ALGERIAN DEMOCRATIC AND POPULAR REPUBLIC

MINSTRY OF HIGHER EDUCATION

AND SCIENTIFIC RESEARCH

20 AUGUST 1955 UNIVERSITY SKIKDA

FACULTY OF SCIENCES

DEPARTMENT OF MATHEMATICS

Probability Course

Karima Kimouche

UNIVERSITY YEAR : 2024-2025

Probability Course

Karima Kimouche
20 august 1955 university Skikda

University year : 2024-2025

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

In a world increasingly driven by data and uncertainty, the study of probability has become a cornerstone of modern science, engineering, finance, and decision-making. This manuscript is an introduction to probability and is intended for second-year mathematics students at the University of 20 August 1955, Skikda. It can also be useful for students from other disciplines seeking to deepen their knowledge of probability.

This first version of the handout is not intended to be a reference work; rather, it is designed to serve as a memory aid for students approaching a probability course for the first time.

The content of this manuscript is structured into three chapters:

The first chapter is devoted to the basic notions of probability, defining random experiments and events, as well as key concepts such as conditional probability, the total probability formula, Bayes' theorem, and the notion of independence of events, which is specific to probability theory.

The second chapter focuses on random variables. After the definition of this notion we study in detail the two large families of random variables, namely discrete variables and continuous variables. We provide definitions and key properties of probability mass function, density function and cumulative distribution functions, including expectation and variance in both cases. Additionally, we explore inequalities in probability.

Given the importance of probability distribution, the third chapter illustrates the definitions and main properties of common probability distributions: discrete probability distributions and continuous probability distributions. We also consider the approximations between the main distribution such as the convergence of a binomial distribution to the Poisson distribution, as well the transformations of random variables.

In writing this manuscript, I relied on various recognized references in the field. It is important to note that this first version does not claim any originality; the content presented is standard and is found in most books on modern probability theory. Like any academic work, it may contain errors, and I wish to express my gratitude and thanks in advance to any colleagues or students who share their feedback and criticisms to assist in the development of this manuscript.

CHAPTER

1

BASIC CONCEPTS OF PROBABILITY

Contents

This chapter introduces a few concepts from probability theory.

1.1 Introduction

Probability, as a mathematical discipline, has fascinated thinkers for centuries due to its profound implications in understanding uncertainty and making informed decisions. Its origins can be traced back to the 16th century, when mathematicians began to explore the concepts of chance through games of dice and cards. Pioneering figures such as Gerolamo Cardano and Pierre de Fermat laid the groundwork by formulating foundational principles. In the 17th century, Blaise Pascal and Fermat further advanced the field by developing methods for calculating probabilities, particularly in gambling scenarios. The 18th century saw the emergence of probability as a formal discipline with contributions from mathematicians like Jacob Bernoulli, who introduced the Law of Large Numbers, and Pierre-Simon Laplace, who extended the theory and applied it to various scientific fields.
Today, probability is integral to diverse areas, it is useful in the biological, physical, actuarial, management, computer sciences, economics, engineering and operations research. It helps in modeling complex systems and in decision-making when there is uncertainty. It can be used to prove theorems in other mathematical fields (such as analysis, number theory, game theory, graph theory, quantum theory and communications theory). Its rich history not only highlights its mathematical beauty but also underscores its relevance in addressing real-world problems.

1.2 Basic concept

1.2.1 Sample space and events

In probability theory, a probability space or a probability triple $(\Omega, \mathcal{F}, P)$ is a mathematical construct that provides a formal model of a random process or experiment. For example, one can define a probability space which models the throwing of a die.
A probability space consists of three elements:

1. A sample space Ω, which is the set of all possible outcomes of a random experiment.

2. An event space, which is a set of events, $\mathcal{F}$ an event being a set of outcomes in the sample space.

3. A probability function, P, which assigns, to each event in the event space, a probability, which is a number between 0 and 1 (inclusive).

1.2.2 Basic probability notation

Set representation	Probabilistic formulation
Ω	sample space
$\{\omega\} \subset \Omega$	elementary event
$A \subset \Omega$	event A
$A \subset B$	occurrence of A implies occurrence of B
$\overline{A} = C_{\Omega}^{A}$	the complement event
$A \cap B$	A and B
$A \cup B$	A or B fulfilled
$A \cap B = \emptyset$	A and B are mutually exclusive or disjoint events
$(\Omega = A \cup B)$ and $(A \cap B = \emptyset)$	A and B form a partition of a sample space

Table 1.1: The set representation with probabilistic formulation.

Example 1.1. *In the example of the throw of a standard die:*

1. *The sample space Ω is typically the set $\{1, 2, 3, 4, 5, 6\}$ where each element in the set is a label that represents the outcome of the die landing on that label. For example,* 1 *represents the outcome that the die lands on* 1.

2. *The event space $\mathcal{F}$ could be the set of all subsets of the sample space, which would then contain simple events such as* 5 *("the die lands on 5"), as well as complex events such as* $\{2, 4, 6\}$ *("the die lands on an even number").*

3. *The probability function P would then map each event to the number of outcomes in that event divided by* 6*. So for example,* $\{5\}$ *would be mapped to* $\frac{1}{6}$*, and* $\{2,4,6\}$ *would be mapped to* $\frac{3}{6}=\frac{1}{2}$*.*

1.3 Probability

1.3.1 Definition

Let Ω be the sample space and A be an event associated with a random experiment. Let $card(\Omega)$ and $card(A)$ be the number of elements of Ω and A. Then the probability of event A occurring is denoted as $P(A)$, is denoted by

$$P(A)=\frac{card(A)}{card(\Omega)}.$$

Definition 1.1. *Let Ω be the sample space and A be an event associated with a random experiment. The probability of the event A, $P(A)$ is defined as a real number satisfying the following axioms:*

1. $0 \leq P(A) \leq 1$.

2. $P(\Omega)=1$.

3. *If A and B are mutually exclusive events,*

$$P(A\cup B)=P(A)+P(B).$$

4. *If $A_1 A_2,\ldots,A_n,\ldots$ are mutually exclusive events,*

$$P(A_1\cup A_2\cup\ldots\cup A_n\ldots)=P(A_1)+P(A_2)+\ldots+P(A_n)\ldots$$

1.3.2 Properties

1. If A is an impossible event, $P(A)=0$.

2. If A is a certain event, $P(A)=1$.

3. If $\overline{A}$ is the complementary event of $A, P\left(\overline{A}\right)=1-P(A)$.

4. In addition, If A and B are any events,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B),$$

and it can be extended to any 3 events, A, B and C

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

5. If $B \subset A, P(B) \leq P(A)$.

1.4 Conditional probability

1.4.1 definition

Definition 1.2. *If A and B are two events associated with the same sample space of a random experiment, then the conditional probability of the event B under that the event A has happened, is denoted by $P(B|A)$ and defined as*

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, \quad \textit{provided} \quad P(A) \neq 0.$$

1.4.2 Properties of conditional probability

Let A, B and C be three events associated with the sample space Ω of an experiment. Then

i) $P(\Omega|A) = P(\Omega|B) = 1$.

ii) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.

iii) $P\left(\overline{A}|B\right) = 1 - P(A|B)$.

1.4.3 Product Theorem on probability

1. Let A and B be two events associated with the sample space Ω of an experiment. Then

$$\begin{aligned} P(A \cap B) &= P(A|B).P(B), P(B) \neq 0 \\ &= P(B|A).P(A), P(A) \neq 0. \end{aligned}$$

2. If A, B and C be three events associated with the sample space Ω of an experiment. Then
$$P(A \cap B \cap C) = P(A).P(B|A)P(C|A \cap B).$$

3. More general, if $A_1, ..., A_n$ a family of n events such that: $P(A_1 \cap A_2 \cap ... \cap A_n) \neq 0$, then
$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1 \cap A_2)...P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1}).$$

1.5 Bayes Theorem

1.5.1 Partition of a sample space

A set of events $B_1, B_2, ..., B_n$ is said to represent a partition of a sample space Ω if

i) $B_i \cap B_j = \emptyset, i \neq j = 1, 2, ..., n,$

ii) $B_1 \cup B_2 \cup ... \cup B_n = \Omega,$

iii) Each $B_i \neq \emptyset$, *i.e.*, $P(B_i) > 0$ for all $i = 1, 2, ..., n$.

1.5.2 Theorem of total probability

Let $\{B_1, B_2, ..., B_n\}$ be a partition of the sample Ω. Let A be any event associated with Ω, then
$$P(A) = \sum_{k=1}^{n} P(A \cap B_k) = \sum_{k=1}^{n} P(A|B_k)P(B_k).$$

Theorem 1.1 (Bayes). *If $B_1, B_2, ..., B_n$ are mutually exclusive and exhaustive events associated with a sample space, and A is any event of non zero probability, then*
$$P(B_k|A) = \frac{P(B_k).P(A|B_k)}{\sum_{k=1}^{n} P(B_k)P(A|B_k)}.$$

1.6 Independent events

If the two events A and B are independent, the product theorem takes the form
$$P(A \cap B) = P(A)P(B).$$

Conversely, if $P(A \cap B) = P(A)P(B)$, the events are said to be independent (pairwise independent). Thus, two events A and B will be independent:

i) $P(A|B) = P(A)$, provided $P(B) \neq 0$

ii) $P(B|A) = P(B)$, provided $P(A) \neq 0$

Furthermore,

$$\begin{aligned} A \text{ and } B \text{ independents} &\Longleftrightarrow A \text{ and } \overline{B} \text{ are independents} \\ &\Longleftrightarrow \overline{A} \text{ and } B \text{ are independents} \\ &\Longleftrightarrow \overline{A} \text{ and } \overline{B} \text{ are independents} \end{aligned}$$

Three events A, B and C are said to be mutually independent if all the following conditions hold:

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \end{aligned}$$

and

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

CHAPTER

2

RANDOM VARIABLES

Contents

All the experiments that are performed can be classified into two broad divisions viz, random experiments and deterministic experiments. Deterministic experiments are those experiments in which the outcome of the experiment remains the same whenever it is performed. But in the case of a random experiment the outcome is found to vary each time the experiment is performed. Here the experimenter may know the set of all possible outcomes of the random experiment but cannot say with certainty which outcome will occur when the experiment is performed. For example: In the case of throwing a die, we may use a variable X for representing the out come of the throw. Thus X will take the values $1, 2, 3, 4, 5$ and 6. But in some cases the outcomes may be qualitative e.g. tossing of a coin which may be head or tail, the colors of balls drawn from an urn may be red, yellow, white etc. But for mathematical convenience, the qualitative outcomes may be expressed in quantitative forms. For example, in tossing of a coin we may denote the outcome "Head" by 1 and "Tail" by 0. In this way each outcome of a random experiment, whether it is qualitative or quantitative, can be expressed by a real number. The real number, which is associated with the outcome of random experiment, is called a random variable [1]. The random variable takes certain values depending on chance, so it is also called a chance variable or a stochastic variable. In other words, a random variable is a function that associates each single outcome of an experiment with a real number. Their significance can be seen from the fact that it is theoretically easier to deal with numbers than outcomes.

[1] The Holy Roman Empire was, in the words of the historian Voltaire, "neither holy, nor Roman, nor an empire". Similarly, a random variable is neither random nor a variable.

2.1 Random variables

2.1.1 Definition

A random variable can be defined as below:

Definition 2.1. *A real valued function X, defined on a sample space Ω of a random experiment, is called a random variable which assigns to each sample point, one and only one real number $X(\omega) = x$ where $\omega \in \Omega$. i.e,*

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}, \\ \omega &\longmapsto X(\omega). \end{aligned}$$

Definition 2.2. *The sample space Ω is the domain of a random variable, and the range space or value set is the set of possible values that a random variable X can take.*
The range space for a random variable X is often denoted $X(\Omega)$:

$$X(\Omega) = \{x \in \mathbb{R} : \exists \omega \in \Omega; \text{ such that } X(\omega) = x\}.$$

Remark 2.1.

1. *We shall represent a random variable by a capital letter (such as X, Y, or W) and any particular value of the random variable by a lower case letter (such as x, y, or w).*
2. *The standard abbreviation for random variable is "r.v.".*

Example 2.1.

1. *Toss coin n times and count the number of heads.*
2. *Roll a die.*
3. *Draw a card from a deck.*
4. *Number of car accidents per week.*
5. *Number of defective items in a given company.*
6. *Number of bacteria per two cubic centimeter of water.*
7. *Height of students at certain college.*

8. *Mark of a student.*

9. *Life time of light bulbs.*

10. *Length of time required to complete a given training.*

The random variables in the above examples are representatives of the two types of random variables:

1. Discrete random variable (random variables have discrete values; the sample space can be discrete, continuous or even mixture of discrete and continuous).

2. Continuous random variable (continuous range of values, it cannot be produced from a discrete sample space or a mixed sample space).

Example 2.2. *Find the range for each of the following random variables:*

1. *I toss a coin 100 times. Let X be the number of heads I observe.*

2. *I toss a coin until the first heads appears. Let Y be the total number of coin tosses.*

3. *The random variable T is defined as the time (in hours) from now until the next earthquake occurs in a certain city.*

Solution:

1. *The random variable X can take any integer from 0 to 100, so $X(\Omega) = \{0, 1, 2, ..., 100\}$.*

2. *The random variable Y can take any positive integer, so $Y(\Omega) = \{1, 2, 3, ...\} = \mathbb{N}$.*

3. *The random variable T can in theory get any nonnegative real number, so $T(\Omega) = [0, \infty)$.*

Remark 2.2. *Mixed random variables have some part taking values over an interval like typical continuous variables, and part of it concentrated on particular values like discrete variables.*

Example 2.3. *Consider the time spent by vehicles waiting at a set of traffic lights before proceeding through the intersection.*
If the light is green on arrival, the wait time is exactly zero (i.e., discrete): the vehicle can drive straight through the intersection. A wait time of zero seconds can be measured exactly. However, if the light is red on arrival, the vehicle needs to wait a continuous amount of time before it turns green.
The time spent waiting is a mixed random variable.

2.1.1.1 Probability of a random variable

The standard probability axioms are the foundations of probability theory introduced by Russian mathematician Andrey Kolmogorov in 1933. These axioms remain central and have direct contributions to mathematics, the physical sciences, and real-world probability cases.
Let $X : \Omega \to \mathbb{R}$ be a random variable such that $X(\Omega) = \{x_1, x_2, ..., x_n\}$, the probability that the variable X is equal to x is the probability of the elements of Ω having the image of the value x in the mapping. This probability measures the degree of credibility of the occurrence of the event.

According to Kolmogorov axioms, we have:

1. $0 \leq P(X = x_i) \leq 1$.
2. $P(X = X(\Omega)) = 1$.
3. If the x_i are incompatible with each other, we have for all $i \neq j$ $P[(X = x_i)$ *or* $(X = x_j)] = P(X = x_i) + P(X = x_j)$.

Remark 2.3. *Conditions for a function to be a random variable:*

- *It not be multivalued.*
- *The set* $\{X \leq x\}$ *shall be an event for any real number* x.
- $P(\{X \leq x\})$ *is equal to the sum of the probabilities of all the elementary events corresponding to* $\{X \leq x\}$.
- *The probabilities of events* $\{X = +\infty\}$ *and* $\{X = -\infty\}$ *is:*

$$P(\{X = +\infty\}) = 0, P(\{X = -\infty\}) = 0.$$

2.1.2 Probability distribution function

In carrying out any experiment, we generate test results which are basically random by nature. Hence, a random variable provides us a means for describing the experimental outcomes using numerical values. In other words, random variables must assume numerical data for statistical interpretation. The distribution obtained by taking the

possible values of a random variable together with their respective probabilities is called a probability distribution.
The probability distribution for a random variable describes how probabilities are distributed over the values of the random variables. A probability distribution can be presented either with the help of a function or in tabular form where values of the random variable and corresponding probability are shown. The probability distribution for a discrete random variable is called as a discrete probability distribution or " **probability mass function**" (pmf) and that of a continuous random variable is called a "**probability density function**" (pdf).
For discrete random variables, we can enumerate all of the possible realizations of the random variable, and associate a specific probability with each possible realization. In contrast, for continuous random variables, we cannot enumerate all of the possible realizations of the random variable, so it is impossible to associate a specific probability with each possible realization. As a result, we must define the probabilities of discrete and continuous random variable occurrences using distinct (but related) concepts.

2.1.3 Cumulative distribution function

Another way of describing random variables is using a distribution function (df), also called a cumulative distribution function (cdf).

Definition 2.3. *Let X denote a random variable and x its particular value from the whole range of all values of X, say $X(\Omega)$. The function defined by*

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto P(X \leq x). \end{aligned}$$

is called the cumulative distribution function (cdf) of the r.v. X.

Example 2.4. *Let X be the number of heads in two independent tosses of a fair coin. In particular $P(X = 0) = P(X = 2) = 1/4$, and $P(X = 1) = 1/2$. Then,*

$$F_X(x) = \begin{cases} 0, & \text{if} \quad x < 0, \\ \frac{1}{4}, & \text{if} \quad 0 \leq x < 1, \\ \frac{3}{4}, & \text{if} \quad 1 \leq x < 2, \\ 1 & \text{if} \quad x \geq 2. \end{cases}$$

Remark 2.4. *We say that two random variables X and Y have the same distribution if they have the same cumulative distribution function $F_X = F_Y$.*

Remark 2.5. *Let I be an interval of $\mathbb{R}$. The event $\{X \leq x\}$ represents the set of values $\omega \in \Omega$ such that $X(\omega)$ is less than x, i.e.$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$.*

Remark 2.6. *We have $P(X \in \mathbb{R}) = 1$, because $P(X \in \mathbb{R}) = P(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\}) = P(\Omega) = 1$.*

2.1.3.1 Properties of cumulative distribution functions

The cumulative distribution function of a random variable are very useful. Thus, we note a few of their basic properties here.

Theorem 2.1. *Let X be a random variable, and let F be its cumulative distribution function (cdf).*

a) $\forall x \in \mathbb{R}, 0 \leq F_X(x) \leq 1$.

b) *Monotonicity: If $x \leq y$, then $F_X(x) \leq F_X(y)$.*

c) *Limiting values: We have $\lim\limits_{x \to +\infty} F_X(x) = 0$ and $\lim\limits_{x \to +\infty} F_X(x) = 1$.*

d) *Right-continuity: For every x, we have $\lim_{y \searrow x} F_X(y) = F_X(x)$.*

Proof.

a) It follows from the definition of probability.

b) Suppose that $x \leq y$. Then, $\{X \leq x\} \subset \{X \leq y\}$, which implies that

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y).$$

c) Since $F_X(x)$ is monotonic in x and bounded below by zero, it converges as $x \to -\infty$, and the limit is the same for every sequence $\{x_n\}$ converging to $-\infty$. So, let $x_n = -n$, and note that the sequence of events $\bigcap_{n=1}^{\infty} \{X \leq -n\}$ converges to the empty set. Using the continuity of probabilities, we obtain

$$\lim_{x \to +\infty} F_X(x) = \lim_{x \to +\infty} F_X(-n) = \lim_{x \to +\infty} P(X \leq -n) = P(\emptyset) = 0.$$

The proof of $\lim\limits_{x \to +\infty} F_X(x) = 1$ is similar, and is omitted.

d) Consider a decreasing sequence $\{x_n\}$ that converges to x. The sequence of events $\{X \le n\}$ is decreasing and $\bigcap_{n=1}^{\infty} \{X \le x_n\} = \{X \le x\}$. Using the continuity of probabilities, we obtain

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(X \le x_n) = P(X \le x) = F_X(x).$$

Since this is true for every such sequence $\{x_n\}$, we conclude that $\lim_{y\searrow x} F_X(y) = F_X(x)$ (see Appendix A).

□

Remark 2.7.

1. *For the complement of $\{X \le x\}$, we have the survival function*

$$\overline{F}_X(x) = P(X > x) = 1 - P(X \le x) = 1 - F_X(x).$$

2. *Choose $a < b$, then the event $\{X \le a\} \subset \{X \le b\}$. Their set theoretic difference*

$$\{X \le b\} \setminus \{X \le a\} = \{a < X \le b\}.$$

Consequently, by the difference rule for probabilities

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a).$$

Proposition 2.1. *Let X be a random variable, and let F be its cumulative distribution function (cdf). For all $a, b \in \mathbb{R}$ with $a < b$, we have:*

1. $P(a < X \le b) = F_X(b) - F_X(a).$
2. $P(a \le X \le b) = F_X(b) - F_X(a) + P(X = a).$
3. $P(a < X < b) = F_X(b) - F_X(a) - P(X = b).$
4. $P(a \le X < b) = F_X(b) - F_X(a) + P(X = a) - P(X = b).$

Proof. The proof of the other properties is similar to the first property. Indeed

1. Choose $a < b$, then the event $\{X \le a\} \subset \{X \le b\}$, Their set theoretic difference

$$\{X \le b\} \setminus \{X \le a\} = \{a < X \le b\}.$$

Consequently, by the difference rule for probabilities,

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F_X(b) - F_X(a).$$

□

2.2 Discrete random variables

2.2.1 Definition

If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers (countable), it is called a discrete sample space.

Definition 2.4. *A random variable is called a discrete random variable if its set of possible outcomes $X(\Omega)$ is finite or countable, i.e. the values can be arranged in a sequence.*

Indeed, by simply listing out all the possible values x such that $P(X = x) > 0$, we obtain a second, equivalent definition, as follows.

Definition 2.5. *A random variable X is discrete if there is a finite or countable sequence $x_1, x_2, \ldots$ of distinct real numbers, and a corresponding sequence $p_1, p_2, \ldots$ of nonnegative real numbers, such that $P(X = x_i) = p_i$ for all i, and $\sum_i p_i = 1$.*

This second definition also suggests how to keep track of discrete distributions. It prompts the following definition.

2.2.2 Probability function of discrete r.v.

Definition 2.6. *Let X be a discrete random variable, we call the probability function of X (or pf) for short, the function defined by*

$$\begin{aligned} p_X : \mathbb{R} &\longrightarrow [0,1] \\ x_i &\longmapsto p_X(x_i) \end{aligned}$$

with

$$p_X(x_i) = \begin{cases} P(X = x_i) & \text{if } x_i \in X(\Omega) \\ 0 & \text{otherwise.} \end{cases}$$

We will often write $p(x)$ instead of $p_X(x)$.

The probability function (pf) is sometimes given the alternative name of probability mass function (pmf).
The probability function has the following basic properties:

Proposition 2.2. *Let X be a random variable with probability (mass) function. Then*

1. $0 \le p_X(x) \le 1$, *for all* $x \in \mathbb{R}$.
2. $\sum_{x \in X(\Omega)} p_X(x) = 1$.
3. $p_X(x) = 0$ *for all* X *outside a discrete set.*

Definition 2.7. *If* $X(\Omega) = \{x_1, x_2, ...\}$ *the pair*

$$\{(x_i, p_X(x_i)); i = 1, 2, ...\}$$

is called the probability distribution of the discrete random variable X.

The probability distribution of a discrete random variable X can be represented by listing each outcome with its probability, giving a formula, using a table or using a graph which displays the probabilities $p(x)$ corresponding to each $x \in X(\Omega)$ as shown in the table below

x_i	x_1	x_2	...	x_n
$p(x_i) = P(X = x_i)$	$p(x_1)$	$p(x_2)$	...	$p(x_n)$

Example 2.5. *We toss a fair coin twice, and let* X *be defined as the number of heads we observe. Find the range of* X, $X(\Omega)$, *as well as its probability (mass) function* P_X.
Solution: *Here, our sample space is given by*

$$\Omega = \{HH, HT, TH, TT\}.$$

The connection between the sample space and H *is shown in the table below. In this case, the range space of* X *is* $X(\Omega) = \{0, 1, 2\}$.

Element of Ω	*Function* $X(\omega)$	*Value de* X	$P(X = x_i)$
TT	$X(\omega_1)$: *number of heads in* ω_1	0	1/4
TH	$X(\omega_2)$: *number of heads in* ω_2	1	1/4
HT	$X(\omega_3)$: *number of heads in* ω_3	1	1/4
HH	$X(\omega_4)$: *number of heads in* ω_4	2	1/4

The probabilities are

$$P(X = 0) = P(\textit{no heads}) = \frac{1}{4},$$
$$P(X = 1) = P(\textit{one heads}) = \frac{2}{4},$$
$$P(X = 2) = P(\textit{two heads}) = \frac{1}{4}.$$

As a function, the probability mass function is

$$p_X(k) = P(X = k) = \begin{cases} \frac{1}{4} & \text{if} \quad k = 0, \\ \frac{1}{2} & \text{if} \quad k = 1, \\ \frac{1}{4} & \text{if} \quad k = 2, \\ 0 & \text{otherwise.} \end{cases}$$

More succinctly,

$$p_X(k) = P(X = k) = \begin{cases} \frac{1}{2} \times \left(\frac{1}{2}\right)^{|k-1|} & \text{for} \quad k = 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

This information can also be presented as a table (Table 2.1) or graph (Figure 2.1)

k	0	1	2
P$(X = k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Table 2.1: The probability mass function of r.v. X: tossing a fair coin twice.

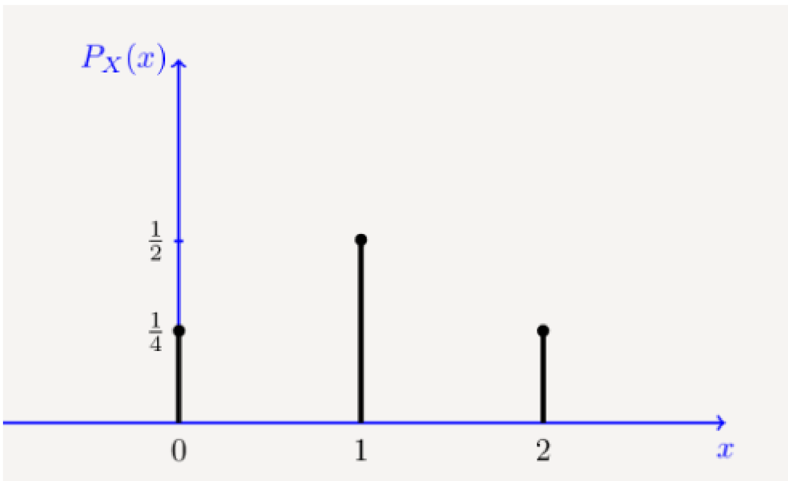


Figure 2.1: The probability mass function of random variable X: tossing a fair coin twice.

Remark 2.8. *If X is discrete random variable then*

1. $P(a < X < b) = \sum_{x=a+1}^{b-1} p(x)$.
2. $P(a \leq X < b) = \sum_{x=a}^{b-1} p(x)$.

3. $P(a < X \leq b) = \sum_{x=a+1}^{b} p(x)$.

4. $P(a \leq X \leq b) = \sum_{x=a}^{b} p(x)$.

Example 2.6. *A random variable X has the following distribution.*

x	1	2	3	4	5	6	7	8
$P(X = x)$	k	$2k$	$3k$	$4k$	$5k$	$6k$	$7k$	$8k$

Find the vales of:

1. $P(X \leq 2)$.

2. $P(2 \leq X \leq 5)$.

Solution: *First we find the value of k, we have and*

$$\begin{aligned}\sum_{i=1}^{8} P(X = x_i) &= k + 2k + 3k + 4k + 5k + 6k + 7k + 8k \\ &= 36k\end{aligned}$$

and

$$\sum_{i=1}^{8} P(X = x_i) = 1$$

which implies that $k = \frac{1}{36}$, *thus*

1.

$$\begin{aligned}P(X \leq 2) &= P(X = 1) + P(X = 2) \\ &= \frac{1}{36} + \frac{2}{36} \\ &= \frac{3}{36} = \frac{1}{12}.\end{aligned}$$

2.

$$\begin{aligned}P(2 \leq X \leq 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} \\ &= \frac{14}{36} = \frac{7}{18}.\end{aligned}$$

Example 2.7. *Let $n \geq 1, C > 0$ and p be the real function defined by*

$$p(x) = \begin{cases} C\frac{x}{n(n+1)} & \text{if} \quad x \in \{1, 2, ..., n\} \\ 0 & \text{otherwise} \end{cases}$$

Under what conditions is p a probability (mass) function?
Solution: *First we check the condition $0 \leq p(x) \leq 1$. A simple calculation shows that this condition is satisfied if $C \leq n+1$. On the other hand we must also have $\sum_{x\in\mathbb{N}} p(x) = 1$. Therefore*

$$1 = \sum_{x\in\mathbb{N}} p(x) = \frac{C}{n(n+1)} \sum_{x=1}^{n} x = \frac{C}{n(n+1)} \times \frac{n(n+1)}{2} = \frac{C}{2}.$$

So, if $C = 2$ both conditions are verified and p is indeed a probability (mass) function.

Indeed, by simply listing out all the possible values x such that $P(X = x) > 0$, we obtain a second, equivalent definition, as follows.

Definition 2.8. *A random variable X is discrete if there is a finite or countable sequence $x_1, x_2, ...$ of distinct real numbers, and a corresponding sequence $p_1, p_2, ...$ of nonnegative real numbers, such that $P(X = x_i) = p_i$ for all i, and $\sum_i p_i = 1$.*

A discrete random variable is simply an application that takes discrete values $\{x_1, x_2, ..., x_n, ...\}$. Hence, if x_1, x_2 are the distinct values such that $P(X = x_i) = p_i$ for all i with $\sum_i p_i = 1$, then

$$p_X(x) = \begin{cases} p_i & \text{if} \quad x = x_i \quad \text{for some } i \\ 0 & \text{otherwise.} \end{cases}$$

A simple way of describing the probabilistic properties of a discrete random variable is in terms of the cumulative distribution function, which we now define.

2.2.3 Cumulative distribution function of discrete r.v.

Definition 2.9. *Given a random variable X, its cumulative distribution function (or cdf for short) is the function $F_X : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = P(X \leq x)$.*

The cumulative distribution $F_X(x)$ is "the total probability you have accumulated when you get to x".
The reason for calling $F_X(x)$ the "cumulative distribution function" is that the full distribution of X can be determined directly from $F_X(x)$. We use the $F_X(x)$ form when we need to make the identity of the random variable clear.

Remark 2.9. *For discrete random variables, the pmf is also called the probability distribution. Thus, when asked to find the probability distribution of a discrete random variable X, we can do this by finding its pmf. The phrase distribution function is usually reserved exclusively for the cumulative distribution function cdf.*

We can compute the cumulative distribution function (cdf) $F_X(x)$ of a discrete random variable from its probability function $p_X(x)$, as follows.

Theorem 2.2. *Let X be a discrete random variable with probability function $p_X(x)$. Then its cumulative distribution function $F_X(x)$ satisfies $F_X(x) = \sum_{y \leq x} p_X(y)$.*

Proof. Let $x_1, x_2, \ldots$ be the possible values of X. Then

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{x_i \leq x} P(X = x_i) \\ &= \sum_{y \leq x} P(X = y) \\ &= \sum_{y \leq x} p_X(y). \end{aligned}$$

□

The cumulative distribution function F_X of a discrete random variable is constant except for jumps. At the jump, F_X is right continuous

$$\lim_{x \to x_0+} F_X(x) = F_X(x_0).$$

If X is a discrete variable with values in $\{x_1, x_2, \ldots, x_n\}$with $x_1 < x_2 < \ldots < x_n$, then for all $x \in \mathbb{R}$

$$F_X(x) = \sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} p_i \text{ with } k \text{ such that } x_k \leq x < x_{k+1}.$$

Similarly, if X takes an infinity of values $\{x_1, x_2, \ldots, x_n, \ldots\}$with $x_1 < x_2 < \ldots < x_n < \ldots$, we have for all $x \in \mathbb{R}$

$$F_X(x) = \sum_{i=1}^{k} P(X = x_i) = \sum_{i=1}^{k} p_i \text{ with } k \text{ such that } x_k \leq x < x_{k+1}.$$

In particular

$$F_X(x) = \begin{cases} 0, & \text{if} \quad x < x_1 \\ p_1, & \text{if} \quad x_1 \le x < x_2 \\ p_1 + p_2, & \text{if} \quad x_2 \le x < x_3 \\ \quad \vdots & \qquad \vdots \\ p_1 + p_2 + ...p_{n-1}, & \text{if} \quad x_{n-1} \le x < x_n \\ 1, & \text{if} \quad x_n \le x < \infty \end{cases}$$

which have the following graphic representation:

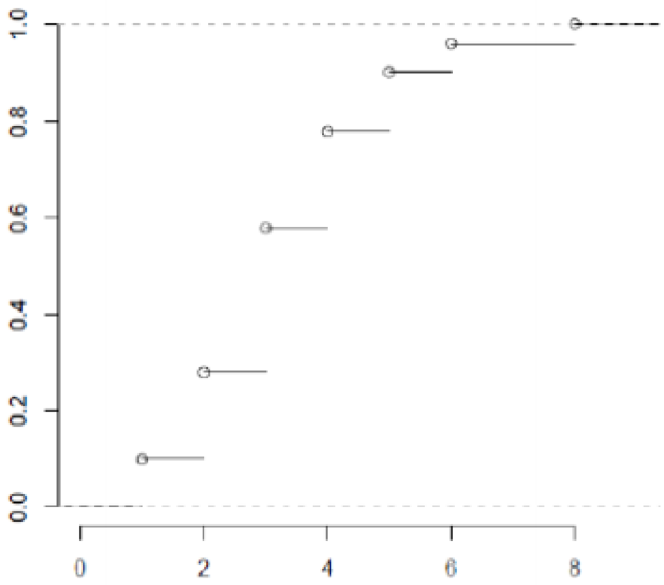


Figure 2.2: The cumulative distribution function F_X of a discrete random variable X.

The jumps of the distribution function $F_X(x)$ occur at points x_i and the height of the jump at point x_i is equal to $P(X = x_i)$. It is therefore sufficient to calculate the distribution function at points x_i .

We will need the relation between the probability (mass) function $p(x)$ and the cumulative distribution function $F_X(x)$. Recall that if $F_X(x)$ is a function of an integer variable x then the backward difference function (discrete derivative) ΔF_X of F_X is defined by

$$\Delta F_X(x) = F_X(x) - F_X(x-1).$$

The relation we want is

$$P(x) = \Delta F_X(x).$$

Proposition 2.3. *If X is discrete-valued in $\{x_1,...,x_n\}$ (or $\{x_1,...,x_n,...\}$), the distribution of X is entirely characterized by $\{P(X=x_i): i \geq 1\}$.*

Example 2.8. *Consider the random experiment of tossing a fair coin three times and observing the result (a Head or a Tail) for each toss. Then the sample space Ω which the set of 3-tuples of heads and tails is*

$$\begin{aligned}\Omega &= \{T,H\} \times \{T,H\} \times \{T,H\} \\ &= \{TTT, THH, HTH, HHT, HTT, THT, TTH, HHH\}.\end{aligned}$$

Let X denote the total number of heads obtained in the three tosses of the coin, so we have

$$\begin{aligned}&X(\{TTT\}) = 0, \\ &X(\{HTT\}) = X(\{THT\}) = X(\{TTH\}) = 1, \\ &X(\{THH\}) = X(\{HTH\}) = X(\{HHT\}) = 2, \\ &X(\{HHH\}) = 3.\end{aligned}$$

Then, the range of X is $X(\Omega) = \{0,1,2,3\}$. Note card $(\Omega) = 8$, we have

$$\begin{aligned}&P(X=0) = P(\{TTT\}) = \frac{1}{8}, \\ &P(X=1) = P(\{HTT\}) + P(\{THT\}) + P(\{TTH\}) = \frac{3}{8}, \\ &P(X=2) = P(\{THH\}) + P(\{HTH\}) + P(\{HHT\}) = \frac{3}{8}, \\ &P(X=3) = P(\{HHH\}) = \frac{1}{8},\end{aligned}$$

we will tabulate this

X	0	1	2	3
$P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Table 2.2: The probability function of discrete variable X: tossing a fair coin three times.

and we also illustrate these by two kinds of graphical representations: line graph and histogram

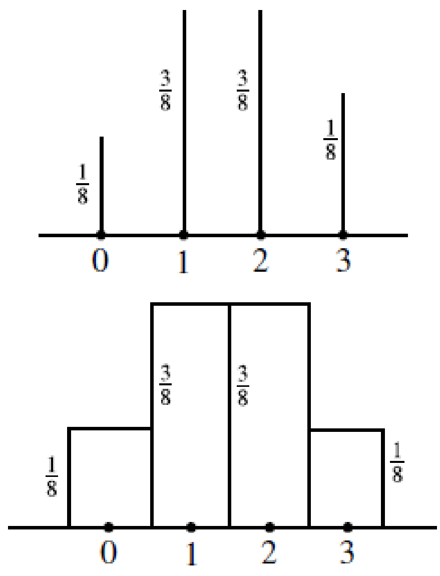


Figure 2.3: Line graph and histogram of discrete random variable X: tossing a fair coin three times.

