nu/2)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} ; -\infty < t < \infty.$$

Properties:

1) **Symmetry:** t-distribution is symmetric about mean (zero). So if $P(T > t_\nu) = \alpha$ then $P(T < -t_\nu) = \alpha$.

2) **Convergence to Normal:** As $n \to \infty$ then the distribution of T_ν approaches the **standard normal distribution**.

3) **Cauchy as special case:** The T_1 distribution is the same as the Cauchy distribution.

7.3 Hypothesis test : Introduction and testing one population parameter

7.3.1 Definition

A statistical hypothesis is a *statement* about the *parameters* of one or more populations.

Example 1: A manufacturer claims that the mean life of a smartphone is more than 1.5 years.

Example 2: A local courier service claims that they deliver a ordered product within 30 minutes on average.

Example 3: A sports drink maker claims that the mean calorie content of its beverages is 72 calories per serving.

7.3.2 Types of hypothesis

Statistical hypothesis are stated in two forms- (i) Null hypothesis (H_0) and (ii) Alternative hypothesis (H_1).

Both null and alternative hypothesis are the written about the parameter of interest based on the claim.

- We will always state the null hypothesis as an **equality claim**.
- However, when the alternative hypothesis is stated with the “<” sign, the implicit claim in the null hypothesis can be taken as ” “ or “=” sign.
- When the alternative hypothesis is stated with the “>” sign, the implicit claim in the null hypothesis can be taken as “ ” or “=” sign.

7.3.3 Developing hypotheses

To develop or state null and alternative hypothesis, at first we have to clearly identify the **“claim”** about population parameter. Now we will see some examples.

Example 1: A manufacturer claims that the mean life of a smartphone is more than 1.5 years.

Hypothesis:

$H_0 : \mu = 1.5$

$H_1 : \mu > 1.5 \;\; (claim)$

Example 2: A local courier service claims that they deliver a ordered product within 30 minutes on average.

Hypothesis:

$H_0 : \mu = 30$

$H_1 : \mu < 30 \ \ (claim)$

Example 3: A sports drink maker claims that the mean calorie content of its beverages is 72 calories per serving.

Hypothesis:

$H_0 : \mu = 72 \ \ (claim)$

$H_1 : \mu \neq 72$

7.3.4 Types of test based on alternative hypothesis H_1

- $H_1 : \mu < \mu_0$ (Lower tailed)
- $H_1 : \mu > \mu_0$ (Upper tailed)
- $H_1 : \mu \neq \mu_0$ (Two-tailed)

7.3.5 Types of error in hypothesis test

While testing a statistical hypothesis concerning population parameter we commit two types of errors.

- **Type I error** occurs when we **reject** a **TRUE** H_0
- **Type II error** occurs when we **FAIL to reject** a **FALSE** H_0
- The **Level of significance** is the probability of comiting **Type I error**. It is denoted by α.

$$\alpha = P(Type \ \ I \ \ error)$$

- The probability of committing a **Type II error**, denoted by β.

$$\beta = P(Type \ \ II \ \ error)$$

! Note

Type I error is more serious than **Type II** error. Because rejecting a TRUE statement is more devastating than FAIL to reject a FALSE statement. So, we always try to keep our probability of Type I error as small as possible (1% or at most 5%).

So, how these hypotheses will be tested?

To test a hypothesis we have to determine

- a **test-statistic**; and
- **Critical/Rejection region** based on the sampling distribution of test-statistic for a given α ;
- If the value of test-statistic **falls** in **Critical/Rejection region**, then we reject Null (H_0) hypothesis; otherwise not.

Another way is to use **P-value.** What is p-value?

- The **P-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.
- The **P-value** is the probability of observing a test statistic as extreme as, or more extreme than, the value calculated from your sample data, *assuming that the null hypothesis is true.*
- **If P-value $\leq \alpha$, reject the null hypothesis otherwise we fail to reject H_0.**
- Most of the statistical softwares routinely compute p-value for any hypothesis test.