The cumulative distribution function is

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{8}, & \text{if } 0 \le x < 1 \\ \frac{4}{8}, & \text{if } 1 \le x < 2 \\ \frac{7}{8}, & \text{if } 2 \le x < 3 \\ 1, & \text{if } 3 \le x < \infty \end{cases}$$

with the graphical representation

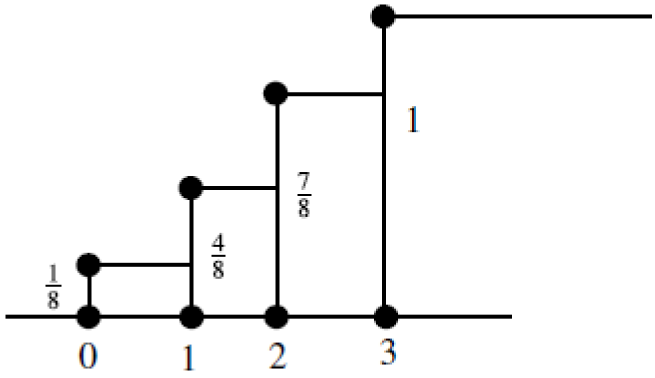


Figure 2.4: The probability distribution function for the number of heads in an experiment of tossing a fair coin three times.

Example 2.9. *Consider rolling one fair six-sided die, so that* $\Omega = \{1,2,3,4,5,6\}$*, with* $P(\omega) = \frac{1}{6}$ *for each* $\omega \in \Omega$*. Let* X *be the number showing on the die divided by 6, so that* $X(\omega) = \frac{\omega}{6}$ *for* $\omega \in \Omega$*.*

- *What is* $F_X(x)$*?*

Since $X(\omega) \le x$ *if and only if* $\omega \le 6x$ *we have that*

$$F_X(x) = P(X \le x) = \sum_{\omega\in\Omega,\omega\le 6x} p(\omega) = \sum_{\omega\in\Omega,\omega\le 6x} \frac{1}{6} = \frac{1}{6}|\omega \in \Omega, \omega \le 6x|.$$

That is, to compute $F_X(x)$*, we count how many elements* $\omega \in \Omega$ *satisfy* $\omega \le 6x$ *and multiply that number by* $\frac{1}{6}$*. Therefore,*

$$F_X(x) = \begin{cases} 0, & \text{if} \quad x < \frac{1}{6} \\ \frac{1}{6}, & \text{if} \quad \frac{1}{6} \le x < \frac{2}{6} \\ \frac{2}{6}, & \text{if} \quad \frac{2}{6} \le x < \frac{3}{6} \\ \frac{3}{6}, & \text{if} \quad \frac{3}{6} \le x < \frac{4}{6} \\ \frac{4}{6}, & \text{if} \quad \frac{4}{6} \le x < \frac{5}{6} \\ \frac{5}{6}, & \text{if} \quad \frac{5}{6} \le x < 1 \\ 1, & \text{if} \quad 1 \le x. \end{cases}$$

In Figure 2.5 we present a graph of the function $F_X(x)$ *and note that this is a step function.*

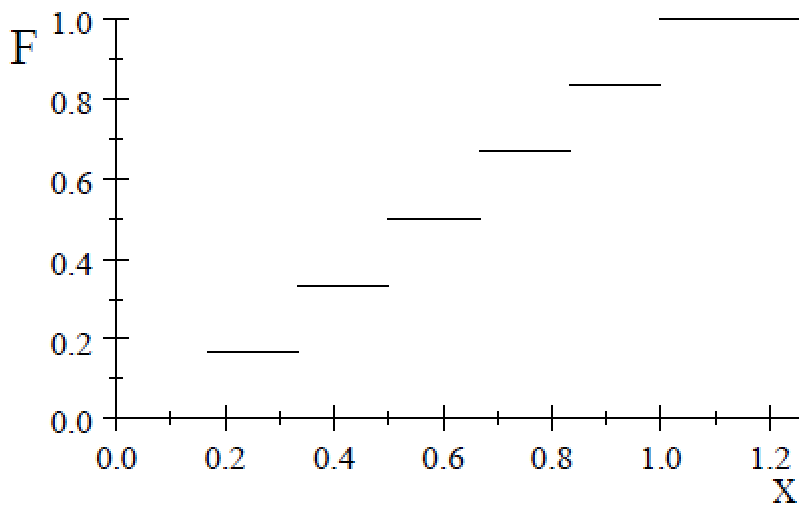


Figure 2.5: The cumulative distribution function of discrete random variable X: rolling one fair six-sided die.

Example 2.10. *We throw two separate dice at the same time and we are interested in the sum of the points. We denote X as this random variable, it is defined by*

$$X : \Omega \longrightarrow \mathbb{R}$$
$$(\omega_1, \omega_2) \longmapsto \omega_1 + \omega_2$$

with $\Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}$ *as shown in the table below*

D_1 \ D_2	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Table 2.3: The outcomes of the experiment: throwing two separate dice at the same time.

In this way, every outcome in the sample space is assigned a real value: $X(\Omega) = \{2, 3, \ldots, 12\}$

Sample space elements	Value of random variable
$(1,1)$	2
$(1,2),(2,1)$	3
$(1,3),(2,2),(3,1)$	4
$\vdots$	$\vdots$
$(6,6)$	12

Table 2.4: The samples space elements with value of random variable.

$$
\begin{aligned}
P(X=2) = P(X=12) &= \frac{1}{36},\\
P(X=3) = P(X=11) &= \frac{2}{36},\\
P(X=4) = P(X=10) &= \frac{3}{36},\\
P(X=5) = P(X=9) &= \frac{4}{36},\\
P(X=6) = P(X=8) &= \frac{5}{36},\\
P(X=7) &= \frac{6}{36}.
\end{aligned}
$$

X *has a value in* $\{2, 3, \ldots, 12\}$*, then* $P(X = k) = 0$ *for* $k \notin \{2, 3, \ldots, 12\}$*, we will tabulate this with the cumulative distribution function in the following table:*

k	2	3	4	5	6	7	8
$P(X=k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36
$F_X(k)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36

9	10	11	12
4/36	3/36	2/36	1/36
30/36	33/36	35/36	1

Table 2.5: The probability and cumulative distribution functions for the sum of the values for two rolls of a die.

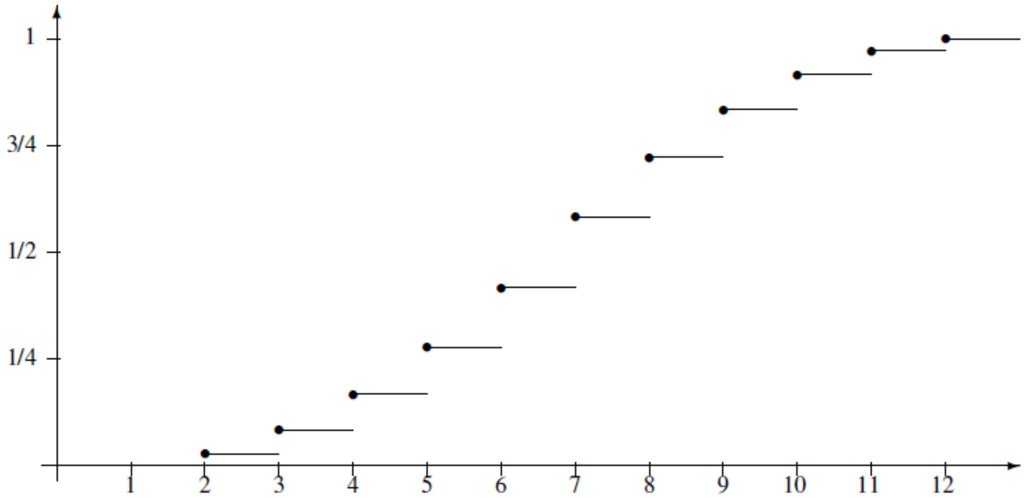


Figure 2.6: The cumulative distribution function for the sum of the values for two rolls of a die.

If we look at the graph of this cumulative distribution function, we see that it is constant in between the possible values for X *and that the jump size at* x *is equal to* $P(X = x)$. *In this example,* $P(X = 5) = 4/36$, *the size of the jump at* $x = 5$. *In addition,*

$$
\begin{aligned}
F_X(5) - F_X(2) &= P\{2 < X \leq 5\} \\
&= P(X = 3) + P(X = 4) + P(X = 5) \\
&= \frac{2}{36} + \frac{3}{36} + \frac{4}{36} \\
&= \frac{9}{36}.
\end{aligned}
$$

Example 2.11. *Two dice are thrown. Let* Y *assign to each point* (ω_1, ω_2) *in* Ω *the maximum number of its number*

$$
\begin{aligned}
Y : \Omega &\longrightarrow \mathbb{R} \\
(\omega_1, \omega_2) &\longmapsto \max(\omega_1, \omega_2).
\end{aligned}
$$

with $\Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\}$. *Find the probability distribution of the random variable*

Y with $Y(\Omega) = \{1, 2, ..., 6\}$

$$P(Y=1) = \frac{1}{36}, P(Y=2) = \frac{3}{36},$$
$$P(Y=3) = \frac{5}{36}, P(Y=4) = \frac{7}{36},$$
$$P(Y=5) = \frac{9}{36}, P(Y=6) = \frac{11}{36}.$$

Y has a value in $\{1, 2, ..., 6\}$, *then* $P(Y = k) = 0$ *for* $k \notin \{1, 2, ..., 6\}$; *We will tabulate this with the cumulative distribution function as*

k	1	2	3	4	5	6
$P(Y=k)$	1/36	3/36	5/36	7/36	9/36	11/36
$F_Y(k)$	1/36	4/36	9/36	16/36	25/36	1

Table 2.6: The probability and cumulative distribution functions for the max of the values for two rolls of a die.

2.2.4 Expectation

In the previous sections, we learned about probability models, random variables, and distributions. There is one more concept that is fundamental to all of probability theory, that of expected value.

The expected value of a discrete random variable X, symbolized as $E(X)$[2], is often referred to as the long-term average or mean (symbolized as μ_X). This means that over the long term of doing an experiment over and over, you would expect this average. For example, let X = the number of heads you get when you toss three fair coins. If you repeat this experiment (toss three fair coins) a large number of times, the expected value of X is the number of heads you expect to get for each of the three tosses on average.

We begin with a definition.

Definition 2.10. *Let X be a discrete random variable with set of possible values* $X(\Omega)$. *The*

[2] $E(X)$ is the whole point for monetary games of chance e.g., lotteries, blackjack, slot machines.
If X =your payoff, the operators of these games make sure $E(X) < 0$. Thorp's card-counting strategy in blackjack changed $E(X) < 0$ (because ties went to the dealer) to $E(X) > 0$ to the dismay of the casinos. See "How to Beat the Dealer" by Edward Thorp (a math professor at UCIrvine).

expected value or mean value of X denote $E(X)$ or μ_X is defined by

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x).$$

Hence, an equivalent definition is the following.

Definition 2.11. *The expectation, expected value, or mean of a discrete random variable X, taking on distinct values $x_1, x_2, ..., x_n$, with probability mas function $P(X = x_i) = p_i$, is given by*

$$E(X) = \sum_{i=1}^{n} x_i p_i.$$

The definition (in either form) is best understood through examples.

Example 2.12. *The expected value for the example "flip a fair coin three times" is:*

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i P(X = x_i) \\ &= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} \\ &= \frac{3}{2}. \end{aligned}$$

The most probable values were 1 and 2 (tied) each with probability $\frac{3}{8}$. We see that the expected value is not the most probable value, so $\frac{3}{2}$ was not even a possible value $P\left(X = \frac{3}{2}\right) = 0$.

Example 2.13. *The expected value for the example "rolling of a die" is:*

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} x_i P(X = x_i) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \\ &= \frac{7}{2}. \end{aligned}$$

Example 2.14. *Suppose that $P(Y = -3) = 2$, and $P(Y = 11) = \frac{7}{10}$, and $P(Y = 31) = \frac{1}{10}$. Then*

$$\begin{aligned} E(Y) &= -3 \times 2 + 11 \times \frac{7}{10} + 31 \times \frac{1}{10} \\ &= \frac{102}{10}. \end{aligned}$$

Example 2.15. *Suppose that $P(Z=-3)=2$, and $P(Z=-11)=\frac{7}{10}$, and $P(Z=31)=\frac{1}{10}$. Then*

$$\begin{aligned} E(Z) &= -3\times 2+(-11)\times\frac{7}{10}+31\times\frac{1}{10} \\ &= -\frac{52}{10}. \end{aligned}$$

In this case, the expected value of Z is negative.

We thus see that, for a discrete random variable X, once we know the probabilities that $X=x$ (or equivalently, once we know the probability function P_X), it is straightforward (at least in simple cases) to compute the expected value of X.

Remark 2.10. *The expected value exists if*

$$\sum_x |x|\,p(x) < \infty.$$

Remark 2.11. *If the random variable X takes a countable number of values $x_1,\dots,x_n,\dots$, its mathematical expectation is then defined by $E(X)=\sum_{i=1}^{\infty} x_i P(X=x_i)$, such that the series converges absolutely.*

Example 2.16. *Let Y be a discrete random variable, with probability function P_Y given by*

$$P_Y(y)=\begin{cases} \frac{1}{2y} & \text{if} \quad y=2,4,8,16,\dots \\ \frac{1}{2|y|} & \text{if} \quad y=-2,-4,-8,-16,\dots \\ 0 & \text{otherwise} \end{cases}$$

That is

$$\begin{aligned} P_Y(2) &= P_Y(-2)=\frac{1}{4}, \\ P_Y(4) &= P_Y(-4)=\frac{1}{8}, \\ P_Y(8) &= P_Y(-8)=\frac{1}{16}, \\ &\quad etc \qquad . \end{aligned}$$

Then it is easily checked that P_Y is indeed a valid probability function (i.e., $P_Y(y)\geq 0$ for all y, with $\sum_y P_Y(y)=1$).

On the other hand, we compute that

$$\begin{aligned} E(Y) &= \sum_{y} P_Y(y) \\ &= \sum_{k=1}^{\infty}\left(2^k\right)\left(\frac{1}{2.2^k}\right)+\sum_{k=1}^{\infty}\left(-2^k\right)\left(\frac{1}{2.2^k}\right) \\ &= \sum_{k=1}^{\infty}\frac{1}{2}-\sum_{k=1}^{\infty}\frac{1}{2} \\ &= \infty-\infty, \end{aligned}$$

which is undefined. We therefore say that $E(Y)$ does not exist, i.e., that the expected value of Y is undefined in this case.

Remark 2.12. *Let X be a discrete a random variable.*

1. *The expectation of X is defined by a sum that can be infinite, which is why it belongs to $\mathbb{R}\cup\{-\infty,+\infty\}$ in general.*
2. *When $E(X)=0$, we say that the random variable X is centered.*
3. *If $E(|X|)<\infty$, we say that the discrete random variable is integrable.*
4. *If $E(|X|)=0$, then $P(X=0)=1$.*
5. *If $X(\Omega)$ is a finite set, then it is not difficult to show that*

$$\min_{\omega\in\Omega} X(\omega) \leq E(X) \leq \max_{\omega\in\Omega} X(\omega).$$

The most important properties of expected value are in following theorem.

Theorem 2.3. *Let X and Y be discrete random variables, let a and b be real numbers, we have*

1. *$E(a)=a$.*
2. *Linearity: $E(aX+bY)=aE(X)+bE(Y)$.*
3. *Monotonicity: If $X\leq Y$ (i.e, $X(\omega)\leq Y(\omega)$ for $\omega\in\Omega$), then $E(X)\leq E(Y)$.*

Proof. We have

1.
$$E(a) = \sum_{i=1}^{n} aP(X = x_i) = a\sum_{i=1}^{n} P(X = x_i) = a \times 1 = a.$$

2. Let $P_{X,Y}$ be the joint probability function of X and Y. Then using the above theorem, we have

$$\begin{aligned}
E(aX + bY) &= \sum_{x,y} (ax + by) P_{X,Y}(x,y) \\
&= a\sum_{x,y} xP_{X,Y}(x,y) + b\sum_{x,y} yP_{X,Y}(x,y) \\
&= a\sum_{x} x \sum_{y} P_{X,Y}(x,y) + b\sum_{y} y \sum_{x} P_{X,Y}(x,y).
\end{aligned}$$

   Because $\sum_y P_{X,Y}(x,y) = P_X(x)$ and $\sum_x P_{X,Y}(x,y) = P_Y(y)$, we have that

$$\begin{aligned}
E(aX + bY) &= a\sum_{x} xP_X(x) + b\sum_{y} yP_Y(y) \\
&= aE(X) + bE(Y).
\end{aligned}$$

3. Let $Z = Y - X$. Then Z is also dicscrete. Furthermore, because $X \le Y$, we have $Z \ge 0$, so that all possible values of Z are nonnegative. Hence, if we list the

$$E(Z) = \sum_{i} z_i P(Z = z_i) \ge 0.$$

   But by linearity, $E(Z) = E(Y) - E(X)$. Hence, $E(Y) - E(X) \ge 0$, so that $E(Y) \ge E(X)$.

□

Proposition 2.4. *Let X be a discrete variable, p_X its mass function and $g : \mathbb{R} \to \mathbb{R}$ a continuous function. Let $Y = g(X)$, then Y is again a discrete random variable and*

$$E(Y) = E(g(X)) = \sum_{k \in X(\Omega)} g(k) P(X = k).$$

Remark 2.13. *In general, it is not true that if $\phi(z)$ is a function (e.g.,$\phi(z) = z^2$), $E(\phi(X)) = \phi(E(X))$.*

2.2.4.1 Moments

Definition 2.12. *Let X be a discrete random variable with set of possible values $X(\Omega) = \{x_1, x_2, ..., x_n\}$. Then for all $r \in \mathbb{N}$, we have*

1. *The $r-$th moment of a random variable X is defined to be $E[X^r]$ and given by*

$$E[X^r] = \sum_{i=1}^{n} x_i^r P(X = x_i).$$

2. *The r-th central moment of X is defined to be $E[(X - E(X))^r]$ and given by*

$$E[(X - E(X))^r] = \sum_{i=1}^{n} (x_i - E(X))^r P(X = x_i).$$

Remark 2.14. *For example, the first moment is the expected value $E[X]$. The second central moment is the variance of X. Similar to mean and variance, other moments give useful information about random variables.*

2.2.5 Variance

The expected value does not tell you everything you went to know about a random variable. So, Two random variables can have equal means but very different patterns of variability. Here is a sketch of the probability functions $p_1(x)$ and $p_2(x)$ of two random variables X_1 and X_2

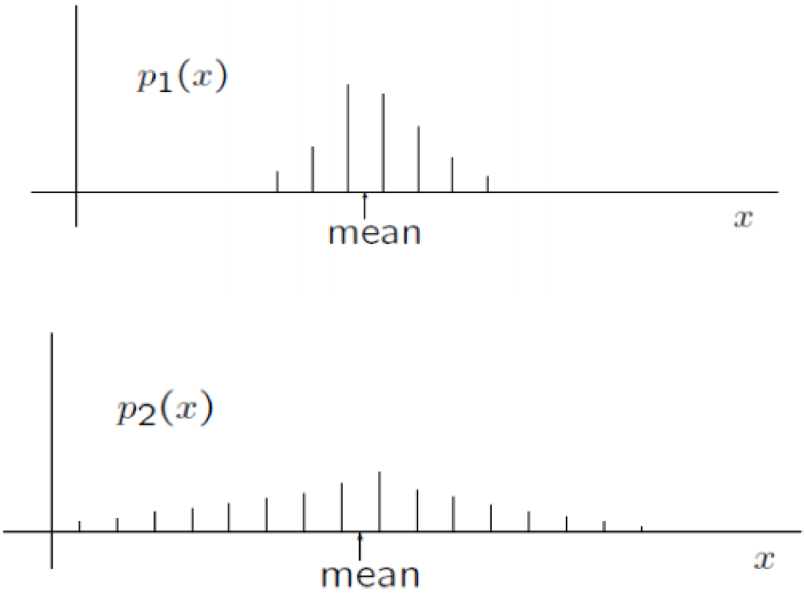


Figure 2.7: The probability functions of two random variables with the same mean.

To distinguish between these, we need a measure of spread or dispersion. There are many possible measures, We look briefly at three plausible ones:

1. Mean difference: $E\{X - E(X)\}$, is attractive superficially, but no use.

2. Mean absolute difference: $E\{|X - E(X)|\}$, is hard to manipulate mathematically.

3. Variance: $E\{X - E(X)\}^2$, is the most frequently-used measure.

Example 2.17. *Suppose you and a friend play the following game of change. Flip a coin. If a head comes up you get 10 DA. If a tail comes up you pay your friend 10 DA. So if X = your payoff*

$$X(H) = +10, X(T) = -10,$$
$$E(X) = (+10) \times \frac{1}{2} + (-10) \times \frac{1}{2} = 0$$

so this is a fair game. Now suppose you play the game changing 10 DA to 200 DA. It is still a fair game

$$E(X) = (+200) \times \frac{1}{2} + (-200) \times \frac{1}{2} = 0$$

but I personally would be very reluctant to play this game. The notion of variance is designed to capture the difference between the two games.

Definition 2.13. *Let X be a discrete random variable with set of possible values $X(\Omega)$ and expected value μ_X (or $E(X)$). Then the variance of X, denoted $Var(X)$ or σ^2 is defined by*

$$\sigma^2 = Var(X) = E\left[(X - E(X))^2\right] = \sum_{x \in X(\Omega)} (x - \mu_X)^2 P(X = x).$$

The standard deviation of X is defined to be the square-root of the variance

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

Intuitively, the variance $Var(X)$ is a measure of how spread out the distribution of X is, or how random X is, or how much X varies, as the following example illustrates.

Example 2.18. *Let X and Y be two discrete random variables, with probability functions*

$$P_X(y) = \begin{cases} 1 & \text{if } x = 10 \\ 0 & \text{otherwise} \end{cases}$$

and

$$P_Y(y) = \begin{cases} \frac{1}{2} & \text{if } y = 5 \\ \frac{1}{2} & \text{if } y = 15 \\ 0 & \text{otherwise} \end{cases}$$

respectively.
Then $E(X) = E(Y) = 10$. However,

$$Var(X) = (10 - 10)^1 (1) = 0,$$

while

$$Var(Y) = (5 - 10)^2 \left(\frac{1}{2}\right) + (15 - 10)^2 \left(\frac{1}{2}\right) = 25.$$

We thus see that, while X and Y have the same expected value, the variance of Y is much greater than that of X. This corresponds to the fact that Y is more random than X; that is, it varies more than X does.

Example 2.19. *Find the variance of the discrete random variable X whose probability distribution is as shown in table*

x	1	2	3	4	5
$P(X = x)$	0.1	0.1	0.3	0.3	0.2

Solution: *We compute the expected value of* X *as:*

$$\begin{aligned} E(X) &= \sum_{i=1}^{5} x_i P(X = x_i) \\ &= 1 \times 0.1 + 2 \times 0.1 + 3 \times 0.3 + 4 \times 0.3 + 5 \times 0.2 \\ &= 3.4. \end{aligned}$$

We compute the variance by then computing the $E\left(X^2\right)$ *as follows*

$$\begin{aligned} E\left(X^2\right) &= \sum_{i=1}^{5} x_i^2 P(X = x_i) \\ &= (1)^2 \times 0.1 + (2)^2 \times 0.1 + (3)^2 \times 0.3 + (4)^2 \times 0.3 + (5)^2 \times 0.2 \\ &= 0.1 + 0.4 + +2.7 + 4.8 + 5 \\ &= 13. \end{aligned}$$

We can the compute the variance using the formula

$$Var(X) = E\left(X^2\right) - (E(X))^2,$$

thus

$$Var(X) = 13 - (3.4)^2 = 1.44.$$

The important properties of the variance is presented in the following theorem:

Theorem 2.4. *Let* X *be any discrete random variable, with expected value* $\mu_X = E(X)$*, and variance* $Var(X)$*. Then the following hold true:*

1. $Var(X) \geq 0.$
2. *If* c *is a constant,* $Var(c) = 0.$
3. *If* a *and* b *are real numbers,* $Var(aX + b) = a^2 Var(X).$
4. $Var(X) = E\left(X^2\right) - (E(X))^2.$
5. $Var(X) \leq E\left(X^2\right).$

Proof.

1. This is immediate, because we always have $(X - \mu_X)^2 \geq 0.$

2. From the definition of variance we have,

$$\begin{aligned} Var(c) &= E(c - E(c))^2 \\ &= E(c-c)^2 \\ &= E(0) \\ &= 0. \end{aligned}$$

3. We note that

$$\mu_{aX+b} = E(aX+b) = aE(X) + b = a\mu_X + b,$$

by linearity. Hence, again using linearity,

$$\begin{aligned} Var(aX+b) &= E\left((aX+b-\mu_{aX+b})^2\right) \\ &= E\left((aX+b-a\mu_X - b)^2\right) \\ &= a^2 E\left((X-\mu_X)^2\right) \\ &= a^2 Var(X). \end{aligned}$$

4. From the linearity property, we have

$$\begin{aligned} Var(X) &= E\left((X-\mu_X)^2\right), \text{ where } \mu = E(X) \\ &= E\left(X^2 - 2X\mu_X + (\mu_X)^2\right) \\ &= E\left(X^2\right) - 2\mu_X E(X) + (\mu_X)^2 \end{aligned}$$

then making the substitution that $E(X) = \mu_X$, we obtain

$$\begin{aligned} Var(X) &= E(X^2) - 2(\mu_X)^2 + (\mu_X)^2 \\ &= E(X^2) - (\mu_X)^2 \\ &= E(X)^2 - (E(X))^2. \end{aligned}$$

5. This follows immediately from part (3) because we have $-(\mu_X)^2 \leq 0$.

□

2.3 Continuous random variables

In Section 2.2 of this chapter, we have defined the discrete random variable as a random variable having countable number of values, i.e. whose values can be arranged in a sequence. But, if a random variable is such that its values cannot be arranged in a sequence, it is called continuous random variable.

2.3.1 Definition

Definition 2.14. *A random variable is called a continous random variable if its set of possible outcomes $X(\Omega)$ is interval or union of intervals of real numbers.*

Example 2.20. *Some examples of continuous random variables:*

1. *Height of students at certain college.*
2. *Mark of a student.*
3. *Temperature of a city at various points of time during a day.*
4. *The volume of waste water treated at a sewage plant per day.*
5. *The weight of hearts in normal rats.*
6. *The lengths of the wings of butterflies.*
7. *The yield of barley from a large paddock.*
8. *The amount of rainfall recorded each year.*
9. *The time taken to perform a psychological test.*
10. *Length of time required to complete a given training.*

The theory of continuous random variables is completely analogous to the theory of discrete random variables. Indeed, Summation and integration have the same meanings but in mathematics there is still difference between the two and that is that the former is used in case of discrete values, i.e. countable values and the latter is used in continuous case. if we want to oversimplify things, we might say the following: take any formula about discrete random variables, and then replace sums with integrals, and replace pmf_s with probability density functions (pdf_s), and you will get the corresponding formula for continuous random variables.

2.3.2 Density function and cumulative distribution of continuous r.v.

Using probability functions to describe the distribution of a continuous random variable is tricky, because probability behaves like mass at points. In the discrete case, mass can be distributed over a number (possibly countably infinite) of distinct points where each

point has non-zero mass. However, in the continuous case, mass cannot be thought of as an attribute of a point but rather of a region surrounding a point. Thus, we need to develop new tools to deal with continuous random variables. So, continuous random variable is represented by different representation known as probability density function unlike the discrete random variable which is represented by probability mass function.
The probability density function (pdf) is a fundamental concept for describing the distribution of continuous random variables. It specifies how the probability is distributed across different values and provides a basis for calculating probabilities over intervals by integrating the pdf. Understanding the nature of the pdf and its continuity properties is crucial for proper modeling and analysis of continuous random variables.
Let $f(x)$ be a continuous function of x . Suppose the shaded region $ABCD$ shown in the following figure represents the area bounded by $y = f(x)$, x -axis and the ordinates at the points x and $x+\delta x$, where δx is the length of the interval $(x, x+\delta x)$.

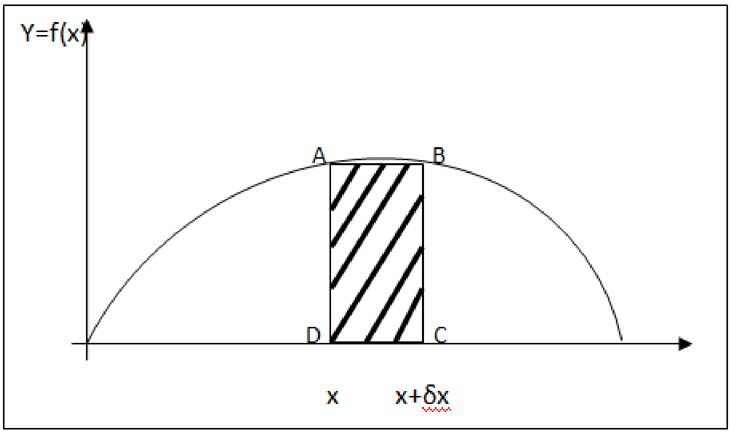


Figure 2.8: The density function of continuous random variable X.

Now, if δx is very-very small, then the curve AB will act as a line and hence the shaded region will be a rectangle whose area will be $AD \times DC$ i.e. $f(x)\delta x$ (Note that: AD = the value of y at x i.e. $f(x)$, DC = length δx of the interval $(x, x+\delta x)$).
Also, this area = probability that X lies in the interval $(x, x+\delta x) = P(x \le X \le x+\delta x)$. Hence

$$P(x \le X \le x+\delta x) = f(x)\delta x$$

which implies that

$$\frac{P(x \le X \le x+\delta x)}{\delta x} = f(x),$$

where δx is very-very small, we have

$$\lim_{\delta x \to 0} \frac{P(x \leq X \leq x + \delta x)}{\delta x} = f(x).$$

$f(x)$, so defined, is called probability density function.
Probability density function has the same properties as that of probability mass function. Indeed,

Definition 2.15. *A continuous random variable X is a real-valued function $X : \Omega \to \mathbb{R}$ for which*

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx \tag{2.1}$$

where f is a function satisfying:

1. $f(x) \geq 0$ *for all* $x \in \mathbb{R}$.
2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

The function f is called the probability density function (pdf).

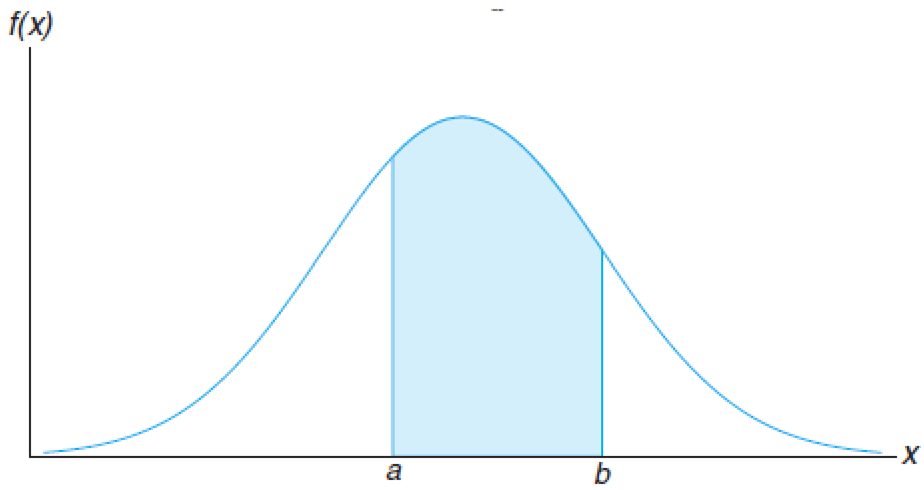


Figure 2.9: Area under the curve of $f(x)$ and over the interval $(a;b)$.

Theorem 2.5. *Let X be an absolutely continuous random variable. Then the probability of specific point in a continuous distribution is zero;*

$$P(X = a) = 0 \text{ for all } a \in \mathbb{R}.$$

Proof. Let a be any real number. Then

$$P(X=a)=P(a\leq X\leq a).$$

On the other hand, setting $a=b$ in (2.1), we see that

$$P(a\leq X\leq a)=\int_a^a f(x)dx=0.$$

Hence, $P(X=a)=0$ for all a, as required. □

Remark 2.15. *For a continous random variable X, we have*

1. $f(x)\neq P(X=x)$ *(in general).*
2. $P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)=P(a\leq X\leq b)$.
3. $P(X\in A)=\int_A f(x)dx$.

Example 2.21.

1. *If* $A=]-\infty,a]: P(X\in A)=P(X\leq a)=\int_{-\infty}^a f(x)dx$.

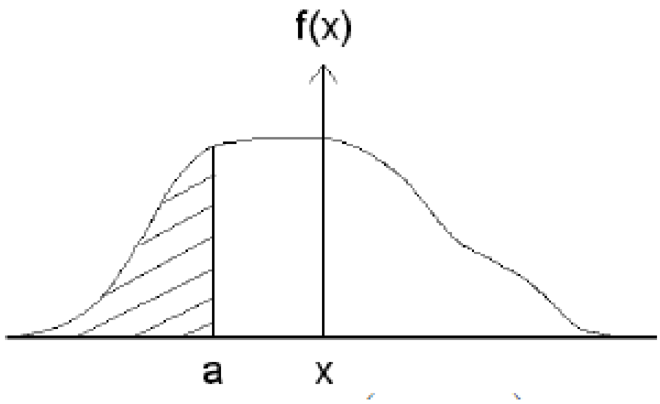


Figure 2.10: The area $P(X\leq a)=\int_{-\infty}^a f(x)dx$.

2. *If* $A=[b,+\infty[: P(X\in A)=P(X\geq b)=\int_b^{+\infty} f(x)dx$.

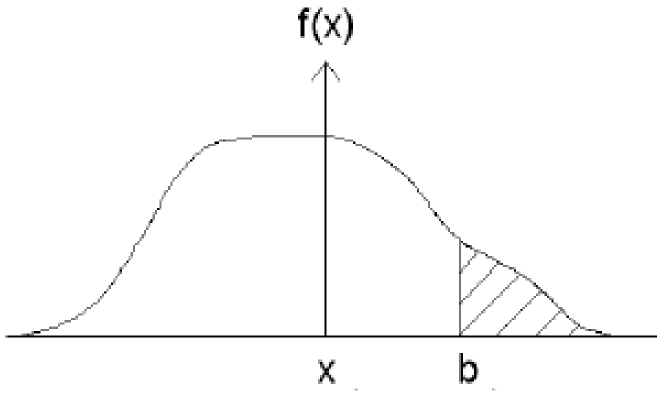


Figure 2.11: The area $P(X \geq b) = \int_b^{+\infty} f(x)dx$.

Remark 2.16.

1. *If the real random variable X follows the density distribution $f(x)$, then the support of X is the set*

$$\Omega_X = \{x \in \mathbb{R} : f(x) > 0\}.$$

2. *Density functions do not need to be bounded.*

Example 2.22. *Let $f(x)$ be the density function for a random variable X given by*

$$f(x) = \begin{cases} 0 & if \quad x \leq 0, \\ \frac{c}{\sqrt{x}} & if \quad 0 < x < 1, \\ 0 & if \quad 1 \leq x. \end{cases}$$

The support of X is the interval $]0, 1[$. Then, to find the value of the constant c, we compute the integral

$$1 = \int_0^1 \frac{c}{\sqrt{t}} dt = 2c\sqrt{t}\Big|_0^1 = 2c,$$

so $c = \frac{1}{2}$.
For $0 \leq a < b \leq 1$,

$$P(a < X \leq b) = \int_a^b \frac{1}{2\sqrt{t}} dt = \sqrt{t}\Big|_a^b = \sqrt{b} - \sqrt{a}.$$

Example 2.23. *A continuous random variable X has the probability density function:*

$$f(x) = \begin{cases} Ax^3 & if \quad 0 \leq x \leq 1 \\ 0 & otherwise. \end{cases}$$

1. *Determine the constant A.*

2. *Calculate* $P(0.2 \le X \le 0.5)$ *and* $P\left(X > \frac{3}{4} \text{ given } X > \frac{1}{2}\right)$.

Solution:

1. *As* $f(x)$ *is probability density function,* $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$, *we have*

$$\int_0^1 f_X(x)\,dx = 1$$
$$\int_0^1 Ax^3 dx = 1$$

which implies that

$$A\left[\frac{x^4}{4}\right]_0^1 = 1$$
$$A\left(\frac{1}{4} - 0\right) = 1$$
$$A = 4.$$

2.

$$\begin{aligned} P(0.2 \le X \le 0.5) &= \int_{0.2}^{0.5} f(x)dx \\ &= \int_{0.2}^{0.5} Ax^3 dx \\ &= 4\left[\frac{x^4}{4}\right]_{0.2}^{0.5} \\ &= (0.5)^4 - (0.2)^4 \\ &= 0.0625 - 0.0016 \\ &= 0.0609. \end{aligned}$$

and

$$\begin{aligned}
P\left(X > \frac{3}{4} \text{ given } X > \frac{1}{2}\right) = P\left(X > \frac{3}{4} | X > \frac{1}{2}\right) &= \frac{P\left(X > \frac{3}{4} \cap X > \frac{1}{2}\right)}{P\left(X > \frac{1}{2}\right)} \\
&= \frac{P\left(X > \frac{3}{4}\right)}{P\left(X > \frac{1}{2}\right)} = \frac{\int_{\frac{3}{4}}^{1} f(x)dx}{\int_{\frac{1}{2}}^{1} f(x)dx} \\
&= \frac{\int_{\frac{3}{4}}^{1} 4x^3 dx}{\int_{\frac{1}{2}}^{1} 4x^3 dx} = \frac{\left[\frac{x^4}{4}\right]_{\frac{3}{4}}^{1}}{\left[\frac{x^4}{4}\right]_{\frac{1}{2}}^{1}} \\
&= \frac{175}{256} \times \frac{16}{15} = \frac{35}{48}.
\end{aligned}$$

Definition 2.16. *The cumulative distribution function (cdf) F_X of a continuous random variable X with probability density function $f(x)$ is given by*

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \text{ for all } x \in \mathbb{R}.$$

(Note the use of the "dummy variable" t in this integral).

Definition 2.17. *A random variable X is called absolutely continuous if its cumulative distribution function (cdf), $F_X(x)$, is absolutely continuous. This means that $F_X(x)$ has a derivative almost everywhere, and this derivative is the probability density function (pdf) $f(x)$ of X.*

By the fundamental theorem of calculus, the density function is the derivative of the distribution function

$$f(x) = \lim_{\Delta x \to 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = F'_X(x).$$

In other words,

$$F_X(x + \Delta x) - F_X(x) \approx f(x)\Delta x.$$

We see that it is also possible to compute a density $f(x)$ once we know the cumulative distribution function $F_X(x)$.

Corollary 2.1. *Let X be a continuous random variable with distribution function $F_X(x)$. The probability density function (pdf) of X is defined as the derivative of the distribution function:*

$$f(x) = F'_X(x) = \frac{\partial}{\partial x} F_X(x).$$

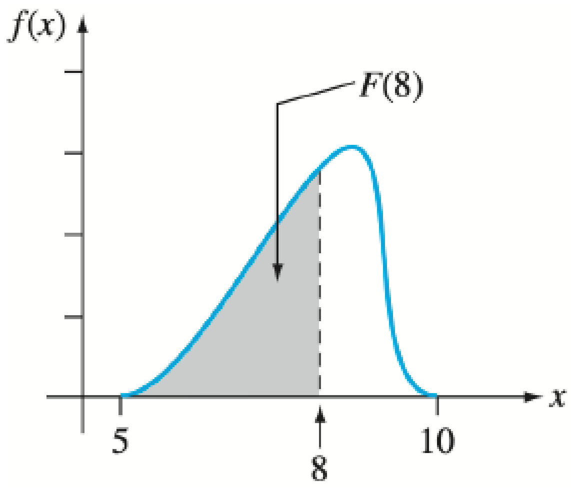


Figure 2.12: The probability density function: example in the value $x = 8$.

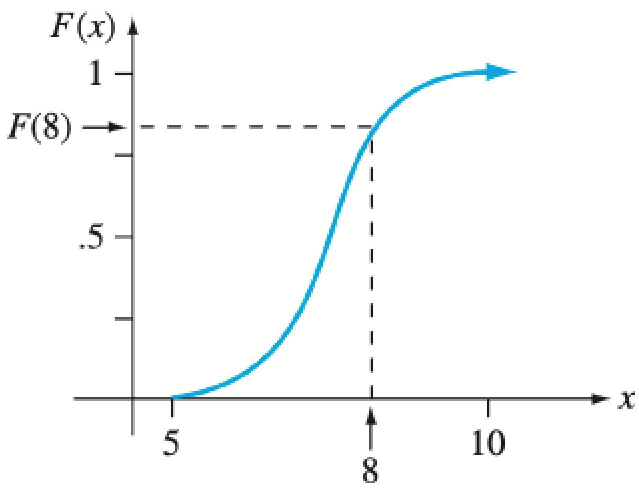


Figure 2.13: The distribution function: example in the value $x = 8$.

Note that the probability density function is highest where the slope of the distribution function is greatest.

Remark 2.17.

1. *A random variable is continuous if and only if its cdf is an everywhere continuous function.*

2. *If a cdf $F_X(x)$ is absolutely continuous, then its derivative, the pdf $f(x)$, exists almost everywhere and is piecewise continuous. However, the pdf $f(x)$ may have jump discontinuities at certain points where $F_X(x)$ has jumps, which are often finite and lead to the*

piecewise continuity of the function. These discontinuities are not problematic as long as the total integral of the pdf remains 1*, ensuring that it remains a valid density function.*

3. *Piecewise continuity of the pdf is significant in practical scenarios. It allows for the modeling of a wide variety of continuous random variables, even if the pdf is not continuous everywhere. Piecewise continuity is often sufficient for theoretical analysis and practical applications.*

Proposition 2.5. *Let X be an absolute continuous random variable with density function $f(x)$. The cumulative distribution function (cdf) of X is the function $F_X : \mathbb{R} \to \mathbb{R}$ satisfying the following properties:*

- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$ *for all* $x \in \mathbb{R}$.
- $\frac{\partial}{\partial x} F_X(x) = f_X(x)$.
- $P(a \leq X \leq b) = F_X(b) - F_X(a)$.
- F_X *is monotone increasing, since* $f_X \geq 0$. *That is,* $F_X(c) \leq F_X(d)$ *for* $c \leq d$.
- $\lim_{x \to -\infty} F_X(x) = P(X \leq -\infty) = 0$.
- $\lim_{x \to +\infty} F_X(x) = P(X \leq +\infty) = 1$.

Proposition 2.6. *Let X be an absolute continuous random, then*

$$\forall x \in \mathbb{R}, P(X = x) = 0.$$

Proof. Let $\epsilon > 0, x \in R$ and let F_X be the distribution function of X, $f(x)$ its density function. Let us first note that

$$P(X = x) = \lim_{\epsilon \to 0} P(x - \epsilon < X \leq x + \epsilon).$$

From Proposition 2.1, we have

$$P(x - \epsilon < X \leq x + \epsilon) = F_X(x + \epsilon) - F_X(x - \epsilon),$$

thus

$$\begin{aligned} P(X = x) &= \lim_{\epsilon \to 0} [F_X(x + \epsilon) - F_X(x - \epsilon)] \\ &= \lim_{\epsilon \to 0} \int_{x-\epsilon}^{x+\epsilon} f(t)dt \\ &= 0. \end{aligned}$$

□

We obtain as an immediate consequence of this proposition:

Corollary 2.2. *for an absolutely continuous random variable X, for all $a, b \in \mathbb{R}$ with $a < b$, we have*

$$\begin{aligned} F_X(b) - F_X(a) &= P(a < X < b) = P(a \le X < b) \\ &= P(a < X \le b) = P(a \le X \le b). \end{aligned}$$

Example 2.24. *The diameter X of a cable is assumed to be a continuous random variable with probability density function (pdf)*

$$f(x) = \begin{cases} 6x(1-x) & \text{if} \quad 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

- *obtain the cumulative distribution function of X.*

Solution: *We have,*

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

1. *If $x < 0$;*

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0dt = 0.$$

2. *If $0 \le x \le 1$;*

$$\begin{aligned} F_X(x) = P(X \le x) &= \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} 6t(1-t)dt \\ &= 6\int_{0}^{x} \left(t - t^2\right)dt = 6\left[\frac{t^2}{2} - \frac{t^3}{3}\right]_0^x = 3x^2 - 2x^3. \end{aligned}$$

3. *If $x > 1$;*

$$\begin{aligned} F_X(x) = P(X \le x) &= \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{1} 6t(1-t)dt + \int_{1}^{x} 0dt \\ &= 6\int_{0}^{1} \left(t - t^2\right)dt = 6\left[\frac{t^2}{2} - \frac{t^3}{3}\right]_0^1 = 1. \end{aligned}$$

Thus

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 3x^2 - 2x^3, & 0 \le x \le 1 \\ 1, & x > 1. \end{cases}$$

Example 2.25. *Suppose that the error in the reaction temperature in C°, for a controlled laboratory experiment is a continuous random variable X having the following probability density function:*

$$f(x) = \begin{cases} \frac{1}{3}x^2, & -1 < x < 2 \\ 0, & elsewhere \end{cases}$$

1. *Verify that $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.*
2. *Find $P(0 < X \leq 1)$.*
3. *Find the cumulative distribution function.*
4. *Using the cdf, find $P(0 < X \leq 1)$.*

Solution: *Let X the error in the reaction temperature in C° ; X is a continuous random variable.*

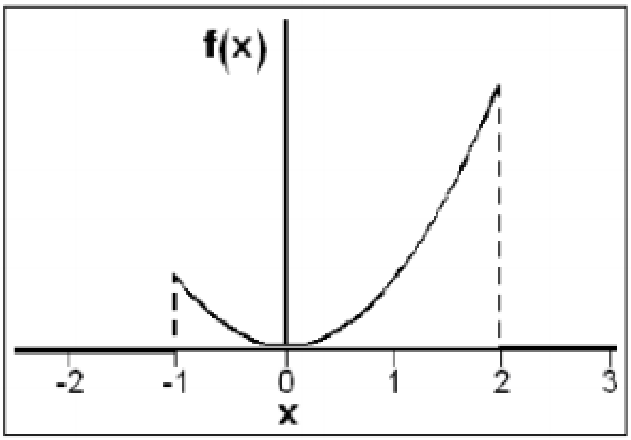


Figure 2.14: The density function of continuous r.v.: the error in the reaction temperature in C°.

1. *$f(x) \geq 0$ because $f(x)$ is a quadratic function.*

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)\,dx &= \int_{-\infty}^{-1} 0dx + \int_{-1}^{2} \frac{1}{3}x^2 dx + \int_{2}^{+\infty} 0dx \\ &= \int_{-1}^{2} \frac{1}{3}x^2 dx = \frac{1}{9}x^3\Big|_{-1}^{2} \\ &= \frac{1}{9}(8-(-1)) = 1.\end{aligned}$$

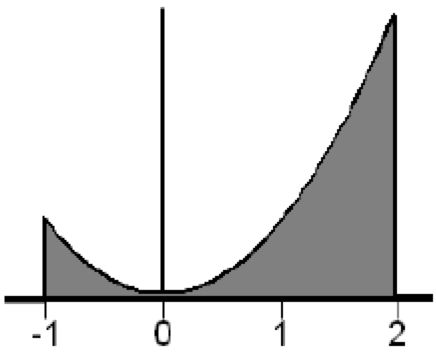


Figure 2.15: The area of $\int_{-\infty}^{+\infty} f(x)dx$.

2.

$$P(0 < X \leq 1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{1}{3}x^2 dx$$
$$= \frac{1}{9}x^3\Big|_0^1 = \frac{1}{9}.$$

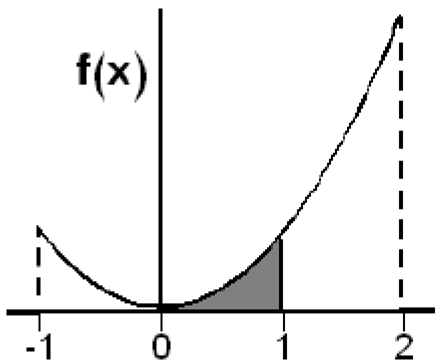


Figure 2.16: The area $P(0 < X \leq 1) = \int_0^1 f(x)dx$.

3. For $x < -1$:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0.$$

For $-1 \le x < 2$:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{x} \frac{1}{3}t^2\,dt$$
$$= \frac{1}{9}t^3\Big|_{-1}^{x} = \frac{1}{9}\left(x^3+1\right).$$

For $x \ge 2$:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{2} \frac{1}{3}t^2\,dt + \int_{2}^{x} 0\,dt$$
$$= \frac{1}{9}t^3\Big|_{-1}^{2} = 1.$$

Therfore, the cdf is:

$$F(x) = P(X \le x) = \begin{cases} 0, & x < -1 \\ \frac{1}{9}\left(x^3+1\right), & -1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

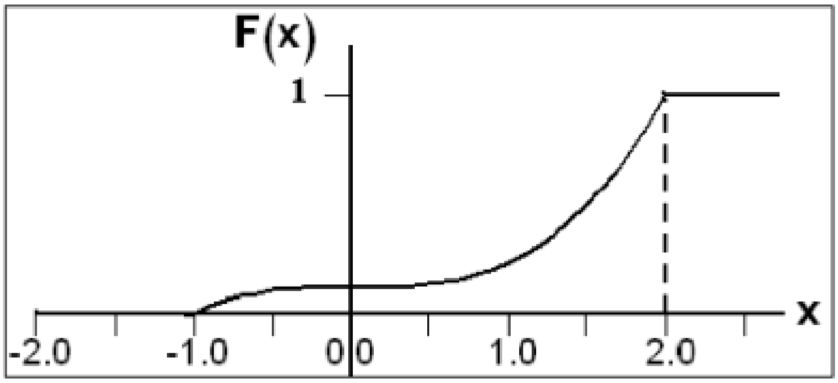


Figure 2.17: The cumulative distribution function of continuous r.v.: the error in the reaction temperature in $C°$.

4. *Using the cdf,*

$$P(0 < X \le 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

2.3.3 Expectation

Definition 2.18. *The expectation, expected value, or mean of a continuous random variable X with probability density function $f(x)$ is given by*

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

Example 2.26. *Let X be a continuous random variable of density function (uniform density) defined by*

$$f(x) = \frac{1}{b-a} \text{ for } x \in [a, b].$$

The mean of X is

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx. \\ &= \int_a^b x \frac{1}{b-a} dx. \\ &= \frac{1}{2(b-a)} x^2 \Big|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{1}{2}(a+b). \end{aligned}$$

Example 2.27. *Let X have density function given by*

$$f(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_1^{+\infty} x . \frac{1}{x^2} dx = \int_1^{+\infty} \frac{1}{x} dx = \log(x)|_1^{+\infty} = \infty.$$

Hence, the expected value of X is infinite.

Similarly to the properties of the expectation of a discrete r.v. we have the following properties of the expectation of continuous random variable:

Theorem 2.6. *Let X and Y be jointly absolutely continuous random variables, and let a and b be real numbers, we have*

1. *Constants: $E(a) = a$.*

2. *Linearity:* $E(aX+bY)=aE(X)+bE(Y)$.

3. *Monotonicity: If* $X \leq Y$, *then* $E(X) \leq E(Y)$.

Proof. We have

1. $E(a)=\int_{-\infty}^{+\infty} af(x)dx = a\int_{-\infty}^{+\infty} f(x)dx = a$.

2. Let $f_{X,Y}$ be the joint probability function of X and Y. Then using the above theorem, we have

$$\begin{aligned}
E(aX+bY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(ax+by)f_{X,Y}(x,y)dxdy \\
&= a\int_{-\infty}^{+\infty} xf_{X,Y}(x,y)dxdy + b\int_{-\infty}^{+\infty} yf_{X,Y}(x,y)dxdy \\
&= a\int_{-\infty}^{+\infty} x\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy\right)dx + b\int_{-\infty}^{+\infty} y\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx\right)dy.
\end{aligned}$$

Because $\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = f_X(x)$ and $\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx = f_Y(y)$, so

$$\begin{aligned}
E(aX+bY) &= a\int_{-\infty}^{+\infty} xf_X(x)dx + b\int_{-\infty}^{+\infty} yf_Y(y)dy \\
&= aE(X)+bE(Y).
\end{aligned}$$

□

Proposition 2.4 remains true in the continuous case, as follows.

Definition 2.19. *Let* X *be an absolute continuous variable, with density function* f_X, *and let* $g:\mathbb{R}\to\mathbb{R}$ *be some function. Then when the expectation of* $g(X)$,

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f_X(x)dx.$$

Proposition 2.7. *Let* X *be an absolute continuous random variable, with density function* f_X, *and let* $g:\mathbb{R}\to\mathbb{R}$ *be continous and positive function. Then we have*

$$E(g(X)) \geq 0.$$

Proof. We have $g(X)\geq 0$, then

$$E(g(X)) = \int_{-\infty}^{+\infty} g(X)f(x)dx \geq 0, \text{ the integral of positive function.}$$

□

Proposition 2.8. *If X is a constant random variable and g is any deterministic function. Then we have*

$$E(g(X)) = g(X).$$

Proof. If $X = c$ is a discret random variable such that $P(X = c) = 1$, then

$$E(g(X)) = \int_{-\infty}^{+\infty} g(c) f(x)dx = g(c) \int_{-\infty}^{+\infty} f(x)dx = g(c) \times 1 = g(X).$$

□

Proposition 2.9. *Let X be an absolute continuous variable, with density function f_X, and $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ two continuous function. Then we have*

1. *Linearity: $E(a(g_1(X)) + b(g_2(X))) = aE(g_1(X)) + bEg_1(X)$ where a, b two real.*
2. *Monotonicity: If $g_1(X) \leq g_2(X)$, then $E(g_1(X)) \leq E(g_2(X))$.*

Proof.

1.

$$\begin{aligned} E(a(g_1(X)) + b(g_2(X))) &= \int_{-\infty}^{+\infty} (a(g_1(X)) + b(g_2(X))) f(x)dx \\ &= \int_{-\infty}^{+\infty} a(g_1(X)) f(x)dx + \int_{-\infty}^{+\infty} b(g_2(X)) f(x)dx \\ &= a\int_{-\infty}^{+\infty} g_1(X) f(x)dx + b\int_{-\infty}^{+\infty} g_2(X) f(x)dx \\ &= aE(g_1(X)) + bEg_1(X) \end{aligned}$$

2. Let the function $g = g_2 - g_1$, the function g is positive and therefore according to proposition 2.7

$$E(g(X)) = E(g_2(X) - g_1(X)) \geq 0$$

and from the liniarity properity, we have

$$E(g_2(X)) - E(g_1(X)) \geq 0.$$

□

Definition 2.20. *A random variable is said to be centered if its expectation is zero.*

2.3.3.1 Moments

Definition 2.21. *Let X be a continuous random variable. Then for all $r \in \mathbb{N}$, we have*

1. *The $r-$th moment of a random variable X is defined to be $E[X^r]$ and given by*
$$E[X^r] = \int_{-\infty}^{+\infty} x^r f(x)\,dx.$$
2. *The r-th central moment of X is defined to be $E[(X-E(X))^r]$ and given by*
$$E[(X-E(X))^r] = \int_{-\infty}^{+\infty} (x-E(X))^r f(x)\,dx.$$

Remark 2.18. 1. *The rth moment about the mean is only defined if $E[(X-E(X))^r]$ exists.*

2. *Note that all odd moments of X around its mean are zero for symmetrical distributions, provided such moments exist.*

Remark 2.19. *Let X be an absolutely continuous random variable.*

1. *If X is centered, then its centered and non-centered moments coincide.*
2. *The centered moment of order 2 of X is exactly its variance.*
3. *The moments and centered moments of an absolutely continuous random variable belong in general to $\mathbb{R} \cup \{-\infty, +\infty\}$.*

2.3.4 Variance

Definition 2.22. *Let X be a continuous random variable and expected value $E(X)$ (i.e, μ_X). Then the variance (second central moment) of X, denoted $Var(X)$ or σ^2 is defined by*
$$\sigma^2 = Var(X) = E\left[(X-E(X))^2\right] = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\,dx.$$
The standard deviation of X is defined to be the square-root of the variance
$$\sigma = \sqrt{Var(X)}.$$

Remark 2.20. *The variance is a measure of the dispersion of the random variable about the mean.*

Theorem 2.7. *Let X be a continuous random variable, with expected value $E(X)$, and variance $Var(X)$. Then the following hold true:*

1. $Var(X) \geq 0$.

2. *If c is a constant,* $Var(c) = 0$.

3. *If a and b are real numbers,* $Var(aX + b) = a^2 Var(X)$.

4. $Var(X) = E(X)^2 - (E(X))^2$.

5. $Var(X) \leq E\left(X^2\right)$.

Example 2.28. *Consider the density function*

$$f(x) = \begin{cases} (p+1)x^p & 0 \leq x \leq 1 \\ 0 & \textit{otherwise} \end{cases}$$

where p is greater than -1. We can compute the $Var(X)$ as follows

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x)\, dx \\ &= \int_0^1 x(p+1)x^p\, dx \\ &= \left.\frac{x^{(p+2)}(p+1)}{(p+2)}\right|_0^1 \\ &= \frac{(p+1)}{(p+2)}. \end{aligned}$$

$$\begin{aligned} E\left(X^2\right) &= \int_{-\infty}^{+\infty} x^2 f(x)\, dx \\ &= \int_0^1 x^2(p+1)x^p\, dx \\ &= \left.\frac{x^{(p+3)}(p+1)}{(p+3)}\right|_0^1 \\ &= \frac{(p+1)}{(p+3)}. \end{aligned}$$

$$\begin{aligned} Var(X) &= E\left(X^2\right) - (E(X))^2 \\ &= \frac{p+1}{p+3} - \left(\frac{p+1}{p+2}\right)^2 \\ &= \frac{p+1}{(p+2)^2(p+3)}. \end{aligned}$$

The values of the mean and variances for various values of p are given in the following table

p	-5	0	1	2	∞
$E(X)$	$\frac{4}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{5}$	1
$Var(X)$	$\frac{2}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{3}{80}$	0

Table 2.7: The mean and the variance of continuous randon variable X.

2.4 Probability inequalities

Inequality has become an essential tool in many areas of mathematical research, for example in probability and statistics where it is frequently used in the proofs. "Probability Inequalities" covers inequalities related with events, distribution functions, characteristic functions, moments and random variables (elements) and their sum. In this section, we give some systematic description of the inequalities.
We begin with a classic result, Markov's inequality, which is very simple but also very useful and powerful.

2.4.1 Markov's inequality

In probability theory, Markov's inequality provides an upper bound on the probability that a non-negative random variable exceeds a certain threshold (positive constant). This inequality is named in honor of Andrei Markov and is very general, applying to any non-negative random variable. The Markov Inequality is particularly useful in scenarios where only the mean of the random variable is known, but not the full distribution. In other words, it provides a simple and general way to estimate upper bounds on probabilities.

Proposition 2.10. *If X is a nonnegative random variable, then, for all $\lambda > 0$ we have*

$$P(X \geq \lambda) \leq \frac{E(X)}{\lambda}.$$

Proof. We only give the proof for a continuous random variable, the case of a discrete random cariable is similar. Suppose X is a positive continuous random variable, we can write

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx,$$

since X is positive-valued and for any $\lambda > 0$ we have

$$\begin{aligned} E(X) &= \int_{0}^{+\infty} x f(x) dx \\ &\geq \int_{\lambda}^{+\infty} x f(x) dx \end{aligned}$$

(since $x > \lambda$ in the integrated region)

$$\begin{aligned} E(X) &\overset{x>\lambda}{\geq} \int_{\lambda}^{+\infty} \lambda f(x) dx, \\ &= \lambda \int_{\lambda}^{+\infty} f(x) dx \\ &= \lambda P(X \geq \lambda). \end{aligned}$$

Therefore

$$\lambda P(X \geq \lambda) \leq E(X).$$

□

Example 2.29. *Let $X \sim B(n,p)$. Using Markov's inequality, find an upper bound on $P(X \geq \alpha n)$, where $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.*
Solution: Note that X is a nonnegative random variable and $EX = np$. By Markov's inequality, we have

$$P(X \geq \alpha n) \leq \frac{E(X)}{\alpha n} = \frac{np}{\alpha n} = \frac{p}{\alpha}.$$

For $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$, we obtain

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{2}{3}.$$

Remark 2.21. *The previous proposition is a quantitative version of the fact that if the expected value of X is small, then it is unlikely that X will be too large.*

Remark 2.22.

1. *The Markov Inequality can be loose or not very informative if the random variable X is highly variable or has a skewed distribution.*

2. *It does not provide exact probabilities but rather an upper bound.*

Example 2.30. *Suppose* $P(X=3)=\frac{1}{2}, P(X=4)=\frac{1}{3}$ *and* $P(X=7)=\frac{1}{6}$. *Then*

$$\begin{aligned} E(X) &= 3\times\frac{1}{2}+4\times\frac{1}{3}+7\times\frac{1}{6} \\ &= 4. \end{aligned}$$

Hence, setting $\lambda=6$, *Markov's inequality says that* $P(X\geq 6)\leq \frac{4}{6}=\frac{2}{3}$. *In fact,* $P(X\geq 6)=\frac{1}{6}<\frac{2}{3}$.

Example 2.31. *Suppose* X *is a non-negative random variable with* $E(X)=10$. *To find an upper bound for* $P(X\geq 5)$, *use the Markov Inequality:*

$$P(X\geq 5)\leq \frac{E(X)}{5}=\frac{10}{5}=2.$$

This bound is actually not useful in this case because probabilities cannot exceed 1. *This highlights the sometimes loose nature of the bound.*

There are several generalizations of Markov's inequality that extend its applicability to a wider range of scenarios. It can be generalized to handle higher moments of a random variable.

Theorem 2.8. *Let* X *be a non-negative random variable and* $r>0$. *For any positive constant* λ, *the generalized Markov inequality is:*

$$P(|X|\geq\lambda)\leq\frac{E(|X|^r)}{\lambda^r}.$$

This inequality provides a bound on the probability that X exceeds a based on the r-th moment of X.
Markov's inequality is also used to prove Chebychev's inequality, perhaps the most important inequality in all of probability theory.

2.4.2 Chebyshev's inequality

Chebyshev's inequality is another fundamental result in probability theory that provides a way to bound the probability that a random variable deviates from its mean. It's a key tool in statistics and probability, particularly for understanding the spread of a distribution. This inequality is very general and applies to any random variable with a finite mean and variance, regardless of the distribution's shape. It does not require the random variable to be normally distributed or follow any specific distribution.

Proposition 2.11. *If X is any nonnegative random variable, with finite mean $E(X)$. Then, for any $\lambda > 0$ we have*

$$P(|X - E(X)| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}.$$

Proof. Let X be any random variable. Define $Y = (X - E(X))^2$, then Y is a nonnegative random variable, so we can apply Markov's inequality to Y. In particular, for any positive real number λ, we have

$$P\left(Y \geq \lambda^2\right) \leq \frac{E(Y)}{\lambda^2},$$

but note that

$$E(Y) = E\left((X - E(X))^2\right) = Var(X),$$

$$\begin{aligned} P\left(Y \geq \lambda^2\right) &= P\left((X - E(X))^2 \geq \lambda^2\right) \\ &= P(|X - E(X)| \geq \lambda). \end{aligned}$$

Thus we conclude that

$$P(|X - E(X)| \geq \lambda) \leq \frac{Var(X)}{\lambda^2},$$

which completes the proof. □

Remark 2.23. *Intuitively, Chebyshev's inequality says that if the variance of X is small, then it is unlikely that X will be too far from its mean value X.*

Now, we consider some examples:

Example 2.32. *Suppose again that $P(X = 3) = \frac{1}{2}, P(X = 4) = \frac{1}{3}$ and $P(X = 7) = \frac{1}{6}$. Then $E(X) = 4$. Also,*

$$\begin{aligned} E\left(X^2\right) &= 3^3 \times \frac{1}{2} + 4^2 \times \frac{1}{3} + 7^2 \times \frac{1}{6} \\ &= 18. \end{aligned}$$

so that $Var(X) = 18 - 4^2 = 2$. Hence, setting $\lambda = 1$, Chebychev's inequality says that $P(|X - 4| \geq 1) \leq \frac{2}{1^2} = 2$, which tells us nothing because we always have $P(|X - 4| \geq 1) \leq 1$. On the other hand, setting $\lambda = 3$, we get $P(|X - 4| \geq 3) \leq \frac{2}{3^2} = \frac{2}{9}$, which tells us nothing because we always have $P(|X - 4| \geq 3) = P(X = 7) = \frac{1}{6} < \frac{2}{9}$.

Example 2.33. *Let $X \sim B(n, p)$. We now will using Chebyshev's inequality, find an upper bound on $P(X \geq qn)$, where $p < q < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $q = \frac{3}{4}$.*

Solution: *One way to obtain a bound is to write*

$$\begin{aligned} P(X \geq qn) &= P(X - np \geq qn - np) \\ &\leq P(|X - np| \geq nq - np) \\ &\leq \frac{Var(X)}{nq - np} \\ &= \frac{p(1-p)}{n(q-p)^2}. \end{aligned}$$

For $p = \frac{1}{2}$ and $q = \frac{3}{4}$, we obtain

$$P\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{n}.$$

Example 2.34. *Let $X \sim N(0,1)$, and $\lambda = 5$. Then by Chebychev's inequality, $P(|X| \geq 5) \leq \frac{1}{5}$.*

Remark 2.24.

1. *The inequality is quite general but can be very loose. It provides a bound but not necessarily a precise measure of how concentrated the distribution is around the mean.*

2. *It doesn't give information about the shape of the distribution or how the probability is distributed within the range.*

2.4.3 Cauchy-Schwarz inequality

The Cauchy-Schwarz inequality is a fundamental result in linear algebra and analysis, with applications spanning various fields including probability theory, statistics, and functional analysis. It provides a bound on the inner product (or dot product) of two vectors in terms of their magnitudes. Let us state and prove the Cauchy-Schwarz inequality for random variables.

Proposition 2.12. *Suppose X and Y are two random variables, then*

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)},$$

and the equality holds if and only if $X = aY$ for some constant $a \in \mathbb{R}$.

proof. Define the random variable $U = (X - aY)^2$ which is a nonegative random variable for any value of $a \in \mathbb{R}$. Then

$$\begin{aligned} 0 \leq E(U) &= E(X - aY)^2 \\ &= E\left(X^2 - 2aXY + a^2Y^2\right) \\ &= E\left(X^2\right) - 2aE(XY) + a^2E\left(Y^2\right). \end{aligned}$$

Define $g(a) = E\left(X^2\right) - 2aE(XY) + a^2E\left(Y^2\right)$ which is a quadratic polynomial in a, then we know that $g(a) \geq 0$ for all a. Moreover, if $g(a) = 0$ for some a, then we have $E(U) = E(X - aY)^2 = 0$, which essentially means $X = aY$ with probability one. To prove the Cauchy-Schwarz inequality, choose $a = \frac{E(XY)}{E(Y^2)}$. We obatin

$$\begin{aligned}
g(a) &= E\left(X^2\right) - 2aE(XY) + a^2E\left(Y^2\right) \\
&= E\left(X^2\right) - 2\frac{E(XY)}{E(Y^2)}E(XY) + \frac{(E(XY))^2}{(E(Y^2))^2}E\left(Y^2\right) \\
&= E\left(X^2\right) - \frac{(E(XY))^2}{E(Y^2)},
\end{aligned}$$

so $g(a) \geq 0$ for all a if and only if

$$E\left(X^2\right) - \frac{(E(XY))^2}{E(Y^2)} \geq 0$$

Thus we conclude

$$E(XY)^2 \leq E\left(X^2\right)E\left(Y^2\right),$$

which imlpies

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}.$$

which is what we needed to show.
To deal with the last claim, observe that if $U > 0$ with probability one, then $g(a) = E(U) > 0$. This happens only if

$$E\left(X^2\right) - \frac{(E(XY))^2}{E(Y^2)} > 0$$

and if $E\left(X^2\right) - \frac{(E(XY))^2}{E(Y^2)} = 0$, then $g\left(\frac{E(XY)}{E(Y^2)}\right) = E(U) = 0$, which only can be true if

$$X - \frac{E(XY)}{E(Y^2)}Y = 0,$$

that is X is a scalar multiple of Y. □

Remark 2.25. *The Cauchy-Schwarz inequality is used in proofs of other inequalities, such as the Hölder's inequality and the Minkowski inequality.*

In linear regression, the Cauchy-Schwarz inequality is used to show that the correlation coefficient ρ between the predictor X and the response Y is bounded by 1.

Example 2.35. *Using the Cauchy-Schwarz inequality, show that for any two random variables X and Y*

$$\left|\rho(X,Y)\right| \le 1.$$

Also, $\left|\rho(X,Y)\right| = 1$ if and only if $Y = aX + b$ for some constants $a, b \in \mathbb{R}$.
***Solution**: Let*

$$U = \frac{X - E(X)}{\sigma_X}, V = \frac{Y - E(Y)}{\sigma_Y},$$

are two normalized random variables. Then $E(U) = E(V) = 0$, and $Var(U) = Var(V) = 1$. Using the Cauchy-Schwarz inequality for U and V, we obtain

$$|E(UV)| \le \sqrt{E(U^2)E(V^2)} = 1.$$

But note that $E(UV) = \rho(X,Y)$, thus we conclude

$$\left|\rho(X,Y)\right| \le 1,$$

where equality holds if and only if $V = \alpha U$ for some constant $\alpha \in \mathbb{R}$. Then

$$\frac{Y - E(Y)}{\sigma_Y} = \alpha \frac{X - E(X)}{\sigma_X},$$

therfore

$$Y = \frac{\alpha\sigma_Y}{\sigma_X} X + \left(E(Y) - \frac{\alpha\sigma_Y}{\sigma_X} E(X)\right),$$

which completes the proof.

Remark 2.26. *The Cauchy-Schwarz inequality is often used to bound the covariance between two random variables. If X and Y are random variables, then:*

$$|Cov(X,Y)| = |E(XY) - E(X)E(Y)| \le \sqrt{Var(X)Var(Y)},$$

where $Var(X) = E\left(X^2\right) - (E(X))^2$ and $Var(Y) = E\left(Y^2\right) - (E(Y))^2$ (i.e, if the variance of X or Y is small, then the covariance of X and Y must also be small).

Example 2.36. *Suppose $X = C$ is a constant. Then $Var(X) = 0$. It follows from the Cauchy-Schawrtz inequality that, for any random variable Y, we must have*
$Cov(X,Y) \le (Var(X)Var(Y))^{1/2} = (0 \times Var(Y))^{1/2} = 0$? so that $Cov(X,Y) = 0$.

2.4.4 Minkowski's inequality

Minkowski's inequality is a fundamental result in mathematics, particularly in the fields of analysis and probability theory. It generalizes the triangle inequality for vector norms to the context of integrals and random variables.

Theorem 2.9. *Let X and Y be ransom variables, and let $p \geq 1$ be a real number. Then the Minkowski's inequality states*

$$E\left(|X+Y|^p\right)^{\frac{1}{p}} \leq E\left(|X|^p\right)^{1/p} + E\left(|Y|^p\right)^{1/p}.$$

Special cases

1. For $p = 1$: Minkowski's inequality reduces to the triangle inequality for absolute values:
$$E\left(|X+Y|\right) \leq E\left(|X|\right) + E\left(|Y|\right).$$
2. For $p = 2$: Minkowski's inequality becomes a special case related to the Cauchy-Schwarz inequality.