7.3.6 Hypothesis testing concerning population mean (μ)

The following two hypotheses tests are used concerning population mean (μ):

1. One sample z-test (with known σ)
2. One sample t-test (with unknown σ)

7.3.6.1 One sample z-test

When sampling is from a **normally distributed population** or **sample size is sufficiently large** and t**he population variance is known**, the test statistic for testing $H_0 : \mu = \mu_0$ at α is

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Decision (Critical value approach): If calculated z falls in rejection region (CR) , then reject H_0 . Otherwise, do not reject H_0.

- For lower tailed test, reject H_0 if $z_0 < -z_\alpha$;
- For upper tailed test, reject H_0 if $z_0 > z_\alpha$;
- For two-tailed test, reject H_0 if $z_0 < -z_{\alpha/2}$ or $z_0 > z_{\alpha/2}$.

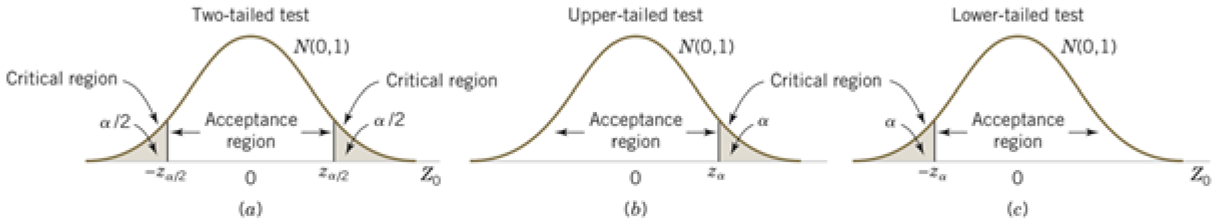


Figure 7.5: Critical region for test of hypothesis (a) Two-tailed test (b) Upper-tailed test (c) Lower-tailed test

Decision (P-value approach): If calculated P-value for the test statistic is less than or equal to α then reject the H_0 otherwise do not reject.

- For lower tailed test, P-value $= P(Z < z_0)$;
- For upper tailed test, P-value $= P(Z > z_0)$;
- For two-tailed test, P-value $P(Z > |z_0|) + P(Z < -|z_0|)]$.

Problem 7.4 State the null and alternative hypothesis in each case.

a) A hypothesis test will be used to potentially provide evidence that the population mean is more than 10.

b) A hypothesis test will be used to potentially provide evidence that the population mean is not equal to 7.

c) A hypothesis test will be used to potentially provide evidence that the population mean is less than 5.

Problem 7.5 (a) For the hypothesis test H0: $= 10$ against H1: >10 and variance known, calculate the P-value for each of the following test statistics.

(i) $z_0 = 2.05$ (ii) $z_0 = -1.84$ (iii) $z_0 = 0.4$

Solution 7.5 (a): Since this is an upper-tailed test so

(i) P-value=$P(Z > 2.05) = P(Z < -2.05) = 0.0202$

(ii) P-value=$P(Z > -1.84) = P(Z < 1.84) = 0.9671$

(iii) P-value=$P(Z > 0.4) = P(Z < -0.4) = 0.3446$

Problem 7.5 (b) For the hypothesis test H0: $= 5$ against H1: <5 and variance known, calculate the P-value for each of the following test statistics. (i) $z_0 = 2.05$ (ii) $z_0 = -1.84$ (iii) $z_0 = 0.4$

Solution 7.5 (b): Since this is an lower-tailed test so

(i) P-value=$P(Z < 2.05) = 0.9798$.

(ii) P-value=$P(Z < -1.84) = 0.03288$.

(iii) do it yourself.

Problem 7.5 (c) For the hypothesis test H0: $= 7$ against H1: 7 and variance known, calculate the P-value for each of the following test statistics.

(i) $z_0 = 2.05$ (ii) $z_0 = -1.84$ (iii) $z_0 = 0.4$

Solution 7.5 (c): Since this is an two-tailed test so

(i) P-value=$P(Z > 2.05) + P(Z < -2.05) = 2P(Z < -2.05) = 0.0404$.

(ii) P-value=$P(Z < -1.84) + P(Z > 1.84) = 2P(Z < -1.84) = 0.0658$.

(iii) do it yourself.