Corollary 2.3. *Let X and Y be ransom variables, and let $p \geq 1$ be a real number. Then the Minkowski's inequality states*

$$E\left(|X+Y|^2\right)^{\frac{1}{2}} \leq E\left(|X|^2\right)^{1/2} + E\left(|Y|^2\right)^{1/2}.$$

Proof. It suffices to show that $E((X+Y)^2)$ is finite if $E(X^2)$ and $E\left(Y^2\right)$ are. Now,

$$E\left((X+Y)^2\right) = E\left(X^2\right) + 2E\left(X.Y\right) + E\left(Y^2\right),$$

and since

$$E\left((X.Y)\right) \leq E\left(X^2\right)^{1/2}.E\left(Y^2\right)^{1/2}$$

we have

$$\begin{aligned} E\left((X+Y)^2\right) &\leq E\left(X^2\right) + E\left(X^2\right)^{1/2}.E\left(Y^2\right)^{1/2} + E\left(Y^2\right) \\ &= \left(E\left(|X|^2\right)^{1/2} + E\left(|Y|^2\right)^{1/2}\right)^2. \end{aligned}$$

□

Remark 2.27. *Minkowski's inequality helps in deriving bounds on the p-th moments of random variables. It is used to show that the p-th moment of the sum of random variables is bounded by the sum of their individual p-th moments.*

Finally, we develop a more advanced inequality that is sometimes very useful.

2.4.5 Jensen's inequality

In mathematics, Jensen's inequality, named after the Danish mathematician Johan Jensen, relates the value of a convex function of an integral to the integral of the convex function. It was proved by Jensen in 1906 [3], building on an earlier proof of the same inequality for doubly-differentiable functions by Otto Hölder in 1889 [4]. Given its generality, the inequality appears in many forms depending on the context. It can be written in two ways: discrete or integral. It appears in particular in analysis, in measure theory and in probability (Rao-Blackwell theorem), but also in statistical physics, in quantum mechanics and in information theory (under the name of Gibbs inequality).
Jensen's inequality deals with a convex (or concave) function applied to the expectation of a random variable. We can state the definition for convex and concave functions in the following way:

Definition 2.23. *Consider a function* $\varphi : \mathbb{R} \to \mathbb{R}$.

- *We say that* φ *is convex function on* $[a,b]$ *if for any* $x, y \in [a,b]$ *and any* $\alpha \in [0,1]$ *we have*

$$\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y).$$

- *We say that* φ *is concave function on* $[a,b]$ *if*

$$\varphi(\alpha x + (1-\alpha)y) \geq \alpha\varphi(x) + (1-\alpha)\varphi(y).$$

Note that for a convex function φ this property holds for any convex linear combination of points in $[a,b]$, that is,

$$\varphi(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n) \leq \alpha_1\varphi(x_1) + \alpha_2\varphi(x_2) + ...\alpha_n\varphi(x_n), \tag{2.2}$$

such that $\alpha_1 + \alpha_2 + ... + \alpha_n = 1, 0 \leq \alpha_1, \alpha_2, ..., \alpha_n \leq 1$, and $x_1, x_2, ..., x_n \in [a,b]$.
If $n = 2$, the above statement is the definition of convex functions. You can extend it to higher values of n by induction.

[3] Jensen, J. L. W. V. (1906). "Sur les fonctions convexes et les inégalit és entre les valeurs moyennes". Acta Mathematica. 30 (1): 175–193.

[4] Guessab, A.; Schmeisser, G. (2013). "Necessary and sufficient conditions for the validity of Jensen's inequality". Archiv der Mathematik. 100 (6): 561–570.

Now, consider a discrete random variable X with n possible values $x_1, x_2, ..., x_n$. In Equation 2.2, we can choose $\alpha_1 = P(X = x_i) = P_X(x_i)$. Then, the left-hand side of 2.2 becomes $\varphi(E(X))$ and the right-hand side becomes $E(\varphi(X))$. So we can prove the Jensen's inequality in this case. Using limiting arguments, this result can be extended to other types of random variables.

Proposition 2.13. *Suppose X is a random variable such that $P(a \leq X \leq b) = 1$. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex on $[a, b]$, then*

$$E(\varphi(X)) \geq \varphi(E(X)).$$

If φ is concave, then

$$E(\varphi(X)) \leq \varphi(E(X)).$$

Proof. If X is constant, then there is nothing to prove, so assume X is not constant. Then we have

$$a < E(X) < b.$$

Denote: $c = E(X)$. Then

$$\varphi(x) \geq AX + B \text{ and } \varphi(E(X)) = AE(X) + B$$

for some $A, B \in \mathbb{R}$. Also note that

$$|\varphi(X)| \leq |A||X| + |B| \max\{|a|, |b|\} |X| + |B|,$$

so $E|\varphi(X)| < \infty$ and therfore $E(\varphi(X))$ is well defined. Now we can use $AX + B \leq \varphi(X)$ to see that

$$\varphi(E(X)) = AE(X) + B \leq E(\varphi(X)).$$

□

Jensen's inequality can be used to provide bounds on the variance of random variables. For example:

Example 2.37. *Remember that variance of every random variable X is a positive value, i.e.,*

$$Var(X) = E\left(X^2\right) - (E(X))^2 \geq 0.$$

Thus

$$E\left(X^2\right) \geq (E(X))^2.$$

if we define $\varphi(x) = x^2$, we can write the above inequality as

$$E(\varphi(x)) \geq \varphi(E(X)).$$

The function $\varphi(x) = x^2$ is an example of convex function. Jensen's inequality states that, for any convex function φ , we have $E(\varphi(x)) \geq \varphi(E(X))$.

Example 2.38. *For a linear function $\varphi(x) = ax + b$, where a and b are constants, Jensen's inequality becomes an equality because linear functions are both convex and concave. Thus:*

$$\varphi(E(X)) = aE(X) + b$$

and

$$E(\varphi(x)) = E(aX + b) = aE(X) + b.$$

In this case, Jensen's inequality simply states:

$$aE(X) + b = E(aX + b)$$

which is always true.

Example 2.39. *Consider $\varphi(x) = e^x$, which is a convex function. Jensen's inequality tells us:*

$$e^{E(X)} \leq E\left(e^X\right).$$

This inequality is particularly useful in probability theory and statistics for bounding the exponential moment of a random variable. It's applied in contexts like large deviation theory.

Example 2.40. *Consider $\varphi(x) = \log(x)$, which is a concave function. For $X > 0$, Jensen's inequality provides:*

$$\log(E(X)) \geq E(\log(X)).$$

This inequality is used in information theory and economics, particularly for analyzing the logarithmic utility in risk-averse scenarios.

More general, to use Jensen's inequality, we need to determine if a function φ is convex. A useful method is the second derivative.

Definition 2.24. *A twice-differentiable function $\varphi : \mathbb{R} \to \mathbb{R}$ is convex on $[a, b]$ if and only if $\varphi''(x) \geq 0$ for all $x \in [a, b]$.*

Example 2.41. *If $\varphi(x) = x^2$, then $\varphi''(x) = 2 \geq 0$,thus $\varphi(x) = x^2$ is convex over $\mathbb{R}$.*

Example 2.42. *Let X be a positive random variable. Compare $E(X^p)$ with $(E(X))^p$ for all values of $p \in \mathbb{R}$.*

Solution: *First note*

$$E(X^p) = 1 = (E(X))^p, \text{ if } p = 0,$$
$$E(X^p) = E(X) = (E(X))^p, \text{ if } p = 1.$$

So let's assume $p \neq 0, 1$. Letting $\varphi(x) = x^p$, we have

$$\varphi''(x) = p(p-1)x^{p-2}.$$

On $(0, \infty)$, we can say $\varphi''(x)$ is positive, if $p < 0$ or $p > 1$. It is negative, if $0 < p < 1$. Therfore we conclude that $\varphi(x)$ is convex, if $p < 0$ or $p > 1$. It is concave, if $0 < p < 1$. Using Jensen's inequality we conclude

$$E(X^p) \geq (E(X))^p, \text{ if } p < 0 \text{ or } p > 1,$$
$$E(X^p) \leq (E(X))^p, \text{ if } 0 < p < 1.$$

In particular,

$$E\left(X^2\right) \geq (E(X))^2,$$

and therfore $E\left(X^2\right) - (E(X))^2 \geq 0$.

CHAPTER

3

PROBABILITY DISTRIBUTION

Contents

3.1 Discrete probability distribution

Discrete probability distribution is a type of probability distribution that shows all possible values of a discrete random variable along with the associated probabilities. In other words, a discrete probability distribution gives the likelihood of occurrence of each possible value of a discrete random variable. These distributions are tools to make solving probability problems easier. Each distribution has its own special characteristics. Bernoulli distributions, binomial distributions and Geometric distributions are some commonly used discrete probability distributions. This section sheds light on the definition of a discrete probability distribution, its formulas, types, and various associated examples. In this part, we will learn about the formula, pmf, cdf, and other aspects of the Bernoulli Distribution

3.1.1 Bernoulli distribution

Bernoulli distribution is a special kind of distribution that is used to model real-life examples and can be used in many different types of applications. It can be used to describe events that can only have two outcomes, that is: success or failure, life or death, hit or miss, male or female, ...etc.

Definition 3.1. *A Bernoulli experiment is an experiment which has two outcomes which we call (by convention) "success" S and failure F. Such an experiment is called a Bernoulli trial.*

Example 3.1.

1. *A pass or fail exam can be modeled by a Bernoulli Distribution.*
2. *Flipping a coin, we will call a head a success and a tail a failure.*

Remark 3.1. *Often we call a "success" something that is in fact for from an actual schusses. E.g., a machine breaking down.*

In order to obtain a Bernoulli random variable if we first assign probabilities to S and F by

$$P(S) = p \quad \text{and} \quad P(F) = q$$

so again $p + q = 1$. Thus the sample space of a a Bernoulli experiment will be denoted Ω, and is given by $\Omega = \{S, F\}$. We then obtain a Bernoulli random variable X on Ω by defining

$$X(S) = 1, X(F) = 0$$

so

$$P(X = 1) = P(S) = p \quad \text{and} \quad P(X = 0) = P(F) = q.$$

Remark 3.2. *Any random variable whose only possible values are* 0 *and* 1 *is called a Bernoulli random variable.*

Definition 3.2. *A Bernoulli distribution is a type of discrete probability distribution where the random variable can either be equal to* 0 *(failure) or be equal to* 1 *(success).*

The probability of getting a success is p and that of a failure is $1-p$. It is denoted as $X \sim B(p)$, with $p \in]0,1[$. The probability mass function is expressed as follows:

$$P(X = x) = \begin{cases} 1-p & \text{if} \quad x = 0 \\ p & \text{if} \quad x = 1 \end{cases}$$

Bernoulli's probability function can be written in the following form:

$$P_X(k) = \begin{cases} p^k(1-p)^{1-k} & \text{if} \quad k \in \{0,1\} \\ 0 & \text{otherwise} \end{cases}$$

Example 3.2. *Suppose there is an experiment where you flip a coin that is fair. If the outcome of the flip is heads then you will win. This means that the probability of getting heads is* $p = 1/2$*. If* X *is the random variable following a Bernoulli Distribution, we get* $P(X = 1) = p = 1/2$.

Example 3.3. *A basketball player can shoot a ball into the basket with a probability of 0.6. What is the probability that he misses the shot?*
Solution: *We know that success probability* $P(X = 1) = p = 0.6$*. Thus, probability of failure is* $P(X = 0) = 1 - p = 1 - 0.6 = 0.4$*. The probability of failure of the Bernoulli distribution is* 0.4.

Remark 3.3. *The cumulative distribution function of a Bernoulli random variable X when evaluated at x is defined as the probability that X will take a value lesser than or equal to x.*

Definition 3.3. *Let X be a Bernoulli random variable, then the cumulative distribution function of X is given by*

$$F_X(x) = \begin{cases} 0 & \textit{if} \quad x < 0 \\ 1-p & \textit{if} \quad 0 \le x < 1 \\ 1 & \textit{if} \quad x \ge 1 \end{cases}$$

Theorem 3.1. *If X is a Bernoulli random variable, then the mean or expected value of the Bernoulli random variable is* $E(X) = p$.

Proof. We know that for X, $P(X = 1) = p$, $P(X = 0) = q$, then

$$\begin{aligned} E(X) &= \sum xP(X = x) \\ &= 1 \times p + 0 \times q \\ &= p. \end{aligned}$$

□

Example 3.4. *If a Bernoulli distribution has a parameter* 0.45 *then find its mean.*
Solution: $X \sim B(0.45)$. $E(X) = p = 0.45$.

Theorem 3.2. *If X is a Bernoulli random variable, then the variance of the Bernoulli random variable is* $Var(X) = pq$ *and the standard deviation is* $\sigma = \sqrt{pq}$.

Proof. The variance can be defined as the difference of the mean of X^2 and the square of the mean of X. Thematically this statement can be written as follows:

$$Var(X) = E\left(X^2\right) - (E(X))^2$$

Using the properties of $E\left(X^2\right)$, we get,

$$\begin{aligned} E\left(X^2\right) &= \sum x^2 P(X = x) \\ &= 1^2 \times p + 0^2 \times q \\ &= p \end{aligned}$$

Substituting this value in $Var(X)$ we have

$$\begin{aligned} Var(X) &= p - p^2 \\ &= p(1-p) \\ &= pq. \end{aligned}$$

□

Example 3.5. *If a Bernoulli distribution has a parameter* 0.72 *then find its variance.*
Solution: $X \sim B(0.45)$. $Var(X) = pq = 0.72 \times 0.28 = 0.2016$.

The graph of a Bernoulli distribution helps to get a visual understanding of the probability density function of the Bernoulli random variable.

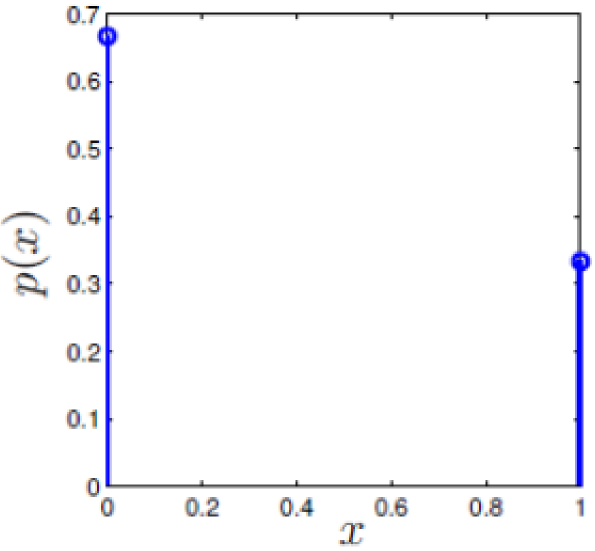


Figure 3.1: Bernoulli distribution graph when $p = \frac{1}{3}$.

The graph shows that the probability of success is $p = \frac{1}{3}$ when $X = 1$ and the probability of failure of X is $(1 - p) = \frac{2}{3}$ or q if $X = 0$.

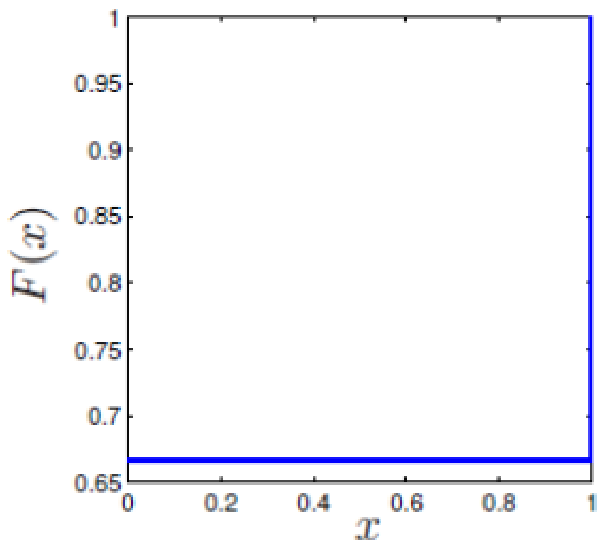


Figure 3.2: Bernoulli cumulative distribution graph when $p = \frac{1}{3}$.

Bernoulli distribution is a simple distribution and hence, is widely used in many industries. Given below are some applications of Bernoulli distribution.

- In medicine, Bernoulli distributions are used to model the events experienced by a single patient. These events could be disease, death, and so on.

- Logistic regressions use Bernoulli distribution to model the occurrence of certain events such as the specific outcome of a dice roll.

- Bernoulli distribution is also used as a basis to derive several other probability distributions that have applications in the engineering, aerospace, and medical industries.

3.1.2 Binomial distribution

In statistics and probability theory, the binomial distribution is the probability distribution that is discrete and applicable to events having only two possible results in an experiment, either success or failure. (the prefix "bi" means two, or twice). A few circumstances where we have binomial experiments are tossing a coin: head or tail, the result of a test: pass or fail, selected in an interview: yes/ no, or nature of the product: defective/non-defective. Such a distribution of a binomial random variable is called a binomial probability distribution.

Definition 3.4. *A binomial distribution is a discrete probability distribution that gives the success probability in n Bernoulli trials.*

Let X denotes k successes in a sequence of n independent trials, the probability of getting a success is given by p ($q := 1-p$ is the probability of failure), we have

Definition 3.5. *A random variable X is said to follow binomial distribution if it assumes only non negative values and its probability mass function is given by*

$$P(X=k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, ..., n.$$

We denote binomial distribution as $B(n,p)$. We say $X \sim B(n,p)$ where the tilde "$\sim$" is read "as distributed as " and p and n called parameters of the distribution such that $p \in]0,1[$ and $n \in N^*$.

Remark 3.4. *When using the binomial formula to solve problems, we have to identify three things:*

1. *The number of trials: n.*

2. *The probability of a success on any one trial p.*

3. *The number of successes desired X.*

Example 3.6. *In a military exercise, a soldier is allowed to shoot at a moving target 10 times. If the probability of hitting the target is 0.7, what is the probability that the soldier hits the target at least 2 times? and at most 3 times?*
Solution: *This random experiment involves repeating the same experiment (shooting at a target) 10 times in a row, which is a binomial experiment. Let X be the random variable that models this experiment, where $X \sim B(10, 0.7)$. We are interested in finding $P(X \geq 2), P(X \leq 3)$.*
To find $P(X \geq 2)$, you would calculate it as:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - [P(X = 0) + P(X = 1)]. \end{aligned}$$

Using the binomial probability formula:

$$C_n^k p^k (1-p)^{n-k}$$

where $n = 10, p = 0.7$ and k is the number of successes. Then we have:

$$\begin{aligned} P(X \geq 2) &= 1 - [P(X = 0) + P(X = 1)] \\ &= 1 - \left[C_{10}^0 (0.7)^0 (1-0.7)^{10-0} + C_{10}^1 (0.7)^1 (1-0.7)^{10-1}\right] \\ &= 0.9998. \end{aligned}$$

To find $P(X \leq 3)$, you would calculate it as:

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= C_{10}^0 (0.7)^0 (1-0.7)^{10-0} + C_{10}^1 (0.7)^1 (1-0.7)^{10-1} \\ &\quad + C_{10}^2 (0.7)^2 (1-0.7)^{10-2} + C_{10}^3 (0.7)^3 (1-0.7)^{10-3} \\ &= 0.01059. \end{aligned}$$

Example 3.7. *Let $X \sim B(10, \frac{1}{3})$, we have*

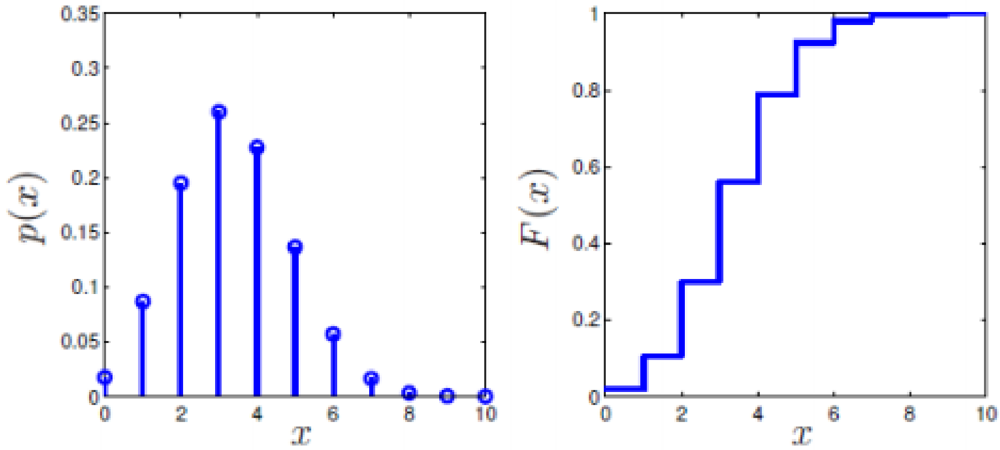


Figure 3.3: Binomial distribution graph of $B\left(10, \frac{1}{3}\right)$.

Example 3.8. *A couple, who are both carriers for a recessive disease, wish to have* 5 *children. They want to know the probability that they will have four healthy kids*

Solution:

$$P(X = 4) = C_5^4 (0.75)^4 (1 - 0.75)^{5-4} = 0.395$$

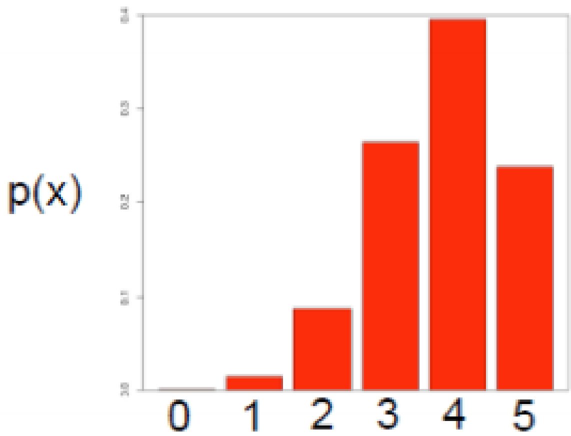


Figure 3.4: Probability mass function of randon variable X for a recessive disease.

Remark 3.5. *In a single experiment when $n = 1$, the binomial distribution is called a Bernoulli distribution.*

Remark 3.6. *The term "binomial" in "binomial distribution" indeed has its roots in the binomial theorem from algebra, and the binomial coefficients C_n^k in the probability mass function of* *the binomial distribution come directly from the expansion of* $(p+q)^n$ *, the expansion is given by:*

$$(p+q)^n = \sum_{k=0}^{n} C_n^k p^{n-k} q^k = 1.$$

Remark 3.7. *There is another way of expressing the binomial distribution that is sometimes*

Remark 3.8. *useful. For example, if* $X_1, X_2, \ldots, X_n$ *are chosen independently and each has the Bernoulli distribution* $B(\theta)$*, and* $Y = X_1 + X_2 + \ldots + X_n$ *, then* Y *will have the Binomial distribution* $B(n, \theta)$*.*

Definition 3.6. *Let* X *be a Binomial random variable, then the cumulative distribution function of* X *is given by*

$$F_X(x) = P(X \leq x) = \sum_{k=0}^{x} C_n^k p^k (1-p)^{n-k}.$$

Theorem 3.3. *If* X *is a binomial random variable, then the mean of* X *is:* $\mu = E(X) = np$*.*

Proof.

$$\begin{aligned}
\mu = E(X) &= \sum_{k=0}^{n} k.P(X=k) \\
&= \sum_{k=0}^{n} k C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}
\end{aligned}$$

let $s = k-1$, so $k = s+1$

$$\begin{aligned}
\mu &= \sum_{s=0}^{n-1} \frac{n!}{s!(n-1-s)!} p^{s+1} (1-p)^{n-1-s} \\
&= \sum_{s=0}^{n-1} \frac{n \times (n-1)!}{s!(n-1-s)!} p \times p^s (1-p)^{n-1-s} \\
&= np \sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-1-s)!} p^s (1-p)^{n-1-s} \\
&= np
\end{aligned}$$

where $\sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-1-s)!} p^s (1-p)^{n-1-s} = 1$, with the random variable following $B(n-1, p)$. □

Theorem 3.4. *If X is a binomial random variable, then the variance of X is:*

$$\sigma^2 = Var(X) = np(1-p)$$

and the standard deviation of X is:

$$\sigma = \sqrt{Var(X)} = \sqrt{np(1-p)}.$$

Proof.

$$\begin{aligned}
\sigma^2 = Var(X) &= E\left(X^2\right) - [E(X)]^2 \\
&= E\left(X^2\right) - E(X) + E(X) - [E(X)]^2, \text{ (we add 0)} \\
&= E\left(X^2 - X\right) + E(X) - [E(X)]^2, \text{ from linearity of the expectation} \\
&= E(X(X-1)) + E(X) - [E(X)]^2.
\end{aligned}$$

From the definition of the expected value of a function, we have

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{n} k(k-1).P(X=k) \\
&= \sum_{k=0}^{n} k(k-1)C_n^k p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k (1-p)^{n-k}.
\end{aligned}$$

The first two terms of the summation equal zero when $k = 0$ and $k = 1$. Therfore, the bottom index on the summation can be changed from $k = 0$ to $k = 2$, as it is here

$$\begin{aligned}
E(X(X-1)) &= \sum_{k=2}^{n} k(k-1)\frac{n!}{k!(n-k)!}p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!}p^k (1-p)^{n-k}
\end{aligned}$$

let $s = k-2$,so $k = s+2$

$$\begin{aligned}
E(X(X-1)) &= \sum_{s=0}^{n-2} \frac{n!}{s!(n-2-s)!}p^{s+2} (1-p)^{n-2-s} \\
&= n(n-1)p^2 \sum_{s=0}^{n-2} \frac{(n-2)!}{s!(n-2-s)!}p^s (1-p)^{n-2-s} \\
&= n(n-1)p^2,
\end{aligned}$$

where $\sum_{s=0}^{n-2} \frac{(n-2)!}{s!(n-2-s)!} p^s (1-p)^{n-2-s} = 1$, with the random variable following $B(n-2, p)$. Thus

$$\begin{aligned}
\sigma^2 &= E(X(X-1)) + E(X) - [E(X)]^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p).
\end{aligned}$$

□

Remark 3.9. *We can proof the above Theorem from properties of mean and variance as*

$$E(X) = E\left(\sum_{k=1}^{n} Y_k\right) = \sum_{k=1}^{n} E(Y_k) = \sum_{k=1}^{n} p = np.$$

$$Var(X) = Var\left(\sum_{k=1}^{n} Y_k\right) = \sum_{k=1}^{n} Var(Y_k) = \sum_{k=1}^{n} pq = npq.$$

3.1.3 Multinomial distribution

The multinomial distribution is a generalization of the binomial distribution and is used to model scenarios where there are more than two possible outcomes. its properties and applications are foundational in statistics, data science, and various fields of research.

Definition 3.7. *The multinomial distribution describes the probabilities of obtaining counts among multiple categories in a fixed number of trials.*

Parameters:

- n: the number of trials (or experiment) (i.e, $\in \{0, 1, 2, \ldots\}$.

- k: the number of categories (or possible outcomes).

- **p**: a vector of probabilities fro each category, where $p_1, p_2, \ldots, p_k$ are the probabilities of each category and must satisfy:

- $p_i \geq 0$ for all i.

- $p_1 + p_2 + ... + p_k = 1$.

Notation:
If $X_1, X_2, ..., X_k$ are the counts of outcomes in each category after n trials, the the random vector $(X_1, X_2, ..., X_k)$ is said to have a multinomial distribution with index n and parameter vector $(p_1, ..., p_k)$ which we denotes as $(X_1, X_2, ..., X_k) \sim \mathcal{M}(n, p_1, ..., p_k)$.
Properties:
The multinomial distribution arises from an experiment with the following properties:

- A fixed number n of trials.
- Each trial is independent of the others.
- Each trial has k mutually exclusive and exhaustive possible outcomes, denoted by $A_1, ..., A_k$.
- On each trial, A_i occurs with probability $p_i, i = 1, ..., k, (i.e,\ P(A_i) = p_i)$.

Definition 3.8. *The probability mass function of the multinomial distribution is given by:*

$$f(x_1, x_2, ..., x_k; n, p_1, ..., p_k) = P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k)$$
$$= \begin{cases} \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k} & when \quad \sum_{i=1}^{k} x_i = n \\ 0 & otherwise, \end{cases}$$

where x_i are non-negative integers representing the counts of each category.

Remark 3.10. *The multinomial coefficient $\frac{n!}{x_1! x_2! ... x_k!}$ is the number of possible ways to put n balls into k boxes.*

Example 3.9. *Suppose that in a three-way election for a large country, candidate A received 20% of the votes, candidate B received 30% of the votes, and candidate C received 50% of the votes. If six voters are selected randomly, what is the probability that there will be exactly one supporter for candidate A, two supporters for candidate B and three supporters for candidate C in the sample?*
Solution: *Since we're assuming that the voting population is large, it is reasonable and permissible to think of the probabilities as unchanging once a voter is selected for the sample.*

$$P(A = 1, B = 2, C = 3) = \frac{6!}{1!2!3!}(0.2)^1 \times (0.3)^2 \times (0.5)^3 = 0.135.$$

Example 3.10. *Suppose a bag contains 10 balls: 4 red, 3 blue and 3 green. You want to draw 5 balls from the bag, and you are interested in the number of balls of each color you draw. Calculate the probability of drawing: 2 red balls, 2 blue balls and 1 green ball.*
Solution: *we have*

$$P(X_R = 2, X_B = 2, X_G = 1) = \frac{5!}{2!2!1!}(0.4)^2 \times (0.3)^2 \times (0.3)^1 = 0.1296.$$

Th probability of drawing 2 red balls, 2 blue balls and 1 green ball when drawing 5 balls from the bag is 0.1296.

Proposition 3.1. *If $(X_1, X_2, ..., X_k) \sim \mathcal{M}(n, p_1, ..., p_k)$. then the expected value of each category is given by*

$$E(X_i) = np_i, i = 1, ..., k.$$

Proposition 3.2. *If $(X_1, X_2, ..., X_k) \sim \mathcal{M}(n, p_1, ..., p_k)$. then the variance of each category count is*

$$Var(X_i) = np_i(1 - p_i), i = 1, ..., k.$$

and the covariance between two different categories X_i and X_j where $(i \neq j)$ is

$$Cov(X_i, X_j) = -np_i p_j.$$

Remark 3.11. *All covariances are negative because for fixed n, an increase in one component of a multinomial vector requires a decrease in another component.*

3.1.4 Hypergeometric distribution

The hypergeometric distribution models the number of successes in a fixed number of draws without replacement from a finite population.

Definition 3.9. *The hypergeometric distribution is a discrete probability distribution that describes the probability of obtaining a specific number of successes in a sample drawn from a population containing a certain number of successes and failures.*

Parameters:

- N: the total number of items in the population.

- K: the total number of successes in the population (the $N - K$ objects are failures).

- n: the number of draws (samples) taken from the population.

- k: the number of observed successes in the drawn sample.

Notation

If X is the random variable representing the number of successes in n draws from a population of N items containing K successes, we write: $\mathcal{H}(N,K,n)$.

Definition 3.10. *The probability mass function of the hypergeometric distribution is given by:*

$$P(X=k)=\begin{cases}\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} & \text{if} \quad k\in\{0,1,\dots,\min(K,n\}\\ 0 & \text{otherwise}\end{cases}$$

where the symbol $\binom{K}{k}$ refers to the number of possible combination of k objects chosen from among K distinct objects.

Remark 3.12. *The formula is valid for k such that*

$$\max(0,n+K-N)\leq k\leq\min(K,n).$$

Remark 3.13. *Using the notation of binomial coefficients, or, using factorial notation, we have*

$$P(X=k)=\frac{n!K!(N-n)!(N-K)!}{N!k!(K-k)!(n-k)!(N-K-n+k)!}$$

Proposition 3.3. *If $X\sim\mathcal{H}(N,K,n)$. Then the expected value or mean of the hypergeometric distribution is*

$$E(X)=np,$$

and the variance (square of the standard deviation) is given by

$$Var(X)=np(1-p)\frac{N-n}{N-1}$$

where $p=\frac{K}{N}$.

Example 3.11. *A school site committee is to be chosen from 4 men and 7 women. If the committee consists of 4 members, Let X be a random variable representing the number of men on this committee. Find the probability distribution of the variable X.*

Solution: *Let X = the number of men on the committee of* 4. *The men are the group of interest (first group). X takes on the values* 0, 1, 2, 3, 4, *where* $N = 11, K = 4$, *and* $n = 4$. $X \sim \mathcal{H}(11, 4, 4)$.

$$P(X=0) = \frac{\binom{4}{0}\binom{11-4}{4-0}}{\binom{11}{4}} = 0.106$$
$$P(X=1) = \frac{\binom{4}{1}\binom{11-4}{4-1}}{\binom{11}{4}} = 0.424$$
$$P(X=2) = \frac{\binom{4}{2}\binom{11-4}{4-2}}{\binom{11}{4}} = 0.38$$
$$P(X=3) = \frac{\binom{4}{3}\binom{11-4}{4-3}}{\binom{11}{4}} = 0.085$$
$$P(X=4) = \frac{\binom{4}{4}\binom{11-4}{4-4}}{\binom{11}{4}} = 0.003$$

From here we can put the probability distribution of the variable X as follows:

k	0	1	2	3	4	Sum
$P(X=k)$	0.106	0.424	0.38	0.085	0.003	1

Table 3.1: The hypergeometric distribution for random variable X.

3.1.5 Poly-hypergeometric distribution

The multivariate hypergeometric distribution generalizes the hypergeometric distribution to multiple categories or types of items. Instead of just counting successes and failures, it accounts for different types of successes in the population.

Definition 3.11. *The multivariate hypergeometric distribution describes the probability of drawing a specific number of successes from multiple categories when sampling without replacement.*

In a multivariate hypergeometric distribution, we have

- Population size (N): total number of items in population.

- Success categories (K): a vector $K = (K_1, K_2, ..., K_m)$ where K_i is the number of items of category i in the population. The total number of items is $N = K_1 + K_2 + ... + K_m + K_0$ (where K_0 is the number of items that do no belong to any category).

- Sample size (n): the total number of items drawn from the population.

- Sample count (k): a vector $k = (k_1, k_2, ..., k_m)$ where each k_i is the number of items drawn from category i.

Notation

If X is the random variable vector representing the number of successes in n draws from a population of N items containing K successes, we write: $X \sim \mathcal{MH}(K_1, ..., K_m; n)$

Remark 3.14. - *The number of items drawn n must not exceed the total number of items N.*

- *The counts of each category in the drawn sample must not exceed the counts in the population:*

$$0 \leq k_i \leq K_i \text{ for all } i$$

- *The total counts in the drawn sample must equal n :*

$$k_1 + k_2 + ... + k_m = n.$$

Definition 3.12. *The probability mass function of the multivariate hypergeometric distribution with parameters N, K and n is given by*

$$P(X = k) = \frac{\prod_{i=1}^{m} \binom{K_i}{k_i}}{\binom{N}{n}},$$

where

- *X is the random vector representing the counts of each category in the drawn sample. (i.e,* $X = (X_1, X_2, ..., X_m)$*).*

- $\binom{K_i}{k_i}$ *is the binomial coefficient representing the number of ways to choose* k_i *successes from* K_i.

- $\binom{N}{n}$ *is the total number of ways to choose n items from N.*

Definition 3.13. *An alternate form of the probability density function of* $(X_1, X_2, ..., X_m)$ *is*

$$P(X_1 = k_1, X_2 = k_2, ..., X_m = k_{m,}) = \binom{n}{k_1, k_2, ..., k_{m,}} \frac{k_1^{K_1} k_2^{K_1} ... k_m^{K_m}}{\binom{N}{n}} \text{ with } \sum_{i=1}^{m} k_i = n.$$

Remark 3.15. - *Dependent sampling: The draws are independent because the sampling occurs without replacement. This is a significant distinction from the binomial distribution.*

- *Marginal distributions: The marginal distribution for any single category follows a hypergeometric distribution.*

Proprieties:

- **Mean:** The expected value for each category i is given by:

$$E(X_i) = n\frac{K_i}{N}.$$

- **Variance:** The variance for each category i is given by:

$$Var(X_i) = n \times \frac{K_i}{N} \times \left(1 - \frac{K_i}{N}\right) \times \frac{N-n}{N-1}.$$

- **Covariance:** For categories i and j :

$$Cov\left(X_i, X_j\right) = -n \times \frac{K_i}{N} \times \frac{K_j}{N} \times \frac{N-n}{N-1}.$$

Example 3.12. *A population of* 100 *voters consists of* 40 *republicans,* 35 *democrats and* 25 *independents. A random sample of* 10 *voters is chosen. Find each of the following:*

1. *The joint density function of the number of republicans, number of democrats, and number of independents in the sample.*
2. *The mean of each variable in (1).*
3. *The variance of each variable in (1).*
4. *The covariance of each pair of variables in (1).*
5. *The probability that the sample contains at least* 4 *republicans, at least* 3 *democrats, and at least* 2 *independents.*

Solution:

1. *We have* $N = 100, n = 10, K_1 = 40, K_2 = 35, K_1 = 25$, *then*

$$P(X = x_, Y = y, Z = z) = \frac{\binom{40}{x}\binom{35}{y}\binom{25}{z}}{\binom{100}{10}} \quad \textit{for} \quad x, y, z \in \mathbb{N} \textit{ with } x + y + z = 10.$$

2. *We have* $E(X_i) = n\frac{K_i}{N}$, *then*

$$E(X) = 10\frac{40}{100} = 4,$$
$$E(X) = 10\frac{35}{100} = 3.5,$$
$$E(X) = 10\frac{25}{100} = 2.5.$$

3. *We have*

$$Var(X_i) = n \times \frac{K_i}{N} \times \left(1 - \frac{K_i}{N}\right) \times \frac{N-n}{N-1}.$$

Then

$$Var(X) = 10 \times \frac{40}{100} \times \left(1 - \frac{40}{100}\right) \times \frac{100-10}{100-1} = 2.1818,$$
$$Var(Y) = 10 \times \frac{35}{100} \times \left(1 - \frac{35}{100}\right) \times \frac{100-10}{100-1} = 2.0682,$$
$$Var(Z) = 10 \times \frac{25}{100} \times \left(1 - \frac{25}{100}\right) \times \frac{100-10}{100-1} = 1.7045.$$

4. *We have*

$$\begin{aligned} Cov(X,Y) &= -n \times \frac{K_1}{N} \times \frac{K_2}{N} \times \frac{N-n}{N-1} \\ &= -10 \times \frac{40}{100} \times \frac{35}{100} \times \frac{100-10}{100-1} \\ &= -1.6346. \end{aligned}$$

$$\begin{aligned} Cov(X,Z) &= -n \times \frac{K_1}{N} \times \frac{K_3}{N} \times \frac{N-n}{N-1} \\ &= -10 \times \frac{40}{100} \times \frac{25}{100} \times \frac{100-10}{100-1} \\ &= -0.9091. \end{aligned}$$

$$\begin{aligned} Cov(Y,Z) &= -n \times \frac{K_2}{N} \times \frac{K_3}{N} \times \frac{N-n}{N-1} \\ &= -10 \times \frac{35}{100} \times \frac{25}{100} \times \frac{100-10}{100-1} \\ &= -0.7955. \end{aligned}$$

5. *We have* $x = 4, y = 3, z = 2$, *then*

$$\begin{aligned} P(X = 4, Y = 3, Z = 2) &= \frac{\binom{40}{4}\binom{35}{3}\binom{25}{2}}{\binom{100}{10}} \quad \text{with} \quad x + y + z = 10 \\ &= 0.2474. \end{aligned}$$

3.1.6 Geometric distribution

Geometric distribution is a another type of discrete probability distribution that represents the probability of the number of successive failures before a success is obtained in a Bernoulli trial. A Bernoulli trial is an experiment that can have only two possible outcomes, i.e., success or failure. In other words, in a geometric distribution, a Bernoulli trial is repeated until the first success is obtained and then stopped.
The geometric probability distribution is widely used in several real-life scenarios. For example, in financial industries, geometric distribution is used to do a cost-benefit analysis to estimate the financial benefits of making a certain decision. In this section, we will study the meaning of geometric distribution, examples, and certain related important aspects.

Definition 3.14. *A random variable X is said to have a geometric distribution if it assumes only non negative values and its probability mass function is given by*

$$P(X=k)=\begin{cases}(1-p)^{k-1}p & \text{for} \quad k=1,2,3,\ldots \\ 0 & \text{otherwise}\end{cases}$$

where k is the number of trials until the first success $(i.e., k=1,2,3,\ldots)$, and p is the probability of success on each trial.

If X is a random variable representing the number of trials until the first success, then X follows a geometric distribution with parameter p, denoted as $X \sim G(p)$, with $p \in]0,1[$.

Example 3.13.

1. *The number of light bulbs tested that work until the first bulb that does not (a working bulb is considered a "failure" for the test).*

2. *The number of at bats without a hit until the first hit for the baseball player both follow the geometric distribution.*

Example 3.14. *In a factory, if the probability that a product is defective is $p=0.1$, the number of items checked until finding the first defective item follows a geometric distribution with $p=0.1$.*

Example 3.15. *We toss a coin such that $p(\{T\})=\theta$ where $0<\theta<1$ and we denote by X: "The number of tosses needed to get the first tails".*

It is clear that $X(\Omega) = \{1, 2, ..\} = \mathbb{N}^*$*and*

$$P(X = 1) = p\{\omega \in \Omega : X(\omega) = 1\} \quad \text{where} \quad \Omega = \{T, H\}$$

then

$$P(X = 1) = p(\{T\}) = \theta.$$

$$P(X = 2) = p\{\omega \in \Omega : X(\omega) = 2\} \quad \text{where} \quad \Omega = \{(r_1, r_2)/r_1, r_2 = T \vee H\}$$

then from the independence between the trials, we have

$$P(X = 2) = p(\{H, T\}) = (1 - \theta)\theta.$$

$$P(X = 3) = p\{\omega \in \Omega : X(\omega) = 3\} \quad \text{where} \quad \Omega = \{(r_1, r_2, r_3)/r_1, r_2, r_3 = T \vee H\}$$

then

$$P(X = 3) = p(\{H, H, T\}) = (1 - \theta)^2\theta.$$

Using the recurring formula we can notice that for all $k \in \mathbb{N}^*$

$$P(X = k) = p(\{H, H, ..., T\}) = (1 - \theta)^{k-1}\theta.$$

Example 3.16. *Figure 2.5 contains the plots of several Geometric probability functions.*

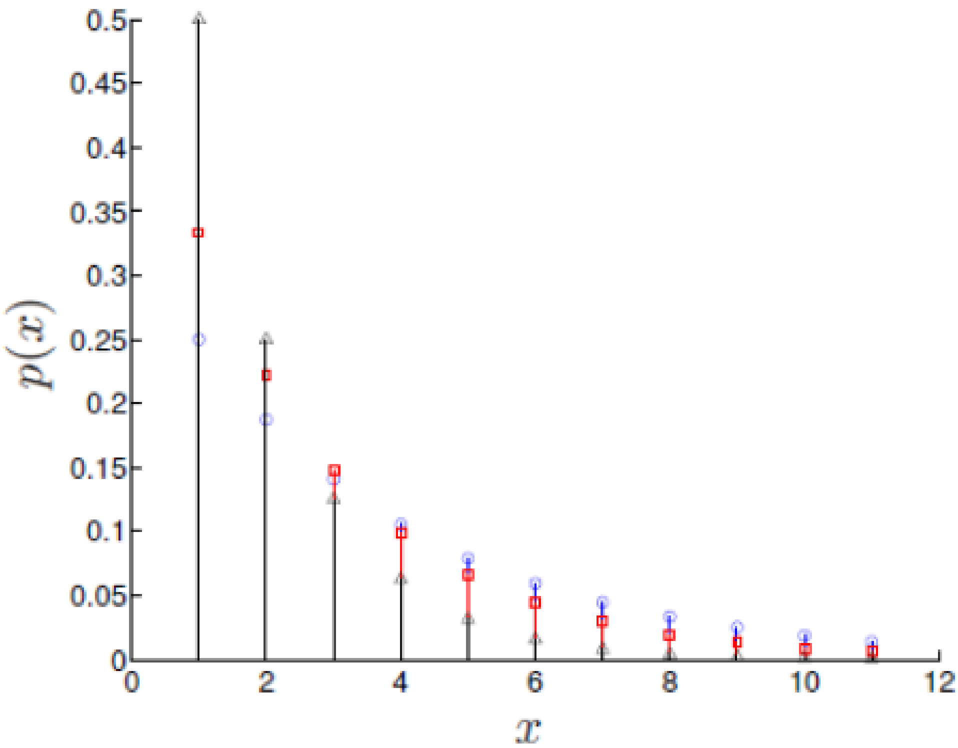


Figure 3.5: Plot of the $G\left(\frac{1}{4}\right), G\left(\frac{1}{3}\right)$ and $G\left(\frac{1}{2}\right)$ probability functions at the value $1, 2, 3, ..., 11$.

Remark 3.16. *The geometric distribution is a special case of the negative binomial distribution, where the number of successes is* 1. *The negative binomial distribution generalizes the geometric distribution to the case where there are multiple successes.*

The cumulative distribution function of a random variable, X, that is evaluated at a point x, can be defined as the probability that X will take a value that is lesser than or equal to x. It is also known as the distribution function. The formula for the geometric distribution cdf is given as follows:

$$P(X \leq k) = 1 - (1-p)^k,$$

for $k = 1, 2, 3, ...$

Proposition 3.4. *If $X \sim G(p)$, then*

$$E(X) = \frac{1}{p} \quad \textit{and} \quad Var(X) = \frac{1-p}{p^2}$$

Proof. We use the definition of expectation for discrete random variables:

$$\begin{aligned} E(X) &= \sum_{k \in X(\Omega)} kP(X = k) = \sum_{k=1}^{+\infty} k(1-p)^{k-1} p \\ &= \sum_{k=1}^{+\infty} kq^{k-1} p = p \sum_{k=1}^{+\infty} kq^{k-1} \quad \text{where} \quad q = 1-p \end{aligned}$$

we use a standard result from series summation. First, let's compute the sum of $\sum_{k=1}^{+\infty} q^k = \frac{q}{1-q} = \frac{1}{p}$ and by differentiating the sum of the geometric series, we have

$$\begin{aligned} &= p \frac{d}{dq} \left(\sum_{k=1}^{+\infty} q^k \right)' = p \frac{d}{dq} \left(\frac{q}{1-q} \right)' \\ &= p \left(\frac{1}{(1-q)^2} \right) = \frac{p}{p^2} \\ &= \frac{1}{p}. \end{aligned}$$

We have

$$\begin{aligned}
E\left(X^2\right) &= \sum_{k\in X(\Omega)} k^2 P(X=k) = \sum_{k=1}^{+\infty} k^2 (1-p)^{k-1} p \\
&= \sum_{k=1}^{+\infty} k^2 q^{k-1} p = p\sum_{k=1}^{+\infty} k^2 q^{k-1} \\
&= p\sum_{k=1}^{+\infty} (k+1-1) k q^{k-1} \\
&= p\sum_{k=1}^{+\infty} (k+1) k q^{k-1} - p\sum_{k=1}^{+\infty} k q^{k-1} \\
&= p\frac{d^2}{dq^2}\left(\sum_{k=1}^{+\infty} q^{k+1}\right) - p\frac{d}{dq}\left(\sum_{k=1}^{+\infty} q^{k}\right) \\
&= p\frac{d^2}{dq^2}\left(\frac{q^2}{1-q}\right) - p\frac{d}{dq}\left(\frac{q}{1-q}\right) \\
&= p\left(\frac{2}{(1-q)^3}\right) - p\left(\frac{1}{(1-q)^2}\right) \\
&= p\left(\frac{2}{p^3} - \frac{1}{p^2}\right) \\
&= p\frac{2-p}{p^3} \\
&= \frac{2-p}{p^2}.
\end{aligned}$$

Now, use the formula for variance

$$Var(X) = E\left(X^2\right) - (E(X))^2,$$

Substitute $E\left(X^2\right)$ and $E(X)$:

$$\begin{aligned}
Var(X) &= \frac{2-p}{p^2} - \frac{1}{p^2} \\
&= \frac{2-p-1}{p^2} \\
&= \frac{1-p}{p^2}.
\end{aligned}$$

□

There are two common ways to parameterize the geometric distribution:

- Number of Trials until First Success: X is the number of trials needed to get the first success. This is the parameterization described above.

- Number of Failures before the First Success: Sometimes the geometric distribution is defined in terms of the number of failures before the first success. In this case, if Y is the number of failures before the first success, then Y follows a geometric distribution with pmf:
$$P(Y = k) = (1 - p)^k . p$$
where $k = 1, 2, 3, ...$

3.1.7 Poisson distribution

The Poisson distribution is a discrete probability distribution that models the number of times an event occurs in a fixed interval of time or space, given that these events occur with a known constant mean rate and are independent of the time since the last event. This distribution was first introduced by Siméon Denis Poisson (1781–1840) and published together with his probability theory in his work Recherches sur la probabilité des jugements en matière criminelle et en matière civile (1837).
The Poisson distribution is used as a distribution of rare events, such as: Arrivals, Accidents, Number of misprints, Hereditary, Natural disasters like earth quake, etc.

Definition 3.15. *A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, and write $X \sim P(\lambda)$, if its probability mass function is given by*

$$P(X = k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!} & \text{if} \quad k = 0, 1, 2, 3, .. \\ 0 & \text{otherwise} \end{cases}$$

where:

- *λ is the average rate (mean) of occurrences in the interval,*

- *k is the number of events (a non-negative integer),*

- *e is the base of the natural logarithm (approximately 2.71828).*

Remark 3.17. *Poisson distribution also used in situations where "events" happen at certain points in time.*

Example 3.17. *a book editor might be interested in the number of words spelled incorrectly in a particular book. It might be that, on the average, there are* 5 *words spelled incorrectly in* 100 *pages. The interval is the* 100 *pages.*

Example 3.18. *Let X be a Poisson random variable with parameter with $\lambda = 4$. Find $P(X > 2)$.*
Solution: *we have*

$$\begin{aligned}
P(X > 2) &= 1 - P(X \le 2) \\
&= 1 - \left[e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} \right] \\
&= 1 - \left[e^{-4} \sum_{k=0}^{2} \frac{4^k}{k!} \right] \\
&= 1 - \left[e^{-4} \frac{4^0}{0!} + e^{-4} \frac{4^1}{1!} + e^{-4} \frac{4^2}{2!} \right] \\
&= 1 - e^{-4}(1 + 4 + 8) \\
&= 1 - 0.238 \\
&= 0.762.
\end{aligned}$$

Definition 3.16. *If X is a random variable which follows a Poisson distribution(i.e, $X \sim P(\lambda)$), The cumulative distribution function is given by*

$$F_X(x) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}, n \le x \le n + 1.$$

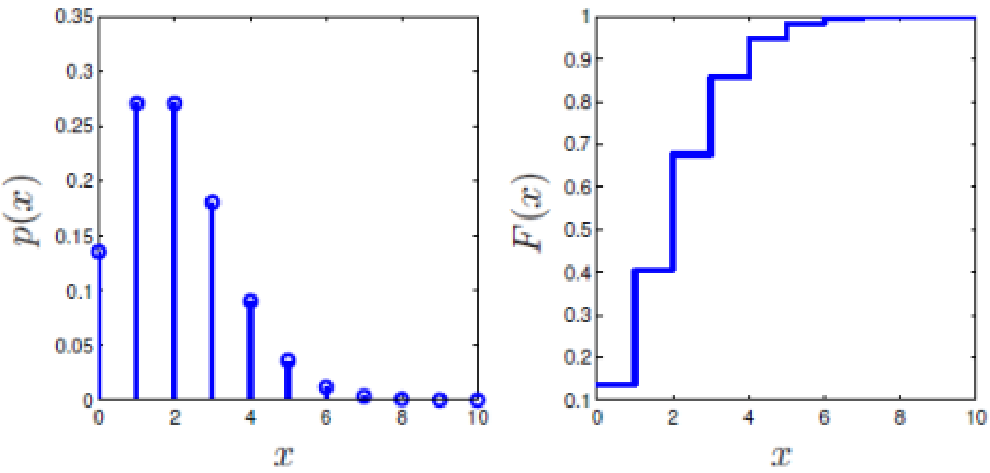


Figure 3.6: The probability mass function and the cumulative distribution of Poisson random variable with $\lambda = 2$.

Remark 3.18. *We note that since* $\sum_{k\in\mathbb{N}} \frac{\lambda^k}{k!} = e^{\lambda}$*, it is indeed true that* $\sum_{k=0}^{\infty} P(X=k) = 1$.

Proposition 3.5. *If* X *has a Poisson distribution with parameter* λ *(i.e,* $X \sim P(\lambda)$*), then*

$$E(X) = Var(X) = \lambda.$$

Proof. These results can be derived directly from the definitions of mean and variance of a random variable with discrete distribution

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kP(X=k) \\ &= \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{s=0}^{\infty} \frac{\lambda^s}{s!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda. \end{aligned}$$

$$\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 P(X=k) \\
&= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{s=1}^{\infty} (s+1) \frac{\lambda^s}{s!} \\
&= \lambda \left[\sum_{s=1}^{\infty} s e^{-\lambda} \frac{\lambda^s}{s!} + \sum_{s=1}^{\infty} e^{-\lambda} \frac{\lambda^s}{s!} \right] \\
&= \lambda(\lambda+1) \\
&= \lambda^2 + \lambda,
\end{aligned}$$

thus

$$\begin{aligned}
Var(X) &= E\left(X^2\right) - (E(X))^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda.
\end{aligned}$$

□

Example 3.19. *Suppose that a rare disease has an incidence of* 1 *in* 1000 *people per year. Assuming that members of the population are affected independently, find the probability of* k *cases in a population of* 10,000 *(followed over* 1 *year) for* $k = 0, 1, 2$.

Solution: *The expected value (mean)* :$\lambda = 0.001 \times 10000 = 10$

$$\begin{aligned}
P(X=0) &= \frac{(10)^0 e^{-10}}{0!} = 0.000454 \\
P(X=1) &= \frac{(10)^1 e^{-10}}{1!} = 0.00454 \\
P(X=2) &= \frac{(10)^2 e^{-10}}{2!} = 0.00227.
\end{aligned}$$

3.2 Continuous probability distribution

3.2.1 Uniform distribution

The continuous uniform distribution is a fundamental probability distribution characterized by its simplicity and the equal likelihood of outcomes within a defined interval.

Its simplicity makes it a useful reference point for understanding more complex distribution, and it serves various applications in statistics, engineering, computer science and other fields.

Definition 3.17. *A continuous random variable X is said to have a uniform distribution, over an interval* $[a,b]$ *and denoted* $X \sim \mathcal{U}([a,b])$, *if its probability density function is given by:*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$

where a and b are two constants(such that a is the minimum value and b is the maximum value).

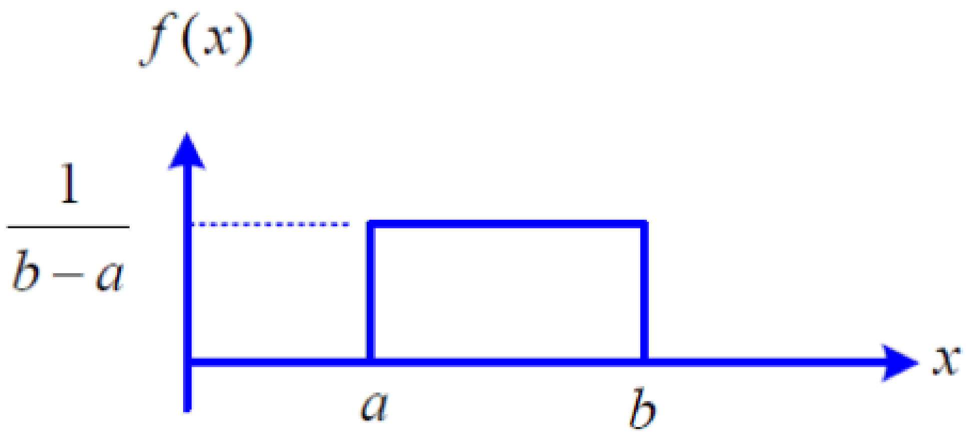


Figure 3.7: The uniform distribution for continous random variable.

Remark 3.19. *X takes any values on an interval* $[a,b]$ *that are equally probable.*

Proposition 3.6. *The cumulative distribution function of a uniform random variable X is given by*

$$F(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > b \end{cases}$$

Proof. We have

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

- for $x < a$,

$$F(x) = P(X \leq x) = 0.$$

- for $a \le x \le b$

$$F(x) = P(X \le x) = \int_a^x f(t)dt = \int_a^x \frac{1}{b-a}dt = \frac{x-a}{b-a}.$$

- for $x > b$,

$$F(x) = P(X \le b) = \frac{b-a}{b-a} = 1$$

□

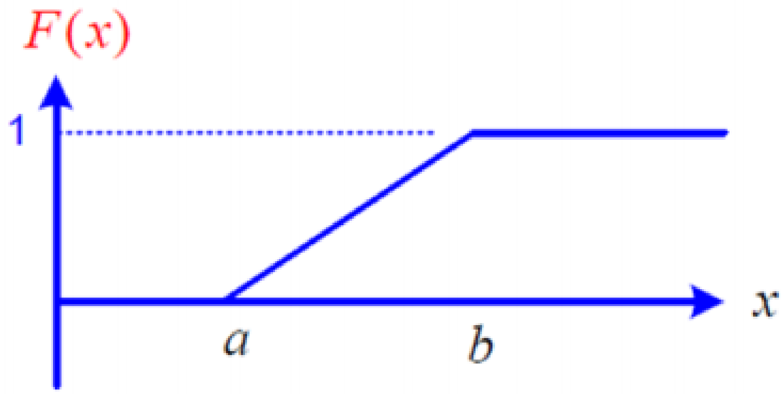


Figure 3.8: The cumulative uniform distribution for continuous random varaible.

Example 3.20. *Let* $X \sim \mathcal{U}([0,2])$, *calculate* $P\left(X \le \frac{3}{2}\right)$ *and* $P(0 \le X \le 2)$.
Solution:

$$P\left(X \le \frac{3}{2}\right) = \int_0^{3/2} \frac{1}{2}dx = \frac{1}{2}\, x|_0^{3/2} = \frac{3}{4}.$$
$$P(0 \le X \le 2) = \int_0^{2} \frac{1}{2}dx = \frac{1}{2}\, x|_0^{2} = 1.$$

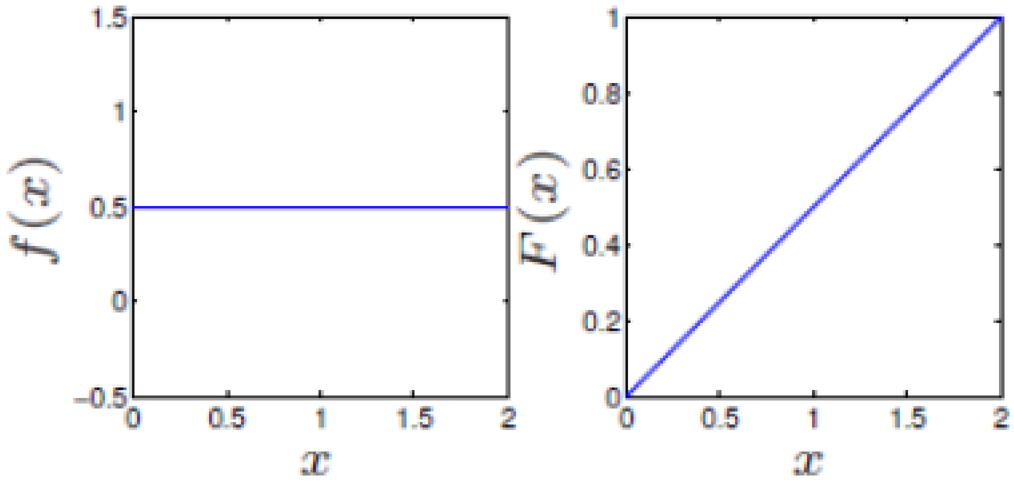


Figure 3.9: The uniform probability and the cumulative distribution functions for continuous r.v.: $x \in [0,2]$.

Theorem 3.5. *The mean of a continous uniform random variable defined over the support $a < x < b$ is*

$$E(X) = \frac{a+b}{2}.$$

Proof. We have

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b x \left(\frac{1}{b-a} \right) dx \\ &= \frac{1}{b-a} \int_a^b x dx = \frac{1}{2(b-a)} x^2 \Big|_a^b \\ &= \frac{1}{2(b-a)} \left(b^2 - a^2 \right) = \frac{1}{2(b-a)} (b-a)(b+a) \\ &= \frac{a+b}{2}. \end{aligned}$$

□

Theorem 3.6. *The variance of a continous uniform random variable defined over the support $a < x < b$ is given by*

$$Var(X) = \frac{(b-a)^2}{12},$$

and the standard deviation is

$$\sigma = \frac{b-a}{\sqrt{12}}$$

Proof. We have

$$Var(X) = E\left(X^2\right) - (E(X))^2$$

$$\begin{aligned}
E\left(X^2\right) &= \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_a^b x^2 \left(\frac{1}{b-a}\right)dx \\
&= \frac{1}{b-a}\int_a^b x^2 dx = \frac{1}{3(b-a)}\, x^3\Big|_a^b \\
&= \frac{1}{3(b-a)}\left(b^3 - a^3\right) = \frac{1}{2(b-a)}(b-a)\left(b^2+ab+a^2\right) \\
&= \frac{b^2+ab+a^2}{3}.
\end{aligned}$$

Thus

$$\begin{aligned}
Var(X) &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} \\
&= \frac{b^2-2ab+a^2}{3} = \frac{(b-a)^2}{12}.
\end{aligned}$$

□

Remark 3.20.

- *If* $X \in [-a, a]$, *then* $E(X) = 0$.
- *If* $X \in [a, b]$, *then* $E(X) \in [a, b]$.
- *In a uniform distribution, the symmetry means that the expected value and the median are indeed equal* $\frac{a+b}{2}$, *while there is no mode (no single value appears more frequently than others; all values in the interval have the same probability).*

Example 3.21. *If* X *is uniformly distributed with Mean* 1 *and Variance* $\frac{4}{3}$, *find* $P(X > 0)$.
Solution: *If* X *is uniformly distribution over* (a, b), *then*

$$E(X) = \frac{a+b}{2} \quad \text{and} \quad Var(X) = \frac{(b-a)^2}{12}$$

which implies that

$$\begin{aligned}
\frac{a+b}{2} = 1 &\quad \text{and} \quad \frac{(b-a)^2}{12} = \frac{4}{3} \\
a+b = 2 &\quad \text{and} \quad (b-a)^2 = 16 \\
a+b = 2 &\quad \text{and} \quad b-a = 4
\end{aligned}$$

then we get

$$a = -1 \quad \text{and} \quad b = 3$$

and the probability function of X is

$$f(x) = \begin{cases} \frac{1}{4} & \text{for} \quad -1 \leq x \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$P(X < 0) = \int_{-1}^{0} \frac{1}{4} dx = \frac{1}{4} x|_{-1}^{0} = \frac{1}{4}.$$

3.2.2 Exponential distribution

The continuous exponential distribution is a probability distribution often used to model the time until an event occurs, such as the time between arrivals in a Poisson process.

Definition 3.18. *A continuous random variable X is said to have an exponential distribution if its range is* $(0, \infty)$ *and its probability density function is given by*

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

for some positive λ *and denoted* $X \sim \mathcal{E}(\lambda)$.

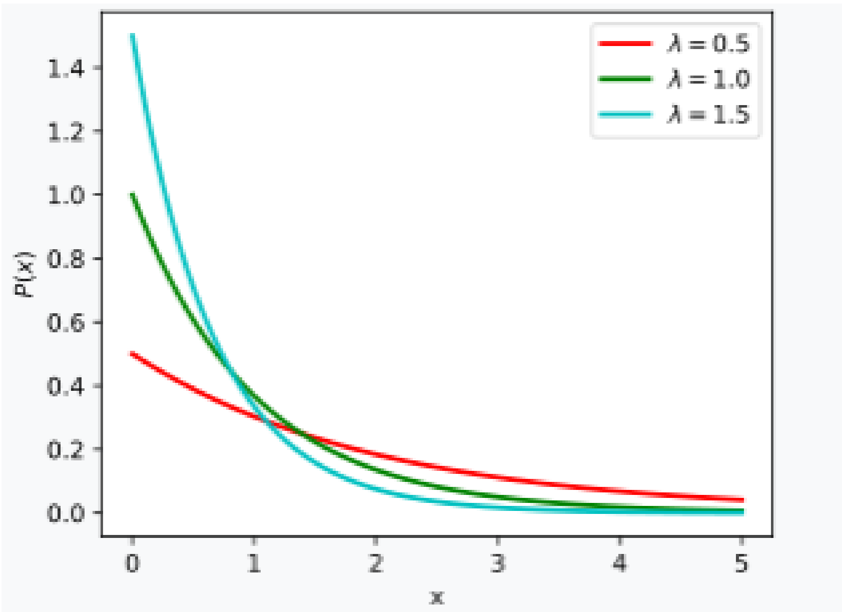


Figure 3.10: The exponentiel distribution of continuous random variable for $\lambda = 0.5, 1, 1.5$.

Remark 3.21.

- *If X is a random variable having the exponential distribution f , then*

$$P(a \leq X \leq b) = \int_a^b \lambda e^{-\lambda x} dx = \left(-e^{-\lambda b}\right) - \left(-e^{-\lambda a}\right) = e^{-\lambda a} - e^{-\lambda b}$$

for $0 \leq a \leq b < \infty$.

- *The area under the curve is*

$$\begin{aligned} &= \int_{-\infty}^{+\infty} f(x)dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx \\ &= \lambda \left[\frac{e^{-\lambda x}}{\lambda}\right]_0^{+\infty} = \frac{\lambda}{-\lambda}[0-1] \\ &= 1. \end{aligned}$$

Proposition 3.7. *The cumulative distribution function is given by*

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for} \quad x \geq 0 \\ 0 & \text{for} \quad x < 0. \end{cases}$$

where λ *is a rate constant and* $\lambda > 0$.

Proof. We have

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \lambda e^{-\lambda t} dt \\ &= \lambda \left[\frac{e^{-\lambda t}}{\lambda}\right]_0^{x} = \frac{\lambda}{-\lambda}\left[e^{-\lambda x} - 1\right] \\ &= 1 - e^{-\lambda x}. \end{aligned}$$

□

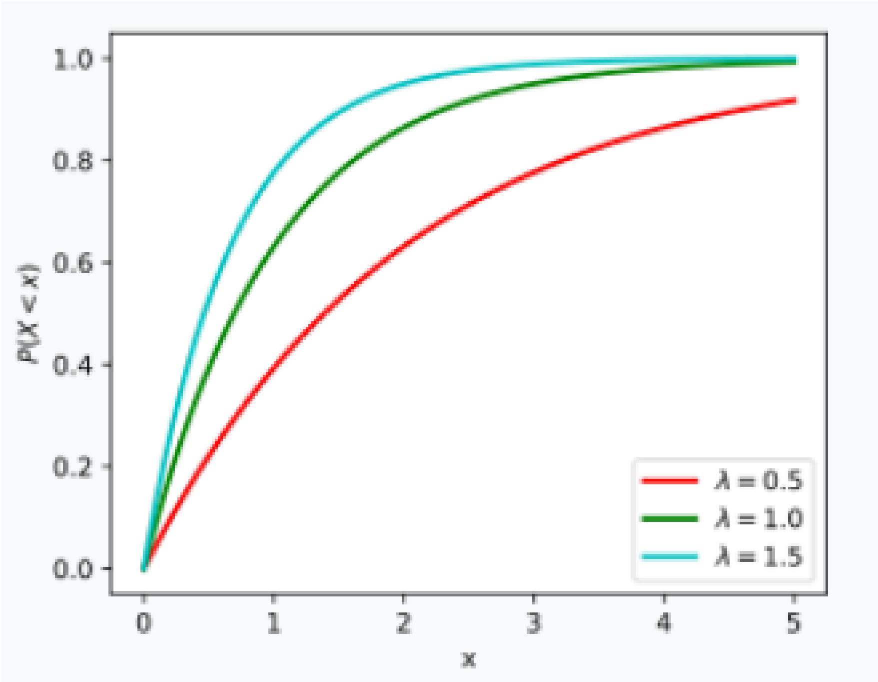


Figure 3.11: The cumulative exponentiel distribution of continuous randon variable for $\lambda = 0.5, 1, 1.5$.

Proposition 3.8. *If X is continuous random variable following exponential distribution, then the expected value, the variance and the standard deviation of r.v. X is given by*

$$E(X) = \frac{1}{\lambda},\ Var(X) = \frac{1}{\lambda^2} \quad and \quad \sigma = \frac{1}{\lambda}.$$

Proof. Substituting the probability density function into the expectation formula, we get

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{0}^{+\infty} x\lambda e^{-\lambda x} dx \\ &= \left[-xe^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx \\ &= 0 + \int_0^{+\infty} e^{-\lambda x} dx = \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{+\infty} \\ &= 0 - \left(-\frac{1}{\lambda}\right) = \frac{1}{\lambda}. \end{aligned}$$

We have

$$Var(X) = E\left(X^2\right) - (E(X))^2$$

and

$$E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx$$
$$= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x} dx$$
$$= 0 + \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}.$$

Thus

$$Var(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}.$$

□

Example 3.22. *On average, a pair of running shoes can last* 18 *months if used every day. The length of time running shoes last is exponentially distributed.*

- *What is the probability that a pair of running shoes last more than* 15 *months?*
- *On average, how long would six pairs of running shoes last if they are used one after the other?*
- *Eighty percent of running shoes last at most how long if used every day?*

Solution: *Let* X = *the amount of time (in months) running shoes can last.* $E(X) = 18$ *then* $\lambda = \frac{1}{18} = 0.056$.

- *The probability that a pair of running shoes last more then 15 months is*

$$P(X > 15) = 1 - P(X < 15)$$
$$= 1 - \left(1 - e^{-1/18.15}\right)$$
$$= e^{-0.84} = 0.4317.$$

- *Six pairs of running shoes would last for* $18 \times 6 = 108$ *months on average.*
- *let* k = *the* 80 *percentile,*

$$P(X < k) = 0.8$$
$$= 1 - e^{-1/18.k}$$

thus

$$e^{-1/18.k} = 1 - 0.8 = 0.2$$

which implies that

$$\begin{aligned} -\frac{1}{18}k &= \ln(2) \\ k &= 18\ln(2) \\ &= 28.97 \text{ months.} \end{aligned}$$

Remark 3.22. *The exponential distribution is often used to model the lengths of gaps between events occurring haphazardly (that is, quite at random, and with no memory) in time.*

- *Births in a hospital*
- *Passage of cars along a road*
- *Arrival of ships at a terminal*

There are close links with the Poisson distribution, which is used to model the number of such events occurring in a fixed time interval.

Example 3.23. *Let X be the number of events occurring in an interval of length t: then X has the Poisson distribution with mean t. Let T be the gap until the first event occurs. Then the events $\{X = 0\}$ and $\{T > t\}$ are identical. We note that*

$$\begin{aligned} P(X=0) &= e^{-\lambda t} \\ P(T>t) &= 1 - F(t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}. \end{aligned}$$

An important fact about the exponential distribution it that it has the memoryless property. Indeed

Proposition 3.9. *The exponential distribution is unique in that it is memoryless, meaning:*

$$P(X > s + t | X > s) = P(X > t) \text{ for all } t, s > 0.$$

Proof. The probability density function of X is

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

we have

$$P(X > k) = \int_k^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_k^{\infty} = e^{-\lambda k}$$

then

$$\begin{aligned} P(X > s+t | X > s) &= \frac{P(X > s+t \cap X > s)}{P(X > s)} \\ &= \frac{P(X > s+t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= P(X > t). \end{aligned}$$

□

Remark 3.23.

- *The sum of n independent and identically distributed (i.i.d) random variables, each following an exponential distribution with rate parameter λ, follows a gamma distribution.*

- *The exponential distribution is widely used in*

the field of reliability (i.e., Reliability deals with the amount of time a product lasts).

3.2.3 Normal distribution

The normal sistribution is a cornerstone of probability and statistics due to its mathematical properties and the central limit theorem. Its applications extend across various fields, making it an essential concept for statisticians and researchers. Understanding how and when to use the normal distribution enables more robust data analysis and inference.
The normal distribution, also known as the Gaussian distribution, is a continuous probability distribution characterized by its symmetric, bell-shape curve. It is defined by two parameters: the mean μ and the standard deviation σ.