Problem 7.6 The life in hours of a battery is known to be approximately normally distributed with standard deviation $\sigma = 1.25$ hours. A random sample of 10 batteries has a mean life of $\bar{x} = 40.5$ hours. **Conduct** a hypothesis test to justify the claim that battery life exceeds 40 hours.

Problem 7.7 A bearing used in an automotive application is supposed to have a nominal inside diameter of 1.5 inches. A random sample of 25 bearings is selected, and the average inside diameter of these bearings is 1.4975 inches. Bearing diameter is known to be normally distributed with standard deviation $\sigma = 0.01$ inch. **Test** the hypothesis $H_0 : \mu = 1.5$. versus $H_1 : \mu \neq 1.5$ using $\alpha = 0.01$.

Solution 7.7:

Let X be the inside diameter of the bearings (in inches)

Given,

Sample size, $n = 25$;

Sample mean $\bar{x} = 1.4975$ inches; Population SD, $\sigma = 0.01$ inch.

Hypotheses:

$$H_0 : \mu = 1.5$$

$$H_1 : \mu \neq 1.5 \quad (two - tailed)$$

Test statistic: Since X follows normal distribution and σ is known so the test statistic is:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.4975 - 1.5}{0.01/\sqrt{25}} = -1.25$$

Critical value: At $\alpha = 0.05$,

$$-z_{\alpha/2} = -1.96 \quad and \quad z_{\alpha/2} = 1.96$$

Decision: Since z_0 does not fall in critical region so do not reject the H_0.

Conclusion: We can conclude the the mean inside diameter of the bearing is 1.5 inches based on thie sample data.

Problem 7.8 The manufacturer of the X-15 steel-belted radial truck tire claims that the mean mileage the tire can be driven before the tread wears out is 60,000 miles. Assume the mileage wear follows the normal distribution and the standard deviation of the distribution is 5,000 miles. Crosset Truck Company bought 48 tires and found that the mean mileage for its trucks is 59,500 miles. Is Crosset's experience different from that claimed by the manufacturer at the 0.05 significance level?

7.3.6.2 One sample t-test

When sampling is from a **normally distributed population** or **sample size is sufficiently large** and t**he population variance is unknown**, the test statistic for testing $H_0 : \mu = \mu_0$ at α is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Test statistic t follows a Student's distribution with $(n-1)$ degrees of freedom.

Decision (Critical value approach): If calculated t falls in rejection region (CR) , then reject H_0 . Otherwise, do not reject H_0.

- For lower tailed test, reject H_0 if $t < -t_\alpha$;
- For upper tailed test, reject H_0 if $t > t_\alpha$;
- For two-tailed test, reject H_0 if $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$.

Problem 10.4 Annual per capita consumption of milk is 21.6 gallons (*Statistical Abstract of the United States: 2006*). Being from the Midwest, you believe milk consumption is higher there and wish to support your opinion. A sample of 16 individuals from the Midwestern town of Webster City showed a sample mean annual consumption of 24.1 gallons with a standard deviation of $s = 4.8$.

a) Develop a hypothesis test that can be used to determine whether the mean annual consumption in Webster City is higher than the national mean.

b) Test the hypothesis at $\alpha = 0.05$.

c) Draw a conclusion.

Problem 10.5 The mean length of a small counterbalance bar is 43 millimeters. The production supervisor is concerned that the adjustments of the machine producing the bars have changed. He asks the Engineering Department to investigate. Engineer selects a random sample of 10 bars and measures each. The results are reported below in millimeters.

42, 39, 42, 45, 43, 40, 39, 41, 40, 42

Is it reasonable to conclude that there has been a change in the mean length of the bars?

7.3.7 Normality test

In parametric (distribution based) hypothesis test the checking normality assumption of study variable is a common practice especially when the sample size is small ($n < 30$). For large samples, the **Central Limit Theorem (CLT)** often makes this test robust to non-normality.