Definition 3.19. *The continuous random variable X follows a normal distribution if its probability density function is defined as*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

where

- *e is the base of the natural logarithm.*
- *μ is the mean*
- *σ is the standard deviation.*

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$ reads as X is following normal distribution with mean μ and variance σ^2, are called parameters.

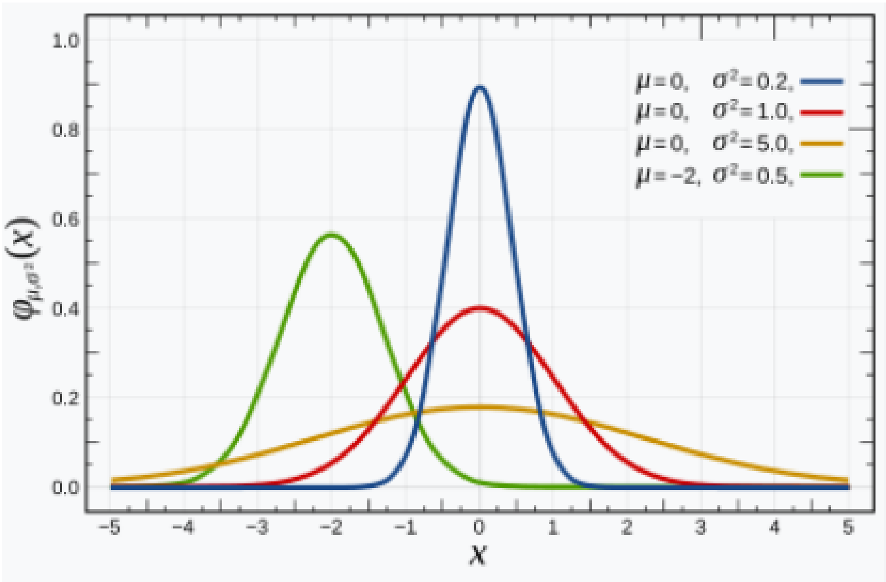


Figure 3.12: The normal distribution of continuous random variable.

Remark 3.24. *A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate.*

Characteristics:

1. All normal curves are bell-shaped with points on inflection at $\mu \pm \sigma$.
2. All normal curves are symmetric about the mean μ.
3. σ is called as shape parameter i.e., its shape is varied as σ varies.
4. The area under the curve is and over the x is unity (i.e., $\int_{-\infty}^{+\infty} f(x)dx = 1$).
5. All normal curves are positive for all x. That is $f(x) > 0$ for all x.

6. The limit of $f(x)$ as x goes to infinity is 0, and the limit of $f(x)$ as x goes to negative infinity is 0. That is

$$\lim_{x\to-\infty} f(x) = 0 \text{ and } \lim_{x\to+\infty} f(x) = 0.$$

7. The height of any normal curve is maximized at $x = \mu$ and its value is $f(x) = \frac{1}{\sigma^2\sqrt{2\pi}}$.

8. The shape of any normal curve depends on its mean μ and standard deviation σ.

Remark 3.25. ***Mean, Median, Mode:*** *in a normal distribution, these three measures of central tendency are all equal and located at the center of the distribution.*

Remark 3.26. *In empirical rule, Approximately:*

- 68.2% *of the data falls within one standard deviation* $(\mu \pm \sigma)$.

- 95.4% *falls within two standard deviations* $(\mu \pm 2\sigma)$.

- 99.7% *falls within three standard deviations* $(\mu \pm 3\sigma)$.

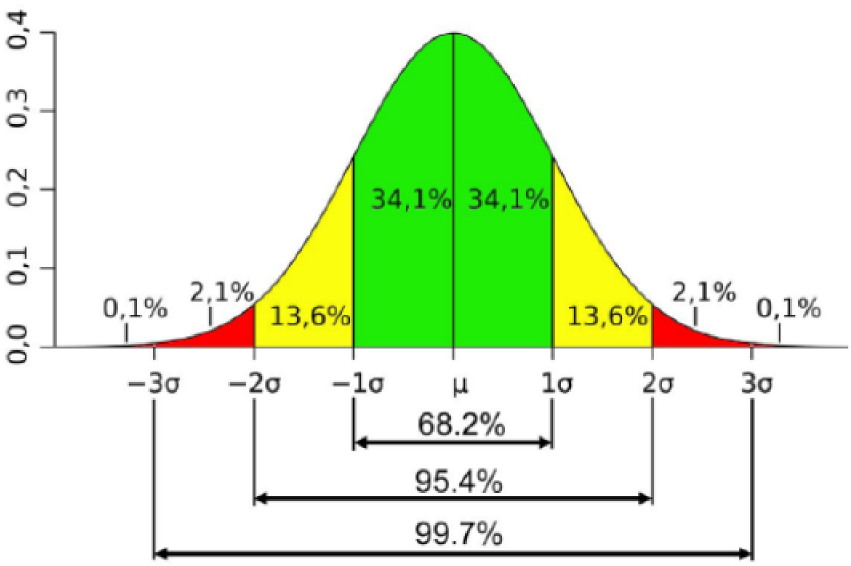


Figure 3.13: The normal curve.

Definition 3.20. *The cumulative distribution function of continous random variable is given by*

$$F(x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where erf is the error function.

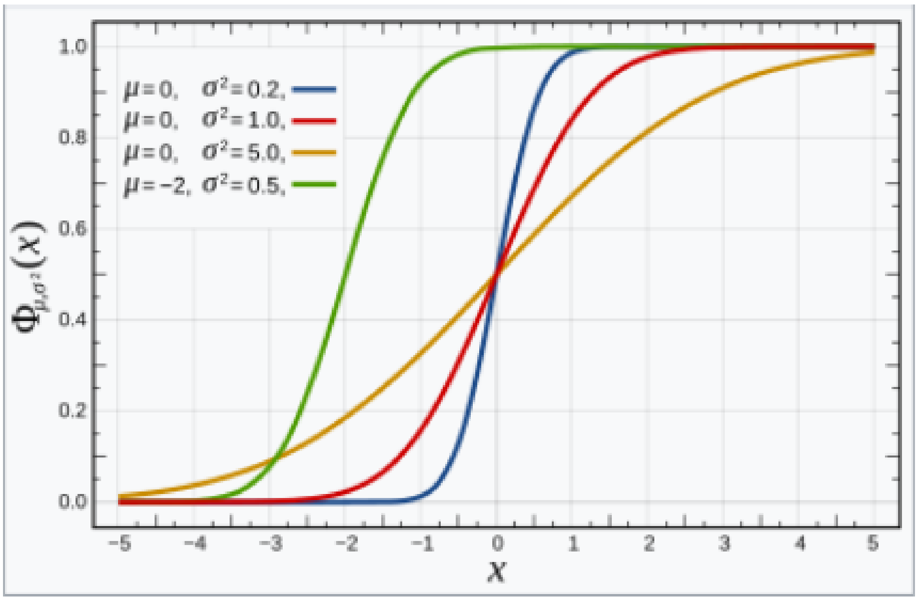


Figure 3.14: The cumulative normal distributiob of continuous random variable.

To facilitate the use of normal distribution, the following distribution known as the standard normal distribution was derived.

3.2.3.1 Standard normal distribution

The simplest case of a normal distribution is known as the standard normal distribution or unit normal distribution. This is a special case of the normal distribution when $\mu = 0$ and $\sigma^2 = 1$.

If X is a normal deviate with parameters μ and σ^2, then this X distribution can be rescaled and shifted via the formula:

$$Z = \frac{X - \mu}{\sigma}$$

to convert it to the standard normal distribution. This variate is also called the standardized form of X. It is described by this probability density function (or density):

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < +\infty$$

and its cumulative distribution function is usually denoted by

$$\Phi(Z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

and $\Phi(-z) = 1 - \Phi(z)$.

Properties of the standard normal distribution:

- Same as a normal distribution, but also mean is zero, variance is one, standard deviation is one.

- Area under the standard normal distribution curve (i.e., $\Phi(x)$) have been tabulated in various ways. The most common ones are the areas between $Z = 0$ and a positive value of Z.

- Given normal distributed random variable X with mean μ and standard deviation σ

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

  which implies

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right),$$

  hence $Z \sim \mathcal{N}(0,1)$ and $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ (means that the cdf of any normal distribution can be obtained from tables of $\mathcal{N}(0,1)$).

Example 3.24. *Let X be a random variable following normal distribution $\mathcal{N}(5,4)$. Find the following $P(1 < X < 7)$.*
Solution : We first consider $Z := \frac{X-5}{\sqrt{4}}$. Then $Z \sim \mathcal{N}(0,1)$:

$$\begin{aligned} P(1X < 7) &= P(\frac{1-5}{2} < \frac{X-5}{2} < \frac{7-5}{2}) \\ &= P(-2 < Z < 1) \\ &= \Phi(1) - \Phi(-2) \\ &= \Phi(1) + \Phi(2) - 1. \end{aligned}$$

according to the table, $\Phi(1) \approx 0.8413$ and $\Phi(2) \approx 0.9772$, thus

$$P(1 < X < 7) \approx 0.8185.$$

3.2.4 log-normal distribution

The log-normal distribution is a continuous probability of a random variable whose logarithm is normally distributed. It is a convenient and useful model for measurements in exact and engineering sciences, as well as medicine, economics and other topics (e.g., energies, concentrations, lengths, prices of financial instruments, and other metrics).

Definition 3.21. *A random variable X is said to follow a log-normal distribution if $Y = \ln(X)$ is normally distributed.*

This means:

$$\text{If } Y \sim \mathcal{N}\left(\mu, \sigma^2\right), \text{ then } X = e^Y \sim \mathcal{LN}\left(\mu, \sigma^2\right).$$

Remark 3.27. *A random variable which is log-normally distributed takes only positive real values.*

Definition 3.22. *The probability of a log-normal distribution is given by*

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

for $x > 0$, where

- *μ is the mean of the natural logarithm of the variable.*

- *σ is the standard deviation of the natural logarithm of the variable.*

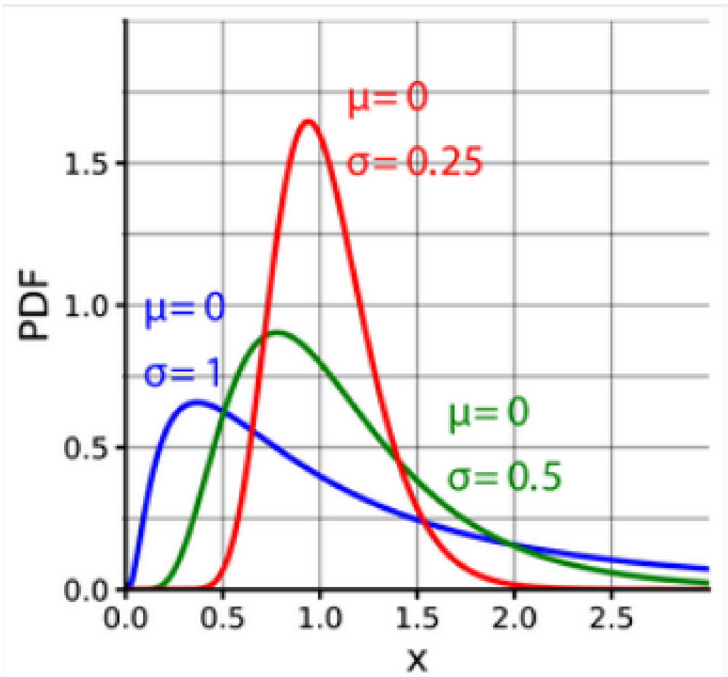


Figure 3.15: The log-normal distribution for identical parameter μ but differing parameters σ.

Definition 3.23. *The cumulative distribution function of log-normal distribution can be expressed as:*

$$F(x;\mu,\sigma) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{\ln x - \mu}{\sigma\sqrt{2}}\right) \textit{ for } x > 0.$$

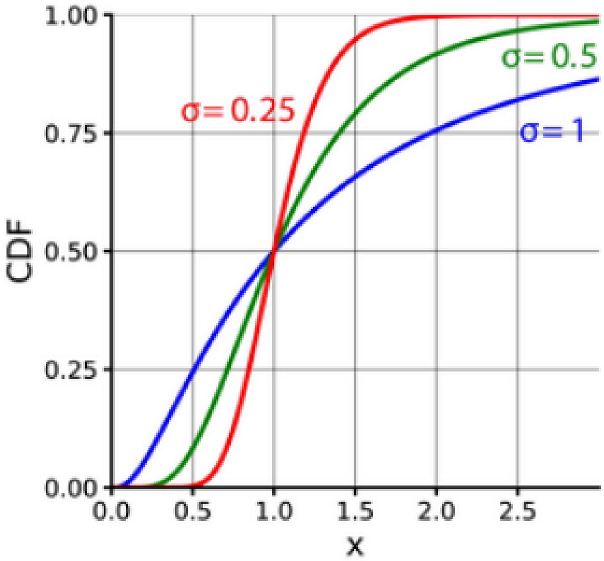


Figure 3.16: The log-normal cumulative distribution for $\mu = 0$.

3.2.4.1 Properties of log-normal distribution

1. **Mean:** The mean of a log-normal random variable X is given by

$$E(X) = e^{\left(\mu + \frac{\sigma^2}{2}\right)}.$$

2. **Variance:** the variance of X is

$$Var(X) = \left(e^{\sigma^2} - 1\right) e^{\left(2\mu + \sigma^2\right)}.$$

3. **Skewness:** the log-normal distribution is positively skewed, meaning it has a longer tail on the right side.

4. **Median:** The median of X is e^{μ}.

5. **Mode:** the mode of X is $e^{\left(\mu - \sigma^2\right)}$.

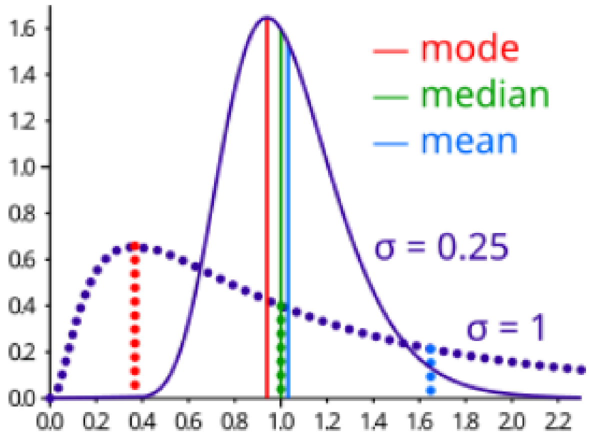


Figure 3.17: Comparison of mean, median and mode of two log-normal distributions with different skewness.

Example 3.25. *If Y is normally with $\mu = 0$ and $\sigma = 1$, then $X = e^Y$ will following a log-normal distribution with the same parameters.*

3.2.5 Gamma distribution

The Gamma distribution is a continuous probability that is commonly used in statistics, particularly in the fields of queuing models, reliability analysis and Bayesian statistics. It utilizes the gamma function to define its Probability density function and ensure proper normalization

3.2.5.1 Gamma function

The gamma function denoted as $\Gamma(n)$, is a fundamental function in mathematics that extends the concept of factorials to complex and real number arguments. It is defined for all complex numbers except for the non-positive integers. The gamma function is defined as

$$\Gamma(n) = \int_0^n t^{n-1} e^{-t} dt, \text{ for } n > 0.$$

3.2.5.1.1 Properties of gamma function

1. **Relation factorials**: for positive integers:

$$\Gamma(n) = (n-1)!.$$

This means

$$\Gamma(1) = 0!, \Gamma(2) = 1!, \Gamma(3) = 2! = 2, \text{ and so on.}$$

2. **Recurrence relation:** The gamma function satisfies the recurrence relation:

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

This property allows for easy computation of Γ for values that are not positive integers.

3. **Values for half-integers:**

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$
$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n)!}{4^n n! \sqrt{\pi}}.$$

Example 3.26. *Calculate* $\Gamma(5)$, $\Gamma\left(\frac{5}{2}\right)$.
Solution: *We have* $\Gamma(n) = (n-1)!$, *then*

$$\Gamma(5) = (5-1)! = 4! = 24,$$

and $\Gamma(n) = (n-1)\Gamma(n-1)$, *thus*

$$\begin{aligned}\Gamma\left(\frac{5}{2}\right) &= \frac{3}{2}\Gamma\left(\frac{3}{2}\right)\\ &= \frac{3}{2}\left(\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right)\\ &= \frac{3}{4}\sqrt{\pi}.\end{aligned}$$

3.2.5.2 Definitions and properties

Definition 3.24. *The probability density function of the gamma distribution is given by*

$$f(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k\Gamma(k)} \text{ for } x > 0,$$

where $\Gamma(k)$ *is the gamma function, which generalizes the factorial function.*

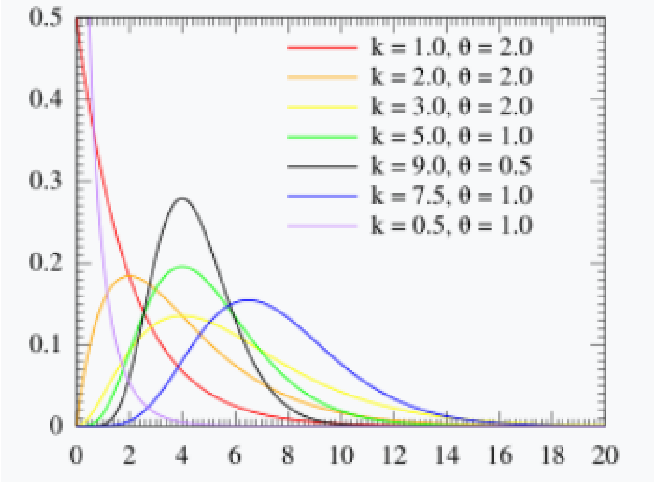


Figure 3.18: The Gamma density function.

Definition 3.25. *The cumulative distribution function is given by*

$$\begin{aligned} F(x;k,\theta) &= \int_0^x f(t;k,\theta)dt \\ &= \frac{1}{\Gamma(k)} \int_0^x t^{k-1} e^{-\frac{t}{\theta}} dt, \end{aligned}$$

which can be computed using special functions, particularly the regularized incomplete gamma function.

$$F(x;k,\theta) = \frac{\gamma\left(k, \frac{x}{\theta}\right)}{\Gamma(k)}.$$

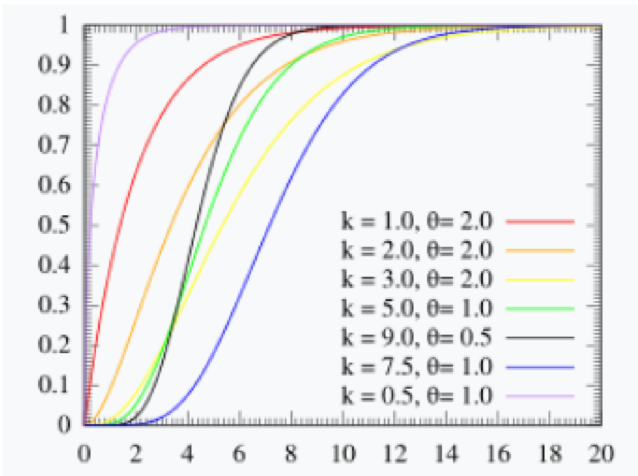


Figure 3.19: The cumulative Gamma distribution function.

Properties:

- **Mean:** The mean of a gamma distribution is $\mu = k\theta$.
- **Variance:** The variance is $\sigma^2 = k\theta^2$.
- **Skewness:** The skewness of the gamma distribution only depends on its shape parameter k, and it is equal to $2/\sqrt{k}$.

Remark 3.28.

- *When $k = 1$, the gamma distribution becomes the exponential distribution.*
- *When k is a positive integer, the gamma distribution can be viewed as the sum of k independent exponential random variables.*

3.2.6 Beta distribution

The Beta distribution is a versatile continuous probability distribution defined on the interval $[0,1]$. It has been applied to model the behavior of random variables limited to intervals of finite length in a wide variety of disciplines.
The beta distribution is a suitable model for the random behavior of percentages and proportions, it commonly used in Bayesian statistics, modeling proportions, and other areas.

Definition 3.26. *The Beta distribution is defined by its probability density function*

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

for $0 < x < 1$, where

- *$\alpha > 0$ is the shape parameter (also known as the first parameter).*
- *$\beta > 0$ is the shape parameter (also known as the second parameter).*
- *$B(\alpha,\beta)$ is the Beta function, defined as*

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

where Γ is the gamma function.

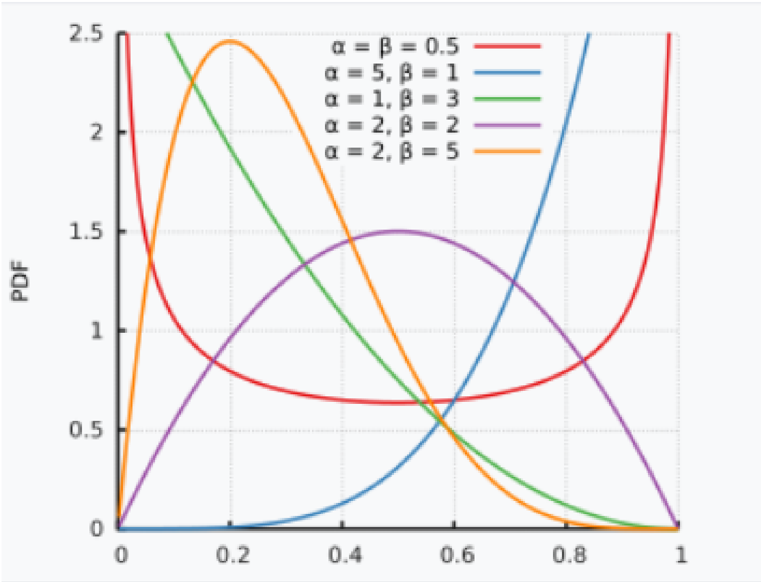


Figure 3.20: The Beta density function for differents values of α and β .

A random variable X beta-distributed with parameters α and β will be denoted: $X \sim Beta(\alpha, \beta)$.

Remark 3.29.

1. *From the relation between B and Γ we can defined The beta distribution as*

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 < x < 1.$$

2. *The beta function B is a normalization constant to ensure that the total probability is* 1.

3. $B(\alpha,\beta) = B(\beta,\alpha)$.

4. $B(1,1) = 1, B\left(\frac{1}{2},\frac{1}{2}\right) = \pi$.

Definition 3.27. *The cumulative distribution function of the Beta distribution is given by*

$$\begin{aligned} F(x;\alpha,\beta) &= \int_0^x f(t;\alpha,\beta)dt \\ &= \frac{1}{B(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt \\ &= \frac{B(x;\alpha,\beta)}{B(\alpha,\beta)} = I_x(\alpha,\beta) \end{aligned}$$

where $B(x;\alpha,\beta)$ is the incomplete beta function and $I_x(\alpha,\beta)$ is the regularized incomplete beta function.

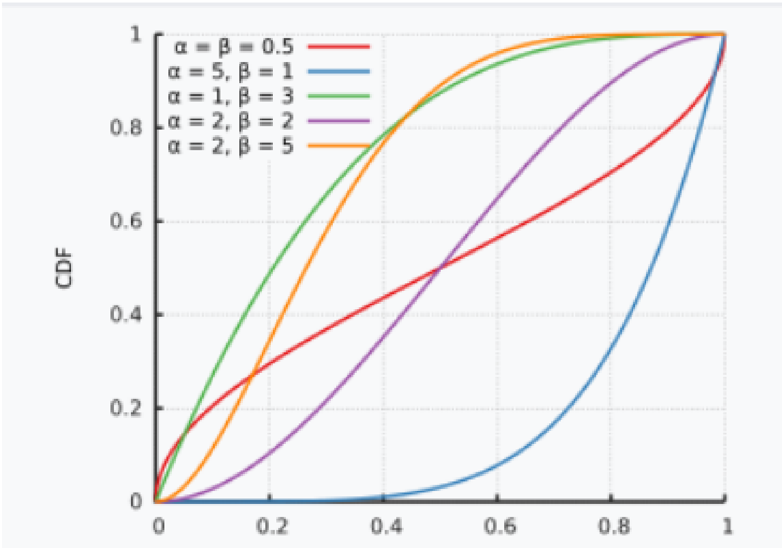


Figure 3.21: The cumulative Beta distribution function for differents values of α and β.

3.2.6.1 Properties of Beta distribution

1. **Mean:** The mean of a Beta-distributed random variable X is

$$E(X) = \frac{\alpha}{\alpha+\beta}.$$

2. **Variance:** The variance is

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

3. **Skewness:** The skewness of the Beta distribution is

$$Skewness = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta)^2(\alpha+\beta+2)}.$$

4. **Kurtosis:** The kurtosis is given by

$$Kutosis = \frac{6\left(\alpha^2+\beta^2\right)+12\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+2)} - 3.$$

5. **Support:** The Beta distribution is defined only on the interval $[0,1]$.

Remark 3.30.

1. *When* $\alpha = 1$ *and* $\beta = 1$*, the Beta distribution reduces to a uniform distribution over* $[0,1]$.(

2. *The Beta distribution is the conjugate prior for the success probability of Bernoulli of Bernoulli or Binomial distribution.*

Example 3.27. *Suppose we want to model the probability of success in a project completion scenario where we expect the probability of success to be skewed towards lower values. We might choose parameters* $\alpha = 2$ *and* $\beta = 5$.
We have

$$E(X) = \frac{2}{2+5} = \frac{2}{7}.$$

$$Var(X) = \frac{2.5}{(2+5)^2(2+5+1)} = \frac{10}{49.8} \simeq 0.0255.$$

3.2.7 Cauchy distribution

The Cauchy distribution, or the Lorentzian distribution named after Augustin-Louis Cauchy, is a continuous probability distribution that is the ratio of two independent normally distributed random variables if the denominator distribution has mean zero. It is one of the few stable distributions with a probability density function that can be expressed analytically, the others being the normal distribution and the Lévy distribution.

Definition 3.28. *The Cauchy distribution is defined by its probability density function*

$$f(x, x_0, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-x_0}{\gamma}\right)^2\right)} \quad for \quad -\infty < x < +\infty$$

where

- x_0 *is the location parameter (the median of the distribution).*

- $\gamma > 0$ *is the scale parameter (related to the spread of the distribution).*

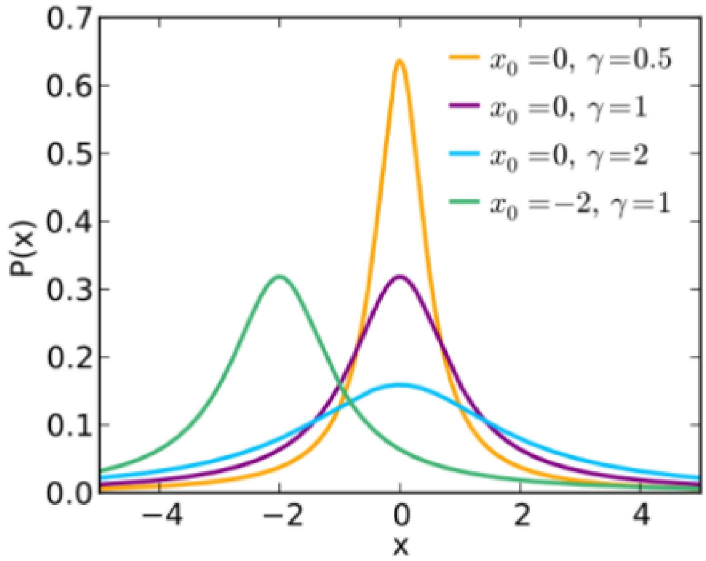


Figure 3.22: The Cauchy density function for differents values of x_0 and γ .

The standard Cauchy distribution is also a special case of the Student's $t-$ distribution with degrees of freedom $\nu = n - 1 = 1$

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\pi(1+x^2)}$$

where $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Definition 3.29. *The cumulative distribution function of the Cauchy distribution is given by*

$$F(x, x_0, \gamma) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x - x_0}{\gamma}\right).$$

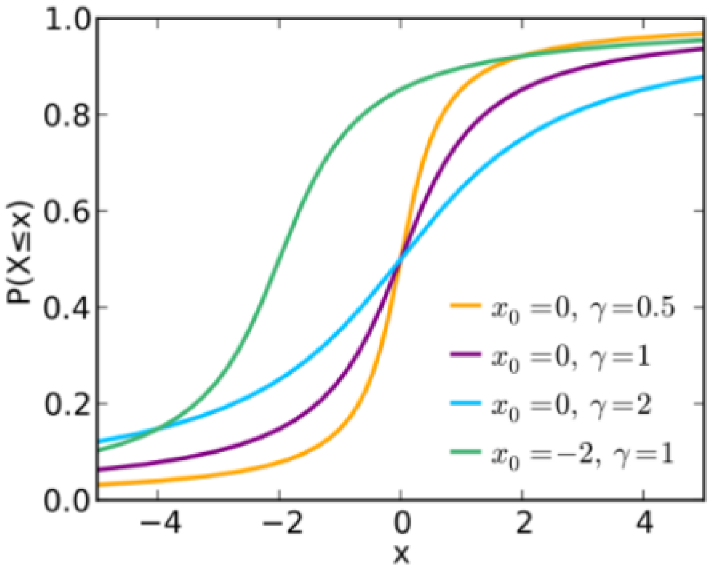


Figure 3.23: The cumulative Cauchy distribution for differents values of x_0 and γ.

3.2.7.1 Properties of Cauchy distribution

1. **Mean and Variance:** The Cauchy distribution does not have a defined mean and variance. This is a key characteristic that distinguishes it from many other distribution.

2. **Median:** The median of the Cauchy distribution is equal to the location parameters x_0.

3. **Mode:** The mode is also x_0.

4. **Symmetric:** The Cauchy distribution is symmetric about the line $x = x_0$. (This is due to the pdf being an even function about $x = x_0$).

5. **Heavy tails:** The tails of the Cauchy distribution decay as $\frac{1}{x^2}$, which means it can produce extreme values. This property makes useful for modeling outliers.

6. **Stability:** The Cauchy distribution is a stable distribution, which means that a linear combination of independent Cauchy-distributed random variables is also Cauchy-distributed.

Example 3.28. *If we have a Cauchy distribution with $x_0 = 0$ and $\gamma = 1$, the pdf simplifies to*

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

This distribution has a peak at 0 and tails that extend infinitely in both directions.

3.2.8 Chi-square distribution

Chi-square (χ_n^2) distributions (of Pearson) are a family of continuous probability distributions that arises in inferential statistics, particularly in the context of hypothesis testing and confidence intervals. They are commonly used in tests of independence in contingency tables, goodness of fit tests and in estimating the variance of a population. The Chi-squared distribution (also chi-square or χ^2-distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. It is characterized by its degrees of freedom, which is typically equal to the number of variables squared.

Definition 3.30. *Let $X_1,...,X_k$ be independent random variables with standard normal distribution $\mathcal{N}(0,1)$. Let*

$$X = \sum_{i=1}^{k} X_i^2,$$

then the sum of their squares X is distributed according to the Chi-squared distribution with k degrees of freedom. This is usually denoted as $X \sim \chi_k^2$.

Definition 3.31. *The probability density function of chi-squared distribution is*

$$f(x;k) = \begin{cases} \frac{x^{k/2-1}e^{-x/2}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} & \text{if} \quad x > 0, \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma\left(\frac{k}{2}\right)$ denotes the gamma function, which has closed-form values for integer k.

Example 3.29. *The graphs below show examples of chi-square distributions with different values of k.*

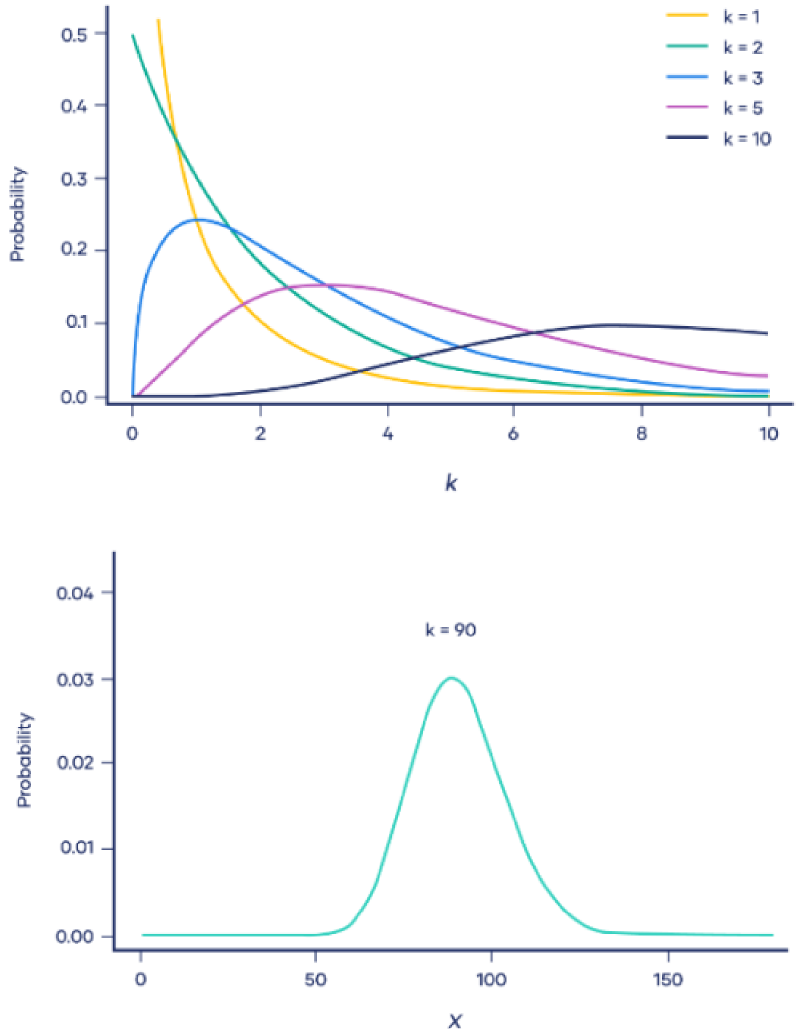


Figure 3.24: The chi-square distribution function χ_k^2 for different valus of k.

We can see how the shape of a chi-square distribution changes as the degrees of freedom (k) *increase by looking at graphs of the chi-square probability density function. For example:*

- *When k is 1 or 2, the chi-square distribution is a curve shaped like a backwards "J", the curve starts out high and then drops off, meaning that there is a high probability that X^2 is close to zero.*

- *When k is greater than two, the chi-square distribution is hump-shaped. The curve starts out low, increases, and then decreases again. There is low probability that X^2 is very close to or very far from zero. The most probable value of X^2 is $X^2 - 2$. When k is only a* *bit greater than two, the distribution is much longer on the right side of its peak than its left.*

- *As k increases, the distribution looks more and more similar to a normal distribution. In fact, when k is 90 or greater, a normal distribution is a good approximation of the chi-square distribution.*

Remark 3.31. *The chi-squared distribution χ_k^2 is a special case of the gamma distribution. If $X \sim \chi_k^2$ then $X \sim Gamma\left(\alpha = \frac{k}{2}, \theta = 2\right)$ (where α is the shape parameter and θ the scale parameter of the gamma distribution).*

Definition 3.32. *The cumulative distribution function is*

$$F(x;k) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} = P\left(\frac{k}{2}, \frac{x}{2}\right),$$

where $\gamma(s,t)$ is the lower incomplete gamma function and $P(s,t)$ is the regularized gamma function.

Remark 3.32. *In a special case of $k = 2$ this function has the simple form:*

$$F(x;1) = 1 - e^{-\frac{x}{2}}$$

which can be easily derived by integrating $f(x;1) = 1 - e^{-\frac{x}{2}}$ directly. The integer recurrence of the gamma function makes it easy to compute $F(x;k)$ for other small, even k.

3.2.8.1 Properties of chi-square distributions

Property	Value
Mean	k
Mode	$k-2$ (when $k > 2$)
Variance	$2k$
Standard deviation	$\sqrt{2k}$
Range	0 to ∞
Symmetry	Asymmetrical (right-skewed), but increasingly symmetrical as k increases.