The normality assumption is checked in two ways:

a) Graphically

b) Numerically using some normality tests

a) Graphical procedure to check normality

We often plot the data (i.e., histogram, density plot, boxplot) to explore so called bell-shaped of the data. But the most popular and effective way to check normality is **Q-Q plot (Quantile-Quantile plot).**

b) Normality test

A number of normality tests are available; of them a common test is **Shapiro-Wilk** test of normality suitable for small to medium sample size (3 to 5000) (**shapiro1965analysis?**; **Royston?**).

Shapiro-Wilk Test Statistic W

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Where,

- $x_{(i)}$: the i^{th} **order statistic** (i.e., the i-th smallest value in the sample)
- $\bar{x}$: the **sample mean**
- a_i : constants calculated based on the **expected values and variances** of order statistics from a **standard normal distribution (**Tabulated in Shapiro Wilk Table)
- n: sample size

Hypotheses:

- **Null Hypothesis** H_0: The data are **normally distributed**.
- **Alternative** H_1: The data are **not normally distributed**.

We reject H_0 if the **p-value** is less than our significance level (e.g., 0.05).

Almost all statistical software and package routinely provide the **Shapiro-Wilk** test.

In **R** Shapiro-Wilk test is available as `shapiro.test` .

```

	Shapiro-Wilk normality test

data:  uniform.data
W = 0.93903, p-value = 0.0001683
```

The p-value<0.05 implies (reject H_0) that the data is not normally distributed.

```

	Shapiro-Wilk normality test

data:  normal.data
W = 0.99212, p-value = 0.83
```

The p-value>0.05 implies (do not reject H_0) that the data is normally distributed.

8 Further topics on random variables

8.1 Transformation of random variables

8.2 Joint Probability Distributions

When events are happened simultaneously, to explore the relationship between two random variables we need **joint probability distributions.**

Definition

The function $f(x,y)$ is said to be a **joint density function** of the continuous random variables X and Y if

1. $f(x,y) \geq 0$, for all (x,y)
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \quad dx \quad dy = 1$
3. $P[(X,Y) \in A] = \int\int_A f(x,y) \quad dx \quad dy$, for any region A in the plane xy.

8.3 R

```r
set.seed(42)

hist(rnorm(1000),freq = FALSE,ylim = c(0,.4),
     breaks = 10,col = "steelblue",main = "Histogram of Normal distribution")
lines(density(rnorm(1000)),col="blue",lwd=2)
```

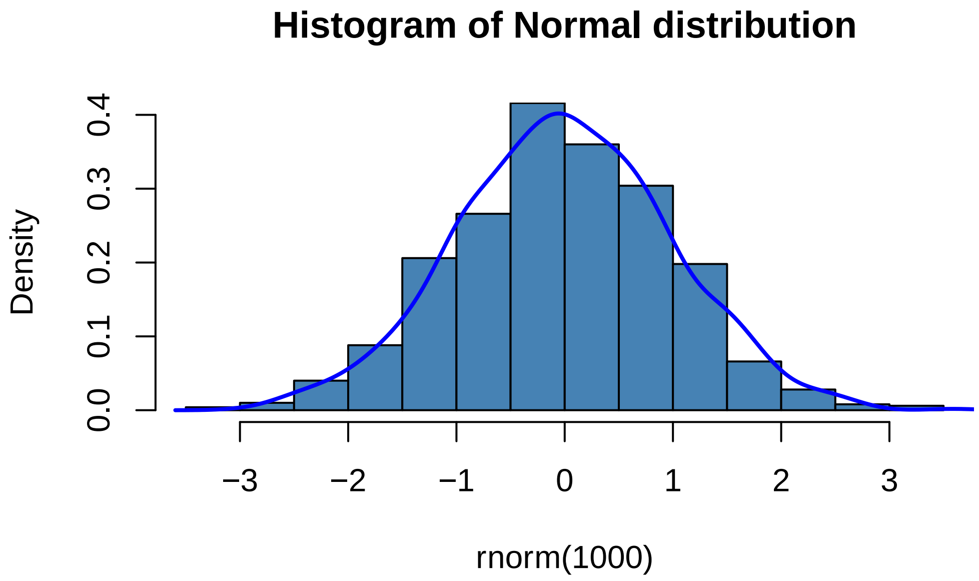


Figure 8.1

8.4 Python

```
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
#import pandas as pd

# Generate 1000 samples
seed = 42

n_rv = np.random.normal(loc=0, scale=1, size=1000)
#print(n_rv)


#plt.clf()  # Clears the current figure

## Using `seaborn`

sns.histplot(n_rv, kde=True,stat="density",bins=10)
```

```
plt.title("Histogram of Normal Distribution")
plt.legend()
plt.show()
```

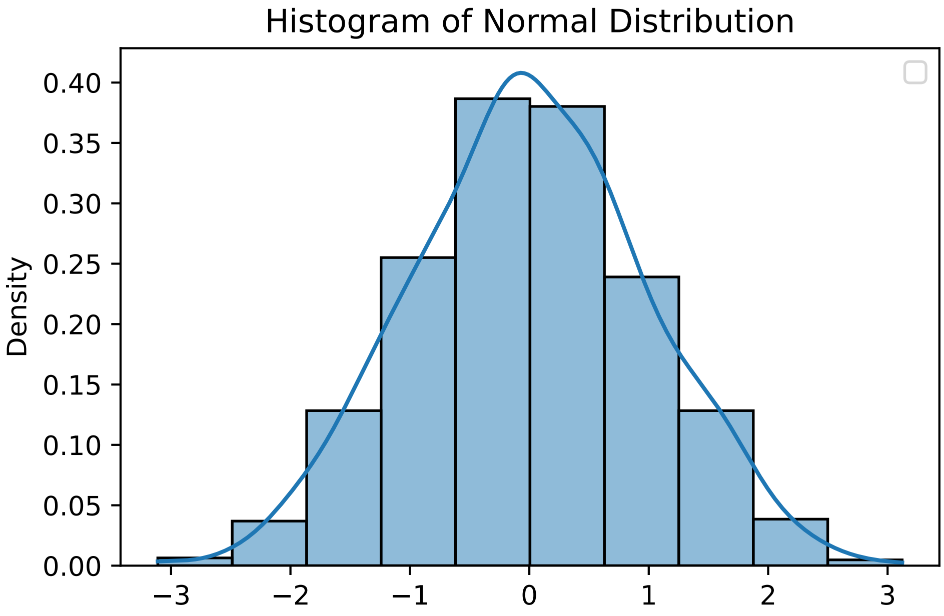


Figure 8.2: Histogram of Normal Distribution

9 Summary

In summary, this book has no content whatsoever.

```
[1] 2
```

References

Baron, Michael. 2019. *Probability and statistics for computer scientists*. Third edition. A Chapman & Hall book. Boca Raton: London.

Bertsekas, Dimitri P., and John N. Tsitsiklis. 2008. *Introduction to probability*. 2nd ed. Optimization and computation series. Belmont: Athena scientific.

Lind, Douglas A., William G. Marchal, and Samuel Adam Wathen. 2012. *Statistical Techniques in Business & Economics*. 15th ed. New York, NY: McGraw-Hill/Irwin.

Montgomery, Douglas C., and George C. Runger. 2014c. *Applied Statistics and Probability for Engineers*. Sixth edition. Hoboken, NJ: John Wiley; Sons, Inc.

———. 2014a. *Applied Statistics and Probability for Engineers*. Sixth edition. Hoboken, NJ: John Wiley; Sons, Inc.

———. 2014b. *Applied Statistics and Probability for Engineers*. Sixth edition. Hoboken, NJ: John Wiley; Sons, Inc.

Navidi, William Cyrus. 2011. *Statistics for Engineers and Scientists*. 3rd ed. New York: McGraw-Hill.

Pishro-Nik, Hossein. 2014. *Introduction to Probability, Statistics, and Random Processes*. Wroclaw: Amazon Fulfillment/Kappa Research.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Walpole, Ronald E., Raymond H. Myers, Sharon L. Myers, and Keying Ye. 2017a. *Probability & statistics for engineers & scientists: MyStatLab update*. Ninth edition. Boston: Pearson.

———, eds. 2017b. *Probability & Statistics for Engineers & Scientists: MyStatLab Update*. Ninth edition. Boston: Pearson.