Proposition 3.10. *The sum of independent chi-square random variables is a Chi-square random variable: If X_i are mutually independent and $X_i \sim \chi^2(k_i)$ for $i = 1, \dots, n$, then*

$$\sum_{i=1}^{n} X_i \sim \chi^2\left(\sum_{i=1}^{n} k_i\right).$$

Example 3.30. *Let X be a chi-square random variable with 3 degrees of freedom. Compute the following probability:*

$$P(0.35 \le X \le 7.81)$$

Solution: *First of all, we need to express the above probability in terms of the distribution function of X :*

$$\begin{aligned} P(0.35 \le X \le 7.81) &= P(X \le 7.81) - P(X \le 0.35) \\ &= F(7.81) - F(0.35) \\ &= 0.95 - 0.05 \\ &= 0.90. \end{aligned}$$

Example 3.31. *Let X_1 and X_2 be two independent normal random variables having mean $\mu = 0$ and variance $\sigma^2 = 16$. Compute the following probability:*

$$P\left(X_1^2 + X_2^2 > 8\right).$$

Solution: *First of all , the two variables X_1 and X_2 can be written as*

$$X_1 = 4Z_1, X_2 = 4Z_2$$

where Z_1 and Z_2 are two standard normal random variables. Thus, we can write

$$\begin{aligned} P\left(X_1^2 + X_2^2 > 8\right) &= P\left(16Z_1^2 + 16Z_2^2 > 8\right) \\ &= P\left(Z_1^2 + Z_2^2 > \frac{8}{16}\right) \\ &= P\left(Z_1^2 + Z_2^2 > \frac{1}{2}\right) \\ &= 1 - F_Y\left(\frac{1}{2}\right) \\ &= 0.2212. \end{aligned}$$

where $F_Y\left(\frac{1}{2}\right)$ is the distribution function of a Chi-square random variable Y with 2 degrees of freedom, evaluated at the point $y = \frac{1}{2}$.

3.2.9 Student distribution

In statistics, the t distribution was first derived as a posterior distribution in 1876 by Helmert and Lüroth and it was developed by Willieam S. Gosset (1908) in his work on "the probable error of a mean," published by him under the nom de plume of Student. Further developments continued with the contributions of Fisher (1925) and others later. The t–Student distribution, often simply the t distribution, is a continuous probability distribution that generalizes the standard normal distribution. It plays a role in a number of widely used statistical analyses, including Student's t test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis.
It is primarily used in statistics when estimating population parameters when the sample size small and the population standard deviation is unknown.

Definition 3.33. *If $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi^2(\nu)$ are independent, then the random variable:*

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

follows a t–distribution with ν degrees of freedom. We write $T \sim t(\nu)$. The probability density function of T is:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}.\left(\frac{t^2}{\nu}+1\right)^{-\frac{\nu+1}{2}} \quad \textit{for} \quad -\infty < t < \infty.$$

where ν is the number of degrees of freedom and Γ is the gamma function.

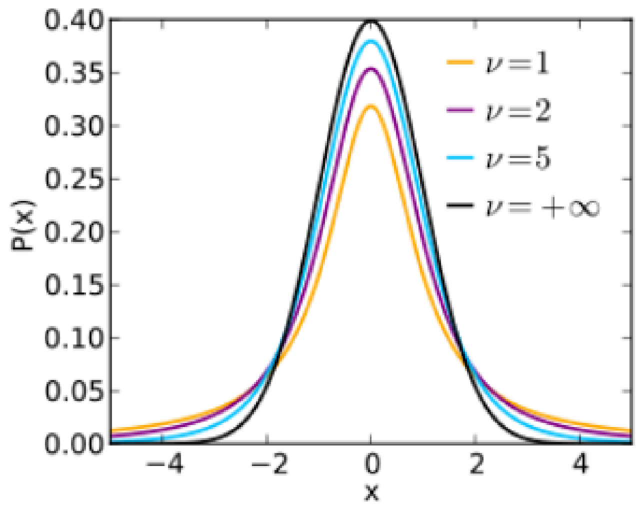


Figure 3.25: The Studen density function for $\nu = 1, 2, 5, +\infty$.

This may also be written as

$$f(t) = \frac{1}{\sqrt{\nu} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} . \left(\frac{t^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}},$$

where B is the beta function. In particular for integer valued degrees of freedom:

- For $\nu > 1$ and even,

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} = \frac{1}{2\sqrt{\nu}} . \frac{(\nu-1).(\nu-3)...5.3}{(\nu-2).(\nu-4)...4.2}$$

- For $\nu > 1$ and odd,

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} = \frac{1}{\pi\sqrt{\nu}} . \frac{(\nu-1).(\nu-3)...5.3}{(\nu-2).(\nu-4)...4.2}$$

3.2.9.1 Properties of the t−student distribution

1. **Degree of freedom** (df)**:** the shape of the t−distribution depends on the degrees of freedom, which is typically related to sample size (df= $n-1$, where n is the sample size). Note that the t−distribution becomes closer to the normal distribution as df increases.

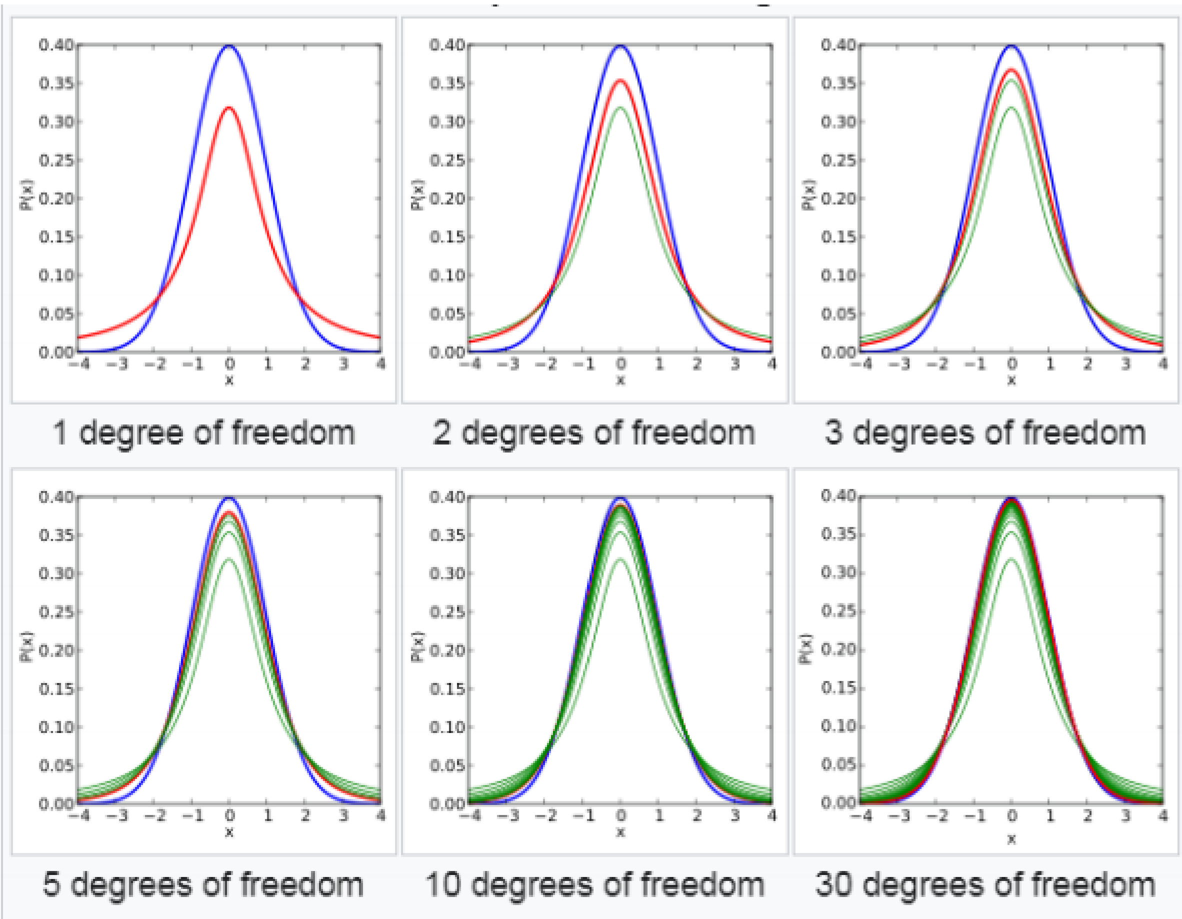


Figure 3.26: Density of t-distribution (red) for 1, 2, 3, 4, 5, 10 and 30 degrees of freedom compared to the standard normal distribution (blue). Previous plots shown in green.

2. **Symmetry:** The $t-$distribution is symmetric around zero.

3. **Heavier tails:** Compared to the normal distribution, the $t-$ distribution has heavier tails, meaning it provides more probability of extreme values. This accounts for the increased uncertainly when estimating the population mean with a small sample.

4. **Mean; Median and Mode:** All are equal to zero.

5. **Variance:** The variance of the $t-$distribution is greater than one and is given by $\frac{\nu}{\nu-2}$ for $\nu > 2$. This variance decreases as the degrees of freedom increase.

6. **Use in Hypothesis testing:** The $t-$distribution is commonly used in $t-$tests to determine if there are significant differences between the means of two groups.

Definition 3.34. *The cumulative distribution function (cdf) can be written in terms of I, the regularized incomplete beta function. For $t > 0$,*

$$F(t) = \int_{-\infty}^{t} f(u)du = 1 - \frac{1}{2} I_{x(t)}\left(\frac{\nu}{2}, \frac{1}{2}\right),$$

where $x(t) = \frac{\nu}{t^2+\nu}$.

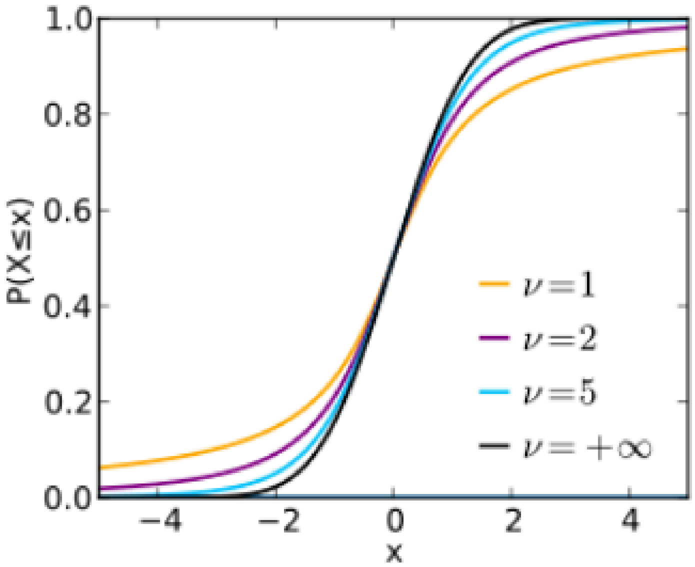


Figure 3.27: The cumulative t-student distribution function for differents values of ν.

Other values would be obtained by symmetry. An alternative formula, valid for $t^2 < \nu$, is

$$\int_{-\infty}^{t} f(u)du = \frac{1}{2} + t \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} F_1^{(2)}\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{t^2}{\nu}\right),$$

where $F_1^{(2)}(.,.;.;.)$ is a particular instance of the hypergeometric function.

Example 3.32. *Suppose a researcher wants to study the average height of a particular plant species. He takes a sample of* 10 *plants and measure their heights (in cm).*
Sample heights: 50, 52, 48, 53, 51, 49, 50, 52, 50, 51.

- *Calculate the sample mean:*

$$\overline{x} = \frac{50 + 52 + 48 + 53 + 51 + 49 + 50 + 52 + 50 + 51}{10} = 50.6$$

- *Calculate the sample standard deviation:*

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$
$$= \sqrt{\frac{(50-50.6)^2 + (52-50.6)^2 + \ldots + (51-50.6)^2}{10-1}}$$
$$\simeq 1.3$$

- *Use the* $t-$*distribution to calculate a confidence interval: For a* 95% *confidence level with* $n-1=9$ *degrees of freedom, (it's approximately* 2.262*).*

- *Calculate the margin of error (*ME*):*

$$ME = t \times \frac{s}{\sqrt{n}} = 2.262 \times \frac{1.3}{\sqrt{10}} \approx 0.93$$

- *Construct the confidence interval:*

$$CI = x \pm ME = 50.6 \pm 0.93 = (49.67, 51.53).$$

you would conclude that you are 95% *confident that the average height of the plant species in the population falls between* 49.67 *cm and* 51.53 *cm. The* $t-$*distribution is key here because the sample size is small.*

3.2.10 Fisher distribution

The Fisher distribution is named after the British statistician Ronald A. Fisher, who developed it in the early 20th Century, where he introduced the distribution in the context of his research on analysis of variance. His methods were aimed at understanding the variance within and between groups in experimental data, particularly in agriculture.
The Fisher distribution also known as the $F-$distribution or $F-$ratio, is a continuous probability distribution that arises in the context of analysis of variance (ANOVA) and regression analysis. It is used primarily to compare variances across different groups.

Definition 3.35. *If* U *and* V *are independent chi-square random variables with* d_1 *and* d_2 *degrees of freedom, respectively, then the random variable* X *defined by*

$$X = \frac{(U/d_1)}{(V/d_2)}$$

follows an F–distribution with d_1 numerator degrees of freedom and d_2 denominator degrees of freedom. We write X˜F (d_1, d_2). The probability density function (pdf) for X is given by

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}}}{x B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$
$$= \frac{1}{B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2} - 1} \left(1 + \frac{d_1}{d_2} x\right)^{-\frac{d_1 + d_2}{2}}$$

for real $x > 0$, where B is the Beta function.

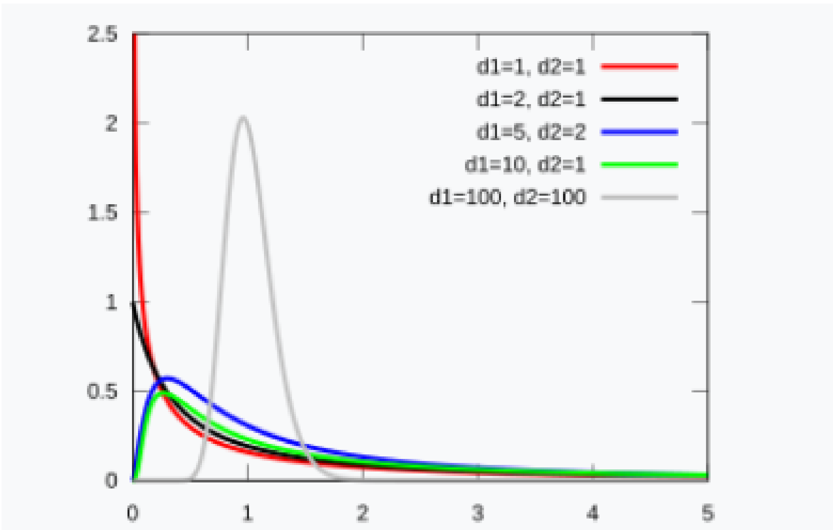


Figure 3.28: The Fisher distribution for diffferents values of d_1 and d_2.

Definition 3.36. *The cumulative distribution function is*

$$F(x; d_1, d_2) = I_{d_1 x/(d_1 x + d_2)}\left(\frac{d_1}{2}, \frac{d_2}{2}\right),$$

where I is the regularized incomplete beta function.

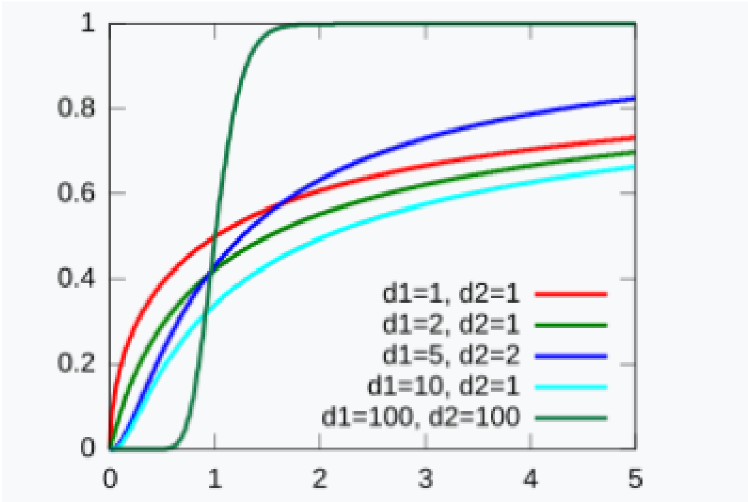


Figure 3.29: The cumulative Fisher distribution for diffferents values of d_1 and d_2.

3.2.10.1 Properties of the F−Fisher distribution

1. **Shape:** The F−distribution is right-skewed, meaning it has a longer tail on the right side. As the degrees of freedom increase, the distribution becomes more symmetric.

2. **Range:** The F−distribution takes on values from 0 to ∞ (it cannot be negative).

3. **Mean:** The mean of the F−distribution is given by

$$Mean = \frac{d_2}{d_2 - 2} \quad \text{for} \quad d_2 > 2.$$

4. **Variance:** The variance is given by

$$Variance = \frac{2.d_2^2(d_1 + d_2 - 2)}{d_1.(d_2 - 2)^2(d_2 - 4)} \quad \text{for} \quad d_2 > 4.$$

5. **Reciprocal property:** If $X \sim F(d_1, d_2)$, then $\frac{1}{X} \sim F(d_2, d_1)$.

6. **Additive property:** The distribution is used in hypothesis testing to determine if the variances of two populations are significantly different.

Remark 3.33. *The F−distribution table typically provides critical values for various combinations of degrees of freedom d_1 and d_2 at specific significance levels (commonly $\alpha = 0.10, \alpha = 0.05$ and $\alpha = 0.01$).*

Example 3.33. $F_{0.95}(1,10) \simeq 4.96$. $F_{0.95}(4,5) \simeq 5.19$ *(see Appendix B for more details to use the $F-$table to find the probability associated with a particular $F-$value.).*

Example 3.34. *Suppose you want to compare the variance of the test scores from two different teaching methods. You collect the scores from two independent groups of students:*

1. *Group 1 (Teaching method A):*

- *Number of students* $n_1 = 10$
- *Variance of scores* $s_1^2 = 20$.

2. *Group 2 (Teaching method B):*

- *Number of students* $n_2 = 12$
- *Variance of scores* $s_2^2 = 15$.

You can use the $F-$test to determine if the variances are significantly different.
Hypothesis: You want to test the null hypothesis (H_0) *that the variance of two groups are equal:*

$$H_0 : \sigma_1^2 = \sigma_2^2$$

against the alternative hypothesis (H_α)*:*

$$H_\alpha : \sigma_1^2 \neq \sigma_2^2$$

The test statistics is calculated as:

$$F = \frac{s_1^2}{s_2^2} = \frac{20}{15} = \frac{4}{3} \approx 1.33$$

The degree of freedom for the group 1 (numerator) and group 2 (denominator) are:

- $d_1 = n_1 - 1 = 10 - 1 = 9$
- $d_2 = n_2 - 1 = 12 - 1 = 11$

Critical value: you need to find the critical value for the $F-$distribution with $d_1 = 9$ and $d_2 = 11$ at your chosen significance level (α) *of* 0.05. *You can look this in an $F-$distribution table.*
Assuming you find a critical values of approximately 3.09 *for a two-tailed test.*
Decision rule:

1. *If the calculated $F-$statistic is greater than the critical value, you reject the null hypothesis.*

2. *If it is less than or equal to the critical value, you fail to reject the null hypothesis.*

In this case:

- *Calculated $F-$statistic:* 1.33

- *Critical value:* 3.09

Since $1.33 < 3.09$ *you fail to reject the null hypothesis.*

Conclusion: *Based on this analysis, there is not enough evidence to suggest that the variances of the test scores from the two teaching methods are significantly different. This means the variability in scores between the two groups mat be similar.*

3.2.11 Weibull distribution

The Weibull distribution is a kind of versatile continuous probability distribution often used in reliability engineering, survival analysis, and modeling life data. It can also fit a huge range of data from many other fields like economics, hydrology, biology, engineering sciences. It is an extreme value of probability distribution which is frequently used to model the reliability, survival, wind speeds and other data.
It is named after Swedish mathematician Waloddi Weibull, who described it in detail in 1939, although it was first identified by René Maurice Fréchet and first applied by Rosin & Rammler (1933) to describe a particle size distribution.

Definition 3.37. *The formula general Weibull Distribution for three-parameter pdf is given as*

$$f(x;\lambda,k,\mu)=\begin{cases}\frac{k}{\lambda}\left(\frac{x-\mu}{\lambda}\right)^{k-1}\exp\left(-\left(\frac{x-\mu}{\lambda}\right)^{k}\right) & \textit{for} \quad x\geq\mu\\ 0 & \textit{for} \quad x<0\end{cases}$$

where

- $\lambda > 0$ *is the scale parameter (also called the characteristic life parameter).*

- $k > 0$ *is the shape parameter (also called as the Weibull slope or the threshold parameter).*

- *μ is the location parameter (also called the waiting time parameter or sometimes the shift parameter).*

Definition 3.38. *The (two-parameter) Weibull distribution is characterized by its shape and scale parameters.*

Definition 3.39. *A random variable X has a Weibull distribution with parameters $k, \lambda \geq 0$, write $X \sim Weibull(k, \lambda)$, if its probability density function is given by*

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^{k}\right) & \textit{for} \quad x \geq 0 \\ 0 & \textit{otherwise} \end{cases}$$

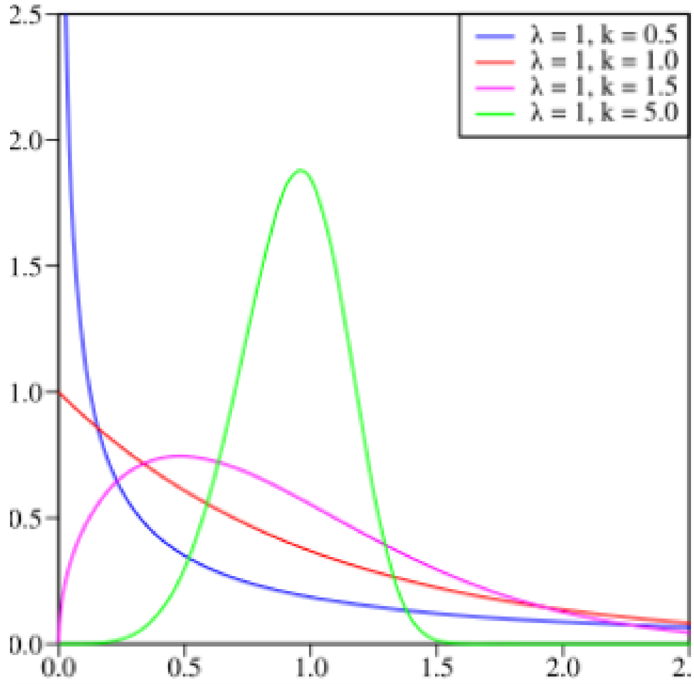


Figure 3.30: The Weibull probability density for diffferents values of λ and k.

Definition 3.40. *The cumulative distribution function of the Weibull,distribution is given by*

$$F(x; \lambda, k) = \begin{cases} 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^{k}\right) & \textit{for} \quad x \geq 0 \\ 0 & \textit{for} \quad x < 0 \end{cases}$$

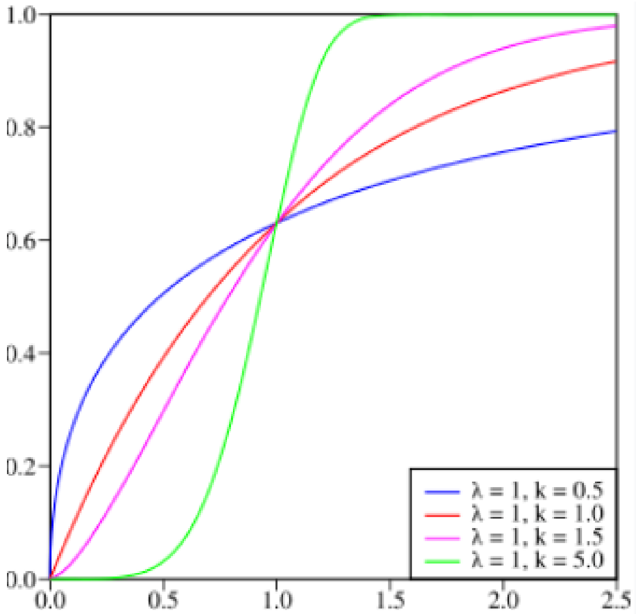


Figure 3.31: The cumulative Weibull distribution for diffferents values of λ and k.

3.2.11.1 Properties of Weibull distribution

1. **Mean:** The mean of a Weibull random variable is:

$$E(X) = \lambda\Gamma\left(1 + \frac{1}{k}\right)$$

   where Γ is the gama function.

2. **Variance:** The variance is given by:

$$Var(X) = \lambda^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right].$$

3. **Shape parameter** k:

   - If $k < 1$: The distribution models decreasing failure rates.
   - If $k = 1$: The distribution simplifies to an exponential distribution, indicating a constant failure rate.
   - If $k > 1$: The distribution models increasing failure rates.

4. **Scale parameter** λ: The scale parameter stretches or compresses the distribution along the x–axis.

Example 3.35. *Suppose we have a product with a scale parameter* $\lambda = 1000$ *hours and a shape parameter* $k = 1.5$*. We can use the Weibull distribution to model the time until failure of the product, allowing for increasing failure rates as it ages.*

3.3 Approximations

3.3.1 Approximation of a binomial distribution by a Poisson distribution

In 1898 Bortkiewicz showed that for large n and small p in a binomial distribution $B(n,p)$, where np is constant, the distribution can be approximated by a Poisson distribution with parameter $\lambda = np$.

Proposition 3.11. *A Poisson distribution with parameter* $\lambda > 0$ *can be derived as a limiting case of binomial distribution* $B(n,p)$*, when* n *is large,* p *is small and* $np = \lambda$*, and note that* $B(n,p) \approx P(np)$

Proof. We have $X \sim B(n,p)$, then for $0 \le k \le n$

$$P(X=k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \frac{(np)^k}{n^k} (1-p)^{n-k}.$$

If we set $np = \lambda$ for some $\lambda > 0$, then $p = \frac{\lambda}{n}, q = 1 - \frac{\lambda}{n}$. then this becomes

$$P(X=k) = \frac{n!}{(n-k)!n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

While keeping k fixed at some nonnegative integer, in that case

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-k+1}{n} \\ &= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right) \end{aligned}$$

$\lim \frac{n!}{(n-k)!n^k} = 1$. When n is large, and if k is small relative to n, we will write:

$$\left(1-\frac{\lambda}{n}\right)^{n-k} \approx \left(1-\frac{\lambda}{n}\right)^{n}.$$

and the exponential factor comes from:

$$\log\left(1-\frac{\lambda}{n}\right)^n = n \log\left(1-\frac{\lambda}{n}\right) = n\left(-\frac{\lambda}{n} - \frac{1}{2}\frac{\lambda^2}{n^2} - \dots\right) \approx -\lambda \text{ if } \frac{\lambda}{n} \approx 0.$$

then

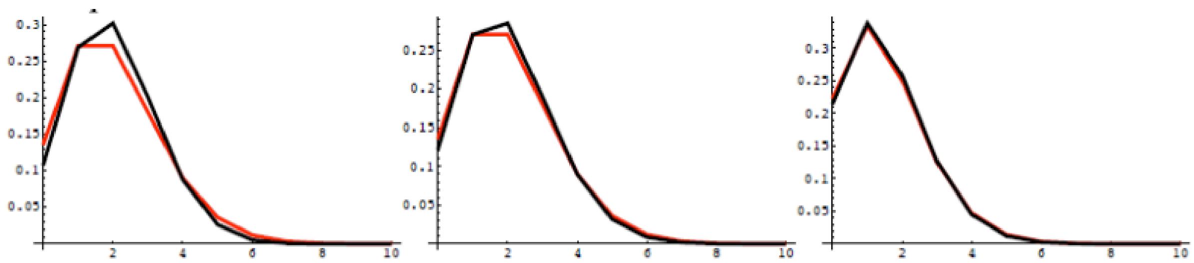
$$\left(1-\frac{\lambda}{n}\right)^{n-k} \longrightarrow e^{-\lambda}.1 = e^{-\lambda}.$$

Taking limit $n \to \infty$, we see that

$$\lim_{n} P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

□

Remark 3.34. *If X follows a Binomial $B(n,p)$, with n is large ($n > 20$) and p is small ($p \leq 0.1$) and $np \leq 5$. Then for all $k \geq 0$: $P(X=k) \simeq P(Y=k)$, where Y follows a Poisson distribution with parameter $\lambda = np$.*



$B(10, 0.2)$ $P(2)$ — $B(20, 0.1)$ $P(2)$ — $B(30, 0.05)$ $P(1.5)$

Figure 3.32: Comparison of Poisson (black) and Binomial (red) distributions: bad approximation ($n = 10$).

Example 3.36. *Let X be a random variable following the binomial distribution $B(100, 0.1)$ is approximate by the Poisson distribution $P(10)$.*

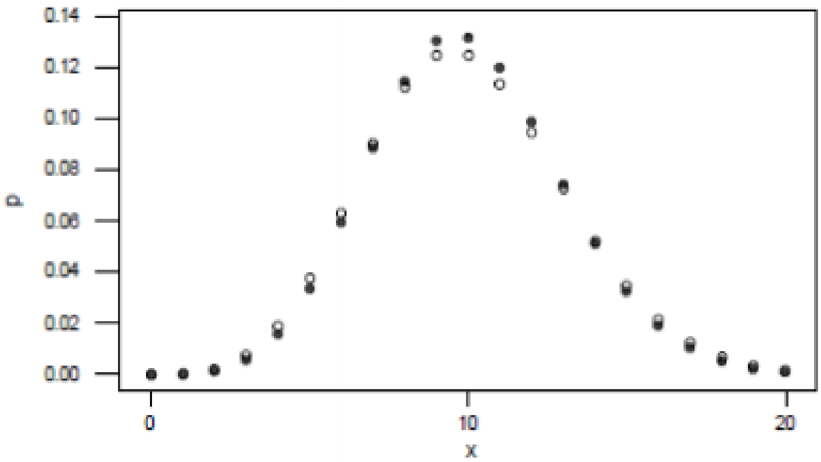


Figure 3.33: Plot of the Binomial $(100, 1/10)$ ($\bullet$) and the Poisson (10) ($\circ$) probability at the values: $0, 1, \ldots, 20$.

Example 3.37. *Find the binomial probability $P(X = 3)$ by using the Poisson distribution if $p = 0.01$ and $n = 200$.*

Solution: *Using Poisson distribution,* $\lambda = np = 0.01 \times 200 = 2$

$$P(X = 3) = \frac{2^3 e^{-2}}{3!} = 0.1804$$

Using Binomial distribution, $n = 200, p = 0.01$

$$P(X = 3) = C_{200}^{3} (0.01)^3 (0.99)^{99} = 0.1814.$$

3.3.2 Approximation of a hypergeometric distribution by a binomial distribution

The hypergeometric distribution approaches the binomial distribution under the following conditions:

- Large population size: the total population N should be much larger than the sample size n (i.e., $N \gg n$).
- Small sample size: the sample size n should be relatively small compared to the population size N.

- Constant probability: as the population size becomes very large, the probability of success on each draw remains approximately constant.

Proposition 3.12. *The hypergeometric distribution approximates the binomial distribution when N tends to infinity.*

Proof. Let $k \in \{0,1,\dots,\min(N,n)\}$, We have

$$\begin{aligned}
P(X=k) &= \frac{C_K^k C_{N-K}^{n-k}}{C_N^n} \\
&= \frac{K!}{k!(K-k)!}\frac{(N-K)!}{(n-k)!(N-K-n+k)!}\frac{n!(N-n)!}{N!} \\
&= \frac{n!}{k!(n-k)!}\frac{K!(N-K)!}{N!(K-k)!}\frac{(N-n)!}{(N-K-n+k)!} \\
&= C_n^k\left(\frac{K}{N}\frac{K-1}{N-1}\cdots\frac{K-k+1}{N-k+1}\right)\left(\frac{N-K}{N-k}\frac{N-K-1}{N-k-1}\right. \\
&\quad \left.\cdots\frac{N-K-n+k+1}{N-n-1}\right)
\end{aligned}$$

if we suppose that $N \to \infty$ and $\frac{K}{N} \to p$, we have

$$\begin{aligned}
\frac{K}{N} \to p, \frac{K-1}{N-1} &= \frac{\frac{K}{N}-\frac{1}{N}}{1-\frac{1}{N}} \to p, \dots \\
\frac{K-k+1}{N-k+1} &= \frac{\frac{K}{N}-\frac{k-1}{N}}{1-\frac{k-1}{N}} \to p; \\
\frac{N-K}{N-k} &= \frac{1-\frac{K}{N}}{1-\frac{k}{N}} \to 1-p, \frac{N-K-1}{N-k-1} \to 1-p, \dots \\
\frac{N-K-n+k+1}{N-n-1} &\to 1-p
\end{aligned}$$

in conclusion if $N \to \infty$ and $\frac{K}{N} \to p$, we have

$$\frac{C_K^k C_{N-K}^{n-k}}{C_N^n} \longrightarrow C_k^n\left(\frac{K}{N}\right)^k\left(1-\frac{K}{N}\right)^{n-k}.$$

□

Remark 3.35. *In practice, the condition for approximating the hypergeometric distribution with the binomial distribution is that N should be at least 10 times larger than n ($N \geq 10n$).*

Example 3.38. *Let X be a random variable following the hypergeometric distribution $\mathcal{H}(100,5,4)$. We will calculate $P(X \geq 1)$.*
Solution: *We have $N = 100 \geq 10n = 10.5 = 50$.*

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{C_0^5 C_4^{95}}{C_4^{100}} \simeq 0.188$$

Approximate calculation: we approximate the distribution $\mathcal{H}(100,5,4)$ the binomial distribution $B(4,0.05)$, then

$$P(X \geq 1) = 1 - P(X = 0) = 1 - C_0^4 (0.05)^0 (0.95)^4 \simeq 0.185.$$

3.3.3 Approximation of a binomial distribution by a normal distribution

The approximation between the binomial distribution and the normal distribution is useful when dealing with large sample sizes.

Theorem 3.7. *If S_n is a binomial variable with parameters n and p, $B(n,p)$, then*

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b \underset{n\to\infty}{\longrightarrow} P(a \leq Z \leq b)\right),$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0,1)$.

In practice, when n large ($n \geq 30$), if $np \geq 5$ and $n(1-p) \geq 5$, then we see that the binomial distribution can be approximated by a normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

Remark 3.36. *When $b - a$ is small, there is a correction that makes things more accurate, namely replace a by $a - \frac{1}{2}$and b by $b + \frac{1}{2}$. This correction never hurts and is sometime necessary. For example, in tossing a coin 100 times, there is positive probability that there are exactly 50 heads, while without the correction, the answer given by the normal approximation would be 0.*

Example 3.39. *Suppose a fair coin is tossed 100 time.*

1. *What is the probability there will be more than 60 heads?*
2. *What is the probability of getting 49, 50, or 51 heads?*

Solution: *we have*

$$\begin{aligned} p &= 0.5, \\ \mu &= np = 100 \times 0.5 = 50, \\ \sigma^2 &= np(1-p) = 100 \times 0.5 \times 0.5 = 25, \\ \sigma &= \sqrt{np(1-p)} = 5. \end{aligned}$$

then from the above theorem let $Z = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{S_n - 50}{5} \sim \mathcal{N}(0,1)$

1. *We have*

$$\begin{aligned} P(S_n \geq 60) &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \geq \frac{60-50}{5}\right) \\ &= P\left(\frac{S_n - 50}{5} \geq 2\right) \\ &\approx P(Z \geq 2) \\ &\approx 0.0228. \end{aligned}$$

2. *We have*

$$\begin{aligned} P(49 \leq S_n \leq 51) &= P(48.5 \leq S_n \leq 51.5) \\ &= P\left(\frac{48.5-50}{5} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{51.5-50}{5}\right) \\ &= P(-0.3 \leq Z \leq 0.3) \\ &= \Phi(0.3) - \Phi(-0.3) \\ &= 2\Phi(0.3) - 1 \\ &\simeq 0.23582. \end{aligned}$$

The normal approximation can be done in three different ways where $S_n = 5Z + 50$*:*

(a)

$$\begin{aligned} P(49 \leq S_n \leq 51) &\approx P(49 \leq 5Z + 50 \leq 51) \\ &= \Phi(0.2) - \Phi(-0.2) \\ &= 2\Phi(0.2) - 1 \\ &\simeq 0.15852. \end{aligned}$$

(b) Or

$$\begin{aligned} P(48 < S_n < 52) &\approx P(48 < 5Z + 50 < 52) \\ &= \Phi(0.4) - \Phi(-0.4) \\ &= 2\Phi(0.4) - 1 \\ &\simeq 0.31084. \end{aligned}$$

(c) Or

$$\begin{aligned} P(48.5 \le S_n \le 51.5) &\approx P(48.5 \le 5Z + 50 \le 51.5) \\ &= \Phi(0.3) - \Phi(-0.3) \\ &= 2\Phi(0.3) - 1 \\ &\simeq 0.23582. \end{aligned}$$

Here all three answers are approximate, and the third one, 0.23582, is the most accurate among these three. We also can compute the precise answer using the binomial formula

$$\begin{aligned} P(48.5 \le S_n \le 51.5) &= \sum_{k=49}^{51} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k} \\ &= \frac{37339688790147532337148742850}{158456325028528675187087900672} \\ &\approx 0.2356465655973331958... \end{aligned}$$

Remark 3.37. *If a continuous distribution such as the normal distribution is used to approximate a discrete one such as the binomial distribution, a continuity correction should be used.*

Example 3.40. *A particular production process used to manufacture ferrite magnets used to operate reed switches in electronic meters is known to give 10% defective magnets on average. If 200 magnets are randomly selected, what is the probability that the number of defective magnets is between 24 and 30?*

Solution: *If X is the number of defective magnets then $X \sim B(200, 0.1)$ and we require*

$$P(24 < X < 30) = P(25 \le X \le 29)$$

Now, $\mu = np = 200 \times 0.1 = 20$ *and* $\sigma = \sqrt{np(1-p)} = \sqrt{200 \times 0.1 \times 0.9} = 4.24$

Note that $np > 5$ *and* $n(1-p) > 5$ *so that approximating* $X \sim B(200, 0.1)$ *by* $Y \sim \mathcal{N}\left(20, 4.24^2\right)$ *is acceptable. We can approximate* $X \sim B(200, 0.1)$ *by the normal distribution* $Y \sim \mathcal{N}\left(20, 4.24^2\right)$ *and use the transformation*

$$Y = \frac{X - 20}{4.24} \sim \mathcal{N}(0, 1)$$

so that with correction we have

$$\begin{aligned}
P(24 < X < 30) &= P(25 \le X \le 29) \\
&\approx P(24.5 \le X \le 29.5) \\
&= P\left(\frac{24.5-20}{4.24} \le \frac{Y-20}{4.24} \le \frac{29.5-20}{4.24}\right) \\
&= P(1.06 \le Z \le 2.24) \\
&= 0.4875 - 0.3554 \\
&= 0.1321.
\end{aligned}$$

Below is a table on how to use the continuity correction for normal approximation to a binomial.

Binomial	Normal
If $P(X = n)$	use $P(n - 0.5 < X < n + 0.5)$
If $P(X > n)$	use $P(X > n + 0.5)$
If $P(X \le n)$	use $P(X < n + 0.5)$
If $P(X < n)$	use $P(< X < n - 0.5)$
If $P(X \ge n)$	use $P(X > n - 0.5)$

Table 3.2: The continuity correction between normal and binomial distributions.

Example 3.41. *From the above example (toss a coin* 100 *times). Since the binomial distribution is discrete and the normal distribution is continuous, apply a continuity correction when using the normal approximation, we can obtain the following normal approximation*

$$\begin{aligned}
P(S_n = 49) &\approx P(48.5 \le 5Z + 50 \le 49.5) \\
&= \Phi(-0.1) - \Phi(-0.3) \\
&= \Phi(0.3) - \Phi(0.1) \\
&\simeq 0.07808.
\end{aligned}$$

$$\begin{aligned}
P(S_n = 50) &\approx P(49.5 \le 5Z + 50 \le 50.5) \\
&= \Phi(0.1) - \Phi(-0.1) \\
&= 2\Phi(0.1) - 1 \\
&\simeq 0.07966.
\end{aligned}$$

$$\begin{aligned} P(S_n = 51) &\approx P(50.5 \leq 5Z + 50 \leq 51.5) \\ &= \Phi(0.3) - \Phi(0.1) \\ &= \Phi(0.3) - \Phi(0.1) \\ &\simeq 0.07808. \end{aligned}$$

Finally, notice that

$$0.07808 + 0.07966 + 0.07808 = 0.23582$$

which is the approximate value for

$$P(48.5 \leq S_n \leq 51.5) \approx P(48.5 \leq 5Z + 50 \leq 51.5).$$

3.3.4 Approximation of a Poisson distribution by a normal distribution

The Poisson distribution ($X \sim P(\lambda)$) can also be approximated by a normal distribution ($X \sim \mathcal{N}\left(\lambda, \sqrt{\lambda}\right)$ under certain conditions, particularly when the mean of the Poisson distribution is sufficiently large, typically $\lambda \geq 20$.

Remark 3.38. *Since Poisson's distribution is also a discrete distribution, we may have to apply the continuity correction.*

Example 3.42. *In a factory there are* 45 *accidents per year and the number of accidents per year follows a Poisson distribution. Use the normal approximation to find the probability that there are more than* 50 *accidents in a year.*
Solution: *Because $\lambda > 20$ a normal approximation can be used.*
Let X be a random variable of the number of accidents per year. To find $P(X > 50)$ apply a coutinuity correction and find $P(X \geq 50.5)$ for the normal approximation $\mu = \lambda = 45$ and $\sigma = \sqrt{\lambda} = \sqrt{45} \simeq 6.71$

$$\begin{aligned} P(X \geq 50.5) &= P\left(\frac{X - 45}{\sqrt{45}} \geq \frac{50.5 - 45}{\sqrt{45}}\right) \\ &\simeq P\left(Z \geq \frac{50.5 - 45}{6.71}\right) \\ &\simeq P(Z \geq 0.820) \\ &\simeq 0.5 - 0.2939 \\ &\simeq 0.2061. \end{aligned}$$

3.4 Transformations of random variables

Transformations of random variables are fundamental concepts in probability theory and statistics, allowing us to derive the distribution of a new random variable based on an existing one. In this section, we study how the distribution of a random variable changes when the variable is transformed in a deterministic way.
To find the distribution of a transformation of a random variable, we can use different techniques depending on whether the original random variable is discrete or continuous.

3.4.1 Transformed variables with discrete distributions

When the transformed variable Y has a discrete distribution, the probability density function of Y can be computed using the following techniques

1. **Identify the transformation:** Define the transformation $Y = g(X)$ where X is a discrete random variable.
2. **Determine the range of Y:** List the possible values that Y can take based on the values of X.
3. **Calculate the probability mass function:** For each possible value of Y, calculate the corresponding probability using the pmf of X :

$$P(Y = y) = P\left(X = g^{-1}(y)\right).$$

If multiple values of X can lead to the same Y, sum their probabilities:

$$P(Y = y) = \sum_{xg(x)=y} P(X = x).$$

Example 3.43. *If X can take values $1, 2$ and 3, and $Y = 2X + 1$.*
We have the values of Y are: $3, 5$ and 7. Then

$$P(Y = 3) = P(X = 1)$$
$$P(Y = 5) = P(X = 2)$$
$$P(Y = 7) = P(X = 3)$$

3.4.2 Transformed variables with continuous distributions

1. **Identify the transformation:** Define the transformation $Y = g(X)$ where X is a continuous random variable.

2. **Determine the range of Y:** Identify the range of Y based on the transformation.

3. **Use the cumulative distribution function:** Calculate the cumulative density function of Y

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P\left(X \leq g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right) \\ &= \int_{-\infty}^{g^{-1}(y)} f(x)dx. \end{aligned}$$

4. **Find the probability density function:** Note that as g is monotonically increasing function, $g(x) \leq y$, then $x \leq g^{-1}(y)$. We can use the change of variables techniques: Let $x = g^{-1}(t)$, so $g'(x)dx = dt$.

$$F_Y(y) = \int_{-\infty}^{y} f_X(g^{-1}(t))\frac{dt}{g'(g^{-1}(t))}.$$

Differentiate the cumulative density function to find the probability density function: we get

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X\left(g^{-1}(y)\right) \\ &= f_X\left(g^{-1}(y)\right)\frac{1}{g'(g^{-1}(y))}. \end{aligned}$$

For g decreasing on the range of X,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P\left(X \geq g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right) \end{aligned}$$

and the density

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = -\frac{d}{dy}F_X\left(g^{-1}(y)\right) \\ &= -f_X\left(g^{-1}(y)\right)\frac{1}{g'(g^{-1}(y))}. \end{aligned}$$

For g decreasing, we also have g^{-1} decreasing and consequently the density of Y is indeed positive, we can combine these two cases to obtain

$$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'(g^{-1}(y))\right|}.$$

Example 3.44. *If X is uniform on $[0,1]$ and $Y = X^2$. We have:*

- *Range of Y : $[0,1]$.*

- *The cumulative density:*

$$F_Y(y) = P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(X \leq \sqrt{y}\right) = \sqrt{y}.$$

- *The probability function*

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}}.$$

APPENDICES

Appendix A

1. If $f : \mathbb{R} \to \mathbb{R}$ and $\lim_{y \downarrow x} f(y) = f(x)$, we say that the function f is right-continuous at x. If this holds for every $x \in \mathbb{R}$, we say that f is right-continuous.

2. If $f : \mathbb{R} \to \mathbb{R}$ and $\lim_{y \uparrow x} f(y) = f(x)$, we say that the function f is left-continuous at x. If this holds for every $x \in \mathbb{R}$, we say that f is left-continuous.

3. For a function $f : \mathbb{R} \to \mathbb{R}$, and some $x \in \mathbb{R}$, it is not hard to show, starting from the above definitions, that f is continuous at x if and only if it is both left-continuous and right-continuous.

Appendix B

Let X be an F random variable with 4 numerator degrees of freedom and 5 denominator degrees of freedom. What is the upper fifth percentile?
Answer:
The upper fifth percentile is the $F-$value x such that the probability to the right of x is 0.05, and therefore the probability to the left of x is 0.95. To find x using the $F-$table, we:

1. Find the column headed by $r_1 = 4$.

2. Find the three rows that correspond to $r_2 = 4$.

3. Find the one row, from the group of three rows identified in the above step, that is headed by $\alpha = 0.05$ (and $P(X \leq x) = 0.95$).

Now, all we need to do is read the $F-$value where the column and the identified row intersect. What do you get?

$$P(F \le f) = \int_0^f \frac{\Gamma[(r_1+r_2)/2](r_1/r_2)^{r_1/2} w^{r_1/2-1}}{\Gamma[r_1/2]\Gamma[r_2/2][1+(r_1 w/r_2)]^{(r_1+r_2)/2}}\, dw$$

α	$P(F \le f)$	Den. d.f r_2	Numerator Degrees of Freedom r_1: 1	2	3	4	5	6	7	8
0.05	0.95	1	161.40	199.50	215.70	224.6	230.20	234.00	236.80	238.90
0.0025	0.975		647.74	799.50	864.16	899.58	921.85	937.11	948.22	956.66
0.01	0.99		4052.00	4999.50	5403.00	5625.00	5764.00	5859.00	5928.00	5981.00
0.05	0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
0.0025	0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
0.01	0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
0.05	0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
0.0025	0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
0.01	0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
0.05	0.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
0.0025	0.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
0.01	0.99		21.20	8.00	16.69	15.98	15.52	15.21	14.98	14.80
0.05	0.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
0.0025	0.975		10.01	8.43	7.76	7.37	7.15	6.98	6.85	6.76
0.01	0.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29

The table tells us that the upper fifth percentile of an F random variable with 4 numerator degrees of freedom and 5 denominator degrees of freedom is 5.19.

Appendix C: Statistical tables

Normal Distribution

Table 3.3: Table of cumulative normal distribution for negative z-values:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table 3.4: Table of cumulative normal distribution for positive z-values:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Poisson Distribution

Table 3.5: Table of Poisson distribution for a given value of λ, entry indicates the probability of a specified value of X.

λ

X	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	0.3679
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	0.1839
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	0.0613
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	0.0153
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	0.0031
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	0.0005
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

λ

X	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2.0**
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.3662	0.3614	0.3543	0.3452	0.3347	0.3230	0.3106	0.2975	0.2842	0.2707
2	0.2014	0.2169	0.2303	0.2417	0.2510	0.2584	0.2640	0.2678	0.2700	0.2707
3	0.0738	0.0867	0.0998	0.1128	0.1255	0.1378	0.1496	0.1607	0.1710	0.1804
4	0.0203	0.0260	0.0324	0.0395	0.0471	0.0551	0.0636	0.0723	0.0812	0.0902
5	0.0045	0.0062	0.0084	0.0111	0.0141	0.0176	0.0216	0.0260	0.0309	0.0361
6	0.0008	0.0012	0.0018	0.0026	0.0035	0.0047	0.0061	0.0078	0.0098	0.0120
7	0.0001	0.0002	0.0003	0.0005	0.0008	0.0011	0.0015	0.0020	0.0027	0.0034
8	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0003	0.0005	0.0006	0.0009
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002

λ

X	**2.1**	**2.2**	**2.3**	**2.4**	**2.5**	**2.6**	**2.7**	**2.8**	**2.9**	**3.0**
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.2572	0.2438	0.2306	0.2177	0.2052	0.1931	0.1815	0.1703	0.1596	0.1494
2	0.2700	0.2681	0.2652	0.2613	0.2565	0.2510	0.2450	0.2384	0.2314	0.2240
3	0.1890	0.1966	0.2033	0.2090	0.2138	0.2176	0.2205	0.2225	0.2237	0.2240
4	0.0992	0.1082	0.1169	0.1254	0.1336	0.1414	0.1488	0.1557	0.1622	0.1680
5	0.0417	0.0476	0.0538	0.0602	0.0668	0.0735	0.0804	0.0872	0.0940	0.1008
6	0.0146	0.0174	0.0206	0.0241	0.0278	0.0319	0.0362	0.0407	0.0455	0.0504
7	0.0044	0.0055	0.0068	0.0083	0.0099	0.0118	0.0139	0.0163	0.0188	0.0216
8	0.0011	0.0015	0.0019	0.0025	0.0031	0.0038	0.0047	0.0057	0.0068	0.0081
9	0.0003	0.0004	0.0005	0.0007	0.0009	0.0011	0.0014	0.0018	0.0022	0.0027
10	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0004	0.0005	0.0006	0.0008
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0002
12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

λ

X	**3.1**	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**	**3.7**	**3.8**	**3.9**	**4.0**
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1397	0.1340	0.1217	0.1135	0.1057	0.0984	0.0915	0.0850	0.0789	0.0733
2	0.2165	0.2087	0.2008	0.1929	0.1850	0.1771	0.1692	0.1615	0.1539	0.1465
3	0.2237	0.2226	0.2209	0.2186	0.2158	0.2125	0.2087	0.2046	0.2001	0.1954
4	0.1734	0.1781	0.1823	0.1858	0.1888	0.1912	0.1931	0.1944	0.1951	0.1954
5	0.1075	0.1140	0.1203	0.1264	0.1322	0.1377	0.1429	0.1477	0.1522	0.1563
6	0.0555	0.0608	0.0662	0.0716	0.0771	0.0826	0.0881	0.0936	0.0989	0.1042
7	0.0246	0.0278	0.0312	0.0348	0.0385	0.0425	0.0466	0.0508	0.0551	0.0595
8	0.0095	0.0111	0.0129	0.0148	0.0169	0.0191	0.0215	0.0241	0.0269	0.0298
9	0.0033	0.0040	0.0047	0.0056	0.0066	0.0076	0.0089	0.0102	0.0116	0.0132
10	0.0010	0.0013	0.0016	0.0019	0.0023	0.0028	0.0033	0.0039	0.0045	0.0053
11	0.0003	0.0004	0.0005	0.0006	0.0007	0.0009	0.0011	0.0013	0.0016	0.0019
12	0.0001	0.0001	0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0005	0.0006
13	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

Binomial Distribution

Table 3.6: Table of Binomial distribution for all values ≤ 0.0005 have been left out:

n	x	p										
		0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
2	0	0.902	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.002
	1	0.095	0.180	0.320	0.420	0.480	0.500	0.480	0.420	0.320	0.180	0.095
	2	0.002	0.010	0.040	0.090	0.160	0.250	0.360	0.490	0.640	0.810	0.902
3	0	0.857	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	
	1	0.135	0.243	0.384	0.441	0.432	0.375	0.288	0.189	0.096	0.027	0.007
	2	0.007	0.027	0.096	0.189	0.288	0.375	0.432	0.441	0.384	0.243	0.135
	3		0.001	0.008	0.027	0.064	0.125	0.216	0.343	0.512	0.729	0.857
4	0	0.815	0.656	0.410	0.240	0.130	0.062	0.026	0.008	0.002		
	1	0.171	0.292	0.410	0.412	0.346	0.250	0.154	0.076	0.026	0.004	
	2	0.014	0.049	0.154	0.265	0.346	0.375	0.346	0.265	0.154	0.049	0.014
	3		0.004	0.026	0.076	0.154	0.250	0.346	0.412	0.410	0.292	0.171
	4			0.002	0.008	0.026	0.062	0.130	0.240	0.410	0.656	0.815
5	0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002			
	1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006		
	2	0.021	0.073	0.205	0.309	0.346	0.312	0.230	0.132	0.051	0.008	0.001
	3	0.001	0.008	0.051	0.132	0.230	0.312	0.346	0.309	0.205	0.073	0.021
	4			0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
	5				0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774
6	0	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001			
	1	0.232	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002		
	2	0.031	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	0.001	
	3	0.002	0.015	0.082	0.185	0.276	0.312	0.276	0.185	0.082	0.015	0.002
	4		0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
	5			0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
	6				0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735
7	0	0.698	0.478	0.210	0.082	0.028	0.008	0.002				
	1	0.257	0.372	0.367	0.247	0.131	0.055	0.017	0.004			
	2	0.041	0.124	0.275	0.318	0.261	0.164	0.077	0.025	0.004		
	3	0.004	0.023	0.115	0.227	0.290	0.273	0.194	0.097	0.029	0.003	
	4		0.003	0.029	0.097	0.194	0.273	0.290	0.227	0.115	0.023	0.004
	5			0.004	0.025	0.077	0.164	0.261	0.318	0.275	0.124	0.041
	6				0.004	0.017	0.055	0.131	0.247	0.367	0.372	0.257
	7					0.002	0.008	0.028	0.082	0.210	0.478	0.698
8	0	0.663	0.430	0.168	0.058	0.017	0.004	0.001				
	1	0.279	0.383	0.336	0.198	0.090	0.031	0.008	0.001			
	2	0.051	0.149	0.294	0.296	0.209	0.109	0.041	0.010	0.001		
	3	0.005	0.033	0.147	0.254	0.279	0.219	0.124	0.047	0.009		
	4		0.005	0.046	0.136	0.232	0.273	0.232	0.136	0.046	0.005	
	5			0.009	0.047	0.124	0.219	0.279	0.254	0.147	0.033	0.005
	6			0.001	0.010	0.041	0.109	0.209	0.296	0.294	0.149	0.051
	7				0.001	0.008	0.031	0.090	0.198	0.336	0.383	0.279
	8					0.001	0.004	0.017	0.058	0.168	0.430	0.663

n	*x*	***p***										
		0.05	***0.1***	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***0.95***
9	**0**	0.630	0.387	0.134	0.040	0.010	0.002					
	1	0.299	0.387	0.302	0.156	0.060	0.018	0.004				
	2	0.063	0.172	0.302	0.267	0.161	0.070	0.021	0.004			
	3	0.008	0.045	0.176	0.267	0.251	0.164	0.074	0.021	0.003		
	4	0.001	0.007	0.066	0.172	0.251	0.246	0.167	0.074	0.017	0.001	
	5		0.001	0.017	0.074	0.167	0.246	0.251	0.172	0.066	0.007	0.001
	6			0.003	0.021	0.074	0.164	0.251	0.267	0.176	0.045	0.008
	7				0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.063
	8					0.004	0.018	0.060	0.156	0.302	0.387	0.299
	9						0.002	0.010	0.040	0.134	0.387	0.630
10	**0**	0.599	0.349	0.107	0.028	0.006	0.001					
	1	0.315	0.387	0.268	0.121	0.040	0.010	0.002				
	2	0.075	0.194	0.302	0.233	0.121	0.044	0.011	0.001			
	3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001		
	4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006		
	5		0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	
	6			0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
	7			0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
	8				0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
	9					0.002	0.010	0.040	0.121	0.268	0.387	0.315
	10						0.001	0.006	0.028	0.107	0.349	0.599
11	**0**	0.569	0.314	0.086	0.020	0.004						
	1	0.329	0.384	0.236	0.093	0.027	0.005	0.001				
	2	0.087	0.213	0.295	0.200	0.089	0.027	0.005	0.001			
	3	0.014	0.071	0.221	0.257	0.177	0.081	0.023	0.004			
	4	0.001	0.016	0.111	0.220	0.236	0.161	0.070	0.017	0.002		
	5		0.002	0.039	0.132	0.221	0.226	0.147	0.057	0.010		
	6			0.010	0.057	0.147	0.226	0.221	0.132	0.039	0.002	
	7			0.002	0.017	0.070	0.161	0.236	0.220	0.111	0.016	0.001
	8				0.004	0.023	0.081	0.177	0.257	0.221	0.071	0.014
	9				0.001	0.005	0.027	0.089	0.200	0.295	0.213	0.087
	10					0.001	0.005	0.027	0.093	0.236	0.384	0.329
	11							0.004	0.020	0.086	0.314	0.569
12	**0**	0.540	0.282	0.069	0.014	0.002						
	1	0.341	0.377	0.206	0.071	0.017	0.003					
	2	0.099	0.230	0.283	0.168	0.064	0.016	0.002				
	3	0.017	0.085	0.236	0.240	0.142	0.054	0.012	0.001			
	4	0.002	0.021	0.133	0.231	0.213	0.121	0.042	0.008	0.001		
	5		0.004	0.053	0.158	0.227	0.193	0.101	0.029	0.003		
	6			0.016	0.079	0.177	0.226	0.177	0.079	0.016		
	7			0.003	0.029	0.101	0.193	0.227	0.158	0.053	0.004	
	8			0.001	0.008	0.042	0.121	0.213	0.231	0.133	0.021	0.002
	9				0.001	0.012	0.054	0.142	0.240	0.236	0.085	0.017
	10					0.002	0.016	0.064	0.168	0.283	0.230	0.099
	11						0.003	0.017	0.071	0.206	0.377	0.341
	12							0.002	0.014	0.069	0.282	0.540

n	*x*	*p*										
		0.05	***0.1***	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***0.95***
13	**0**	0.513	0.254	0.055	0.010	0.001						
	1	0.351	0.367	0.179	0.054	0.011	0.002					
	2	0.111	0.245	0.268	0.139	0.045	0.010	0.001				
	3	0.021	0.100	0.246	0.218	0.111	0.035	0.006	0.001			
	4	0.003	0.028	0.154	0.234	0.184	0.087	0.024	0.003			
	5		0.006	0.069	0.180	0.221	0.157	0.066	0.014	0.001		
	6		0.001	0.023	0.103	0.197	0.209	0.131	0.044	0.006		
	7			0.006	0.044	0.131	0.209	0.197	0.103	0.023	0.001	
	8			0.001	0.014	0.066	0.157	0.221	0.180	0.069	0.006	
	9				0.003	0.024	0.087	0.184	0.234	0.154	0.028	0.003
	10				0.001	0.006	0.035	0.111	0.218	0.246	0.100	0.021
	11					0.001	0.010	0.045	0.139	0.268	0.245	0.111
	12						0.002	0.011	0.054	0.179	0.367	0.351
	13							0.001	0.010	0.055	0.254	0.513
14	**0**	0.488	0.229	0.044	0.007	0.001						
	1	0.359	0.356	0.154	0.041	0.007	0.001					
	2	0.123	0.257	0.250	0.113	0.032	0.006	0.001				
	3	0.026	0.114	0.250	0.194	0.085	0.022	0.003				
	4	0.004	0.035	0.172	0.229	0.155	0.061	0.014	0.001			
	5		0.008	0.086	0.196	0.207	0.122	0.041	0.007			
	6		0.001	0.032	0.126	0.207	0.183	0.092	0.023	0.002		
	7			0.009	0.062	0.157	0.209	0.157	0.062	0.009		
	8			0.002	0.023	0.092	0.183	0.207	0.126	0.032	0.001	
	9				0.007	0.041	0.122	0.207	0.196	0.086	0.008	
	10				0.001	0.014	0.061	0.155	0.229	0.172	0.035	0.004
	11					0.003	0.022	0.085	0.194	0.250	0.114	0.026
	12					0.001	0.006	0.032	0.113	0.250	0.257	0.123
	13						0.001	0.007	0.041	0.154	0.356	0.359
	14							0.001	0.007	0.044	0.229	0.488
15	**0**	0.463	0.206	0.035	0.005							
	1	0.366	0.343	0.132	0.031	0.005						
	2	0.135	0.267	0.231	0.092	0.022	0.003					
	3	0.031	0.129	0.250	0.170	0.063	0.014	0.002				
	4	0.005	0.043	0.188	0.219	0.127	0.042	0.007	0.001			
	5	0.001	0.010	0.103	0.206	0.186	0.092	0.024	0.003			
	6		0.002	0.043	0.147	0.207	0.153	0.061	0.012	0.001		
	7			0.014	0.081	0.177	0.196	0.118	0.035	0.003		
	8			0.003	0.035	0.118	0.196	0.177	0.081	0.014		
	9			0.001	0.012	0.061	0.153	0.207	0.147	0.043	0.002	
	10				0.003	0.024	0.092	0.186	0.206	0.103	0.010	0.001
	11				0.001	0.007	0.042	0.127	0.219	0.188	0.043	0.005
	12					0.002	0.014	0.063	0.170	0.250	0.129	0.031
	13						0.003	0.022	0.092	0.231	0.267	0.135
	14							0.005	0.031	0.132	0.343	0.366
	15								0.005	0.035	0.206	0.463

n	*x*	*p*										
		0.05	***0.1***	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***0.95***
16	**0**	0.440	0.185	0.028	0.003							
	1	0.371	0.329	0.113	0.023	0.003						
	2	0.146	0.275	0.211	0.073	0.015	0.002					
	3	0.036	0.142	0.246	0.146	0.047	0.009	0.001				
	4	0.006	0.051	0.200	0.204	0.101	0.028	0.004				
	5	0.001	0.014	0.120	0.210	0.162	0.067	0.014	0.001			
	6		0.003	0.055	0.165	0.198	0.122	0.039	0.006			
	7			0.020	0.101	0.189	0.175	0.084	0.019	0.001		
	8			0.006	0.049	0.142	0.196	0.142	0.049	0.006		
	9			0.001	0.019	0.084	0.175	0.189	0.101	0.020		
	10				0.006	0.039	0.122	0.198	0.165	0.055	0.003	
	11				0.001	0.014	0.067	0.162	0.210	0.120	0.014	0.001
	12					0.004	0.028	0.101	0.204	0.200	0.051	0.006
	13					0.001	0.009	0.047	0.146	0.246	0.142	0.036
	14						0.002	0.015	0.073	0.211	0.275	0.146
	15							0.003	0.023	0.113	0.329	0.371
	16								0.003	0.028	0.185	0.440
17	**0**	0.418	0.167	0.023	0.002							
	1	0.374	0.315	0.096	0.017	0.002						
	2	0.158	0.280	0.191	0.058	0.010	0.001					
	3	0.041	0.156	0.239	0.125	0.034	0.005					
	4	0.008	0.060	0.209	0.187	0.080	0.018	0.002				
	5	0.001	0.017	0.136	0.208	0.138	0.047	0.008	0.001			
	6		0.004	0.068	0.178	0.184	0.094	0.024	0.003			
	7		0.001	0.027	0.120	0.193	0.148	0.057	0.009			
	8			0.008	0.064	0.161	0.185	0.107	0.028	0.002		
	9			0.002	0.028	0.107	0.185	0.161	0.064	0.008		
	10				0.009	0.057	0.148	0.193	0.120	0.027	0.001	
	11				0.003	0.024	0.094	0.184	0.178	0.068	0.004	
	12				0.001	0.008	0.047	0.138	0.208	0.136	0.017	0.001
	13					0.002	0.018	0.080	0.187	0.209	0.060	0.008
	14						0.005	0.034	0.125	0.239	0.156	0.041
	15						0.001	0.010	0.058	0.191	0.280	0.158
	16							0.002	0.017	0.096	0.315	0.374
	17								0.002	0.023	0.167	0.418

n	x	p										
		0.05	***0.1***	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***0.95***
18	**0**	0.397	0.150	0.018	0.002							
	1	0.376	0.300	0.081	0.013	0.001						
	2	0.168	0.284	0.172	0.046	0.007	0.001					
	3	0.047	0.168	0.230	0.105	0.025	0.003					
	4	0.009	0.070	0.215	0.168	0.061	0.012	0.001				
	5	0.001	0.022	0.151	0.202	0.115	0.033	0.004				
	6		0.005	0.082	0.187	0.166	0.071	0.015	0.001			
	7		0.001	0.035	0.138	0.189	0.121	0.037	0.005			
	8			0.012	0.081	0.173	0.167	0.077	0.015	0.001		
	9			0.003	0.039	0.128	0.185	0.128	0.039	0.003		
	10			0.001	0.015	0.077	0.167	0.173	0.081	0.012		
	11				0.005	0.037	0.121	0.189	0.138	0.035	0.001	
	12				0.001	0.015	0.071	0.166	0.187	0.082	0.005	
	13					0.004	0.033	0.115	0.202	0.151	0.022	0.001
	14					0.001	0.012	0.061	0.168	0.215	0.070	0.009
	15						0.003	0.025	0.105	0.230	0.168	0.047
	16						0.001	0.007	0.046	0.172	0.284	0.168
	17							0.001	0.013	0.081	0.300	0.376
	18								0.002	0.018	0.150	0.397
19	**0**	0.377	0.135	0.014	0.001							
	1	0.377	0.285	0.068	0.009	0.001						
	2	0.179	0.285	0.154	0.036	0.005						
	3	0.053	0.180	0.218	0.087	0.017	0.002					
	4	0.011	0.080	0.218	0.149	0.047	0.007	0.001				
	5	0.002	0.027	0.164	0.192	0.093	0.022	0.002				
	6		0.007	0.095	0.192	0.145	0.052	0.008	0.001			
	7		0.001	0.044	0.153	0.180	0.096	0.024	0.002			
	8			0.017	0.098	0.180	0.144	0.053	0.008			
	9			0.005	0.051	0.146	0.176	0.098	0.022	0.001		
	10			0.001	0.022	0.098	0.176	0.146	0.051	0.005		
	11				0.008	0.053	0.144	0.180	0.098	0.071		
	12				0.002	0.024	0.096	0.180	0.153	0.044	0.001	
	13				0.001	0.008	0.052	0.145	0.192	0.095	0.007	
	14					0.002	0.022	0.093	0.192	0.164	0.027	0.002
	15					0.001	0.007	0.047	0.149	0.218	0.080	0.011
	16						0.002	0.017	0.087	0.218	0.180	0.053
	17							0.005	0.036	0.154	0.285	0.179
	18							0.001	0.009	0.068	0.285	0.377
	19								0.001	0.014	0.135	0.377

n	x	p										
		0.05	***0.1***	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***0.95***
20	**0**	0.358	0.122	0.012	0.001							
	1	0.377	0.270	0.058	0.007							
	2	0.189	0.285	0.137	0.028	0.003						
	3	0.060	0.190	0.205	0.072	0.012	0.001					
	4	0.013	0.090	0.218	0.130	0.035	0.005					
	5	0.002	0.032	0.175	0.179	0.075	0.015	0.001				
	6		0.009	0.109	0.192	0.124	0.037	0.005				
	7		0.002	0.055	0.164	0.166	0.074	0.015	0.001			
	8			0.022	0.114	0.180	0.120	0.035	0.004			
	9			0.007	0.065	0.160	0.160	0.071	0.012			
	10			0.002	0.031	0.117	0.176	0.117	0.031	0.002		
	11				0.012	0.071	0.160	0.160	0.065	0.007		
	12				0.004	0.035	0.120	0.180	0.114	0.022		
	13				0.001	0.015	0.074	0.166	0.164	0.055	0.002	
	14					0.005	0.037	0.124	0.192	0.109	0.009	
	15					0.001	0.015	0.075	0.179	0.175	0.032	0.002
	16						0.005	0.035	0.130	0.218	0.090	0.013
	17						0.001	0.012	0.072	0.205	0.190	0.060
	18							0.003	0.028	0.137	0.285	0.189
	19								0.007	0.058	0.270	0.377
	20								0.001	0.012	0.122	0.358

Chi-square Distribution

Table 3.7: Table of Chi-square distribution for differents values of k :

df	χ^2 .995	χ^2 .990	χ^2 .975	χ^2 .950	χ^2 .900	χ^2 .100	χ^2 .050	χ^2 .025	χ^2 .010	χ^2 .005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Student Distribution

Table 3.8: Table of Student distribution for differents values of t :

cum. prob	**t.50**	**t.75**	**t.80**	**t.85**	**t.90**	**t.95**	**t.975**	**t.99**	**t.995**	**t.999**	**t.9995**
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	ConfidenceLevel										

Fisher Distribution

Table 3.9: Table of Fisher distribution for differents values of α:

	$\alpha = 0.005$																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.63	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.25
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.29
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.24
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

	α = 0.01																		
	d.f.N.																		
d.f.D.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*12*	*15*	*20*	*24*	*30*	*40*	*60*	*120*	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99,48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

	$\alpha = 0.02$																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.69	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

d.f.D.	α = 0.05																		
	d.f.N.																		
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

	$\alpha = 0.10$																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

BIBLIOGRAPHY

[1] Ahsanullah, M., Golam Kibria. B.M. and Shakil, M. 2014. Normal and Student's *t* Distributions and Their Applications. Atlantis Press.

[2] Bertsekas, P., Tsitsiklis, J. N. 2000. Introduction to Probability. Course 6.041-6.431 M.I.T. FALL.

[3] Bogaert, P. Probabilités pour scientifiques et ingé nieurs, Introduction au calcul des probabilités, Edition de Boeck.

[4] Boukhari, F. Probabilités. Université Abou Bekr Belkaid Tlemcen.

[5] Cacoullos, T. 1989. Exercises in Probability. Springer-Verlag.

[6] Casella, G. And R.L. Berger. 2002. Statistical Inference. Pacific Grove, CA: Duxbury.

[7] Chalabi, E. 2021. Cours probabilités et statistique avec exercices corrigés. Université Salah Boubnider Constantine 3.

[8] Degrave, C., Degrave, D. Précis de mathématiques Probabilités-Statistiques 1re et 2eme années, Cours –Mé thodes-Exercices résolus, édition Bréal.

[9] Durrett, R. 2019. Probability : theory and examples. Cambridge University Press. ISBN 978-0-521-76539-8

[10] Evans, M.J and Rosenthal, J.S. 2023. Probability and Statistics: The Science of Uncertainty. University of Toronto.

[11] Faber, M.H. 2007. Statistics and probability theory. Lecture notes.

[12] Calot. G., Cours de calcul des Probabilités. Dunod. 1982.

[13] Gordon, H. 1997. Discrete Probability. Springer-Verlag New York, Inc.

[14] Gut, A. 2005. Probability : a graduate course. Springer texts in statistics.

[15] Jacod, J., Protter, P. 2003. Probability-Essentials. Springer-Verlag Berlin and Heidelberg. Second edition.

[16] Lecoutre, J.P. Statistique et probabilités, Manuel et exercices corrigés. Edition DUNOD.

[17] Lefebvre, M. 2000. Basic Probability theory with applications. Springer.

[18] Li, X. 2015. Fall STAT-3611.Lecture notes. Course en ligne.

[19] Redjdal, K. 2004. Probabilités, OPU Alger.

[20] Tian, Y. 2022. Calculus-based Introduction to Statistics: Lecture notes. Columbia University.

[21] Tropp, JA. 2023. Probability theory and computational mathematics. Caltech CMS. Lecture Notes 2023-01, Pasadena. https://doi.org/10.7907/q75sz-e1e79.

[22] Tsun, A. 2020. Probability and statistics with applications to computing. E-book. https://www.alextsun.com/files/Prob_Stat_for_CS_Book.pdf.