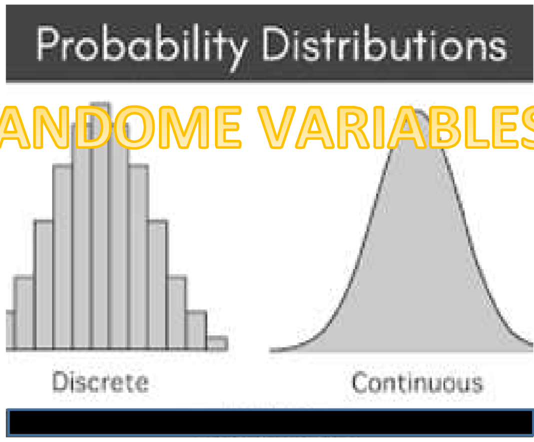
MOHAMED KHIDER UNIVERSITY, BISKRA

FACULTY OF EXACT SCIENCES

DEPARTMENT OF MATHEMATICS

COURSE OF PROBABILITY WITH CORRECTED EXERCISES

PROBABILITY AND RANDOME VARIABLES



Presented by

Dr. LAIADI ABDELKADER

September 2025

Contents

List of Figures

Introduction

These courses is part of the official program of the Biostatistics module intended mainly for students in the second year of a biology degree, but may possibly be useful for students in other mathematics modules (within the framework of the L.M.D. system), and anyone wishing to know probability Fundamentals with discrete and continuous random variables.
The required mathematical level is that of the second year of a Mathematics, economics, engineering degree. The content of this subject is the basis of any introduction to probability theory. It allows the student to acquire the maximum of mathematical techniques necessary for most of the subjects studied throughout their course, namely: probability, discrete random variables, continuous random variables, mathematical programming and...etc.
The theory of probability had its origin in gambling and games of chance. It owes much to the curiosity of gamblers who pestered their friends in the mathematical world with all sorts of questions. Unfortunately, this association with gambling contributed to very slow and sporadic growth of probability theory as a mathematical discipline. The mathematicians of the day took little or no interest in the development of any theory but looked only at the combinatorial reasoning involved in each problem.
The first attempt at some mathematical rigor is credited to Laplace. In his monumental work, Theorie analytique des probabilites (1812), Laplace gave the classical definition of the probability of an event that can occur only in a finite number of ways as the proportion of the number of favorable outcomes to the total number of all possible outcomes, provided that all the outcomes are equally likely. According to this definition, computation of the probability of events was reduced to combinatorial counting problems. Even in those days, this definition was found inadequate. In addition to being circular and restrictive, it did not

answer the question of what probability is; it only gave a practical method of computing the probabilities of some simple events.

An extension of the classical definition of Laplace was used to evaluate the probabilities of sets of events with infinite outcomes. The notion of equal likelihood of certain events played a key role in this development. According to this extension, if Ω is some region with a well-defined measure (length, area, volume, etc.), the probability that a point chosen at random lies in a subregion A of Ω is the ratio measure(A)/measure(Ω). Many problems of geometric probability were solved using this extension. The trouble is that one can define at random in any way one pleases, and different definitions lead to different answers.

The mathematical theory of probability as we know it today is of comparatively recent origin. It was A. N. Kolmogorov who axiomatized probability in his fundamental work, Foundations of the Theory of Probability (Berlin), in 1933. According to this development, random events are represented by sets and probability is just a normed measure defined on these sets. This measure-theoretic development not only provided a logically consistent foundation for probability theory but also joined it to the mainstream of modern mathematics.

This handout contains three main chapters, where are presented

1) Introduction to probability Fundamentals.

2) Some definitions and theorems on discrete random variables.

3) Some definitions and theorems on continuous random variables.

Each chapter ends with some corrected exercises to check the acquisition of the essential concepts that have been introduced.

I could not end this foreword without a great, more personal tribute to my colleague teachers who have seriously examined this handout. I would also like to thank our readers in particular. Finally, errors may be found, please report them to the author.

Chapter 1

Probability Fundamentals

1.1 Basic probability concepts

1.1.1 Introduction

The mathematical theory of probability gives us the basic tools for constructing and analyzing mathematical models for random phenomena. In studying a random phenomenon, we are dealing with an experiment of which the outcome is not predictable in advance. Experiments of this type that immediately come to mind are those arising in games of chance. In fact, the earliest development of probability theory in the fifteenth and sixteenth centuries was motivated by problems of this type.

Our interest in the study of a random phenomenon is in the statements we can make concerning the events that can occur. Events and combinations of events thus play a central role in probability theory. The mathematics of events is closely tied to the theory of sets, and we give in this section some of its basic concepts and algebraic operations.

1.1.2 Random experiment

Definition 1. A random (or statistical) experiment is an experiment in which:

(a) All outcomes of the experiment are known in advance.

(b) Any performance of the experiment results in an outcome that is not known in advance.

(c) The experiment can be repeated under identical conditions.

1.1.3 The sample space

Definition 2. A sample space Ω is the set of possible outcomes of a random experiment.

Example 1.

- When flipping a coin, $\Omega = \{$ heads, tails$\} = \{H, T\}$.
- When throwing a dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$
- When analysing financial data, Ω could be the set of all possible values of all stocks in

all of the world's stockmarkets at all dates and times in the past, present and future.

- When designing a communication system, Ω could be the set of all the files a user might

ever consider transmitting and all the random behaviours (noise) the transmission medium (channel) may possibly exhibit in the past, present or future.

- When studying a physical process, Ω could be every observable quantity in the universe.

A sample space can be a discrete finite set, a discrete countably infinite set such as the set of integers, or a continuous set such as the set of real numbers. Most examples in probability textbooks tend to concentrate on simple random experiments such as flipping a coin or throwing a dice. However, most applications of probability theory typically concern much larger sets Ω. In any case, in order for probability theory to make sense, your sample space must include all random quantities you may ever want to examine jointly.

Sample point

Definition 3. Each possible outcome is called a **sample point**.

1.1.4 Events

Definition 4. **An event** is generally referred to as a subset of the sample space Ω having one or more sample points as its elements.

Example 2. When throwing a dice

- The event $A = \{2.4.6\} \subset \Omega$ that the outcome is even.
- The event $B = \{1, 2, 3\}$ that the outcome is 3 or less.
- The event $C = \{4\}$ that the outcome is 4. This is both an outcome and an event. Some call it an atomic" event because it contains just one outcome.
- The empty event $D = \emptyset$
- The certain event $E = \Omega$

The idea that an event is a subset of possible outcomes is not everyone's cup of tea. Why call it event" if it's just a subset? Most people would not intuitively equate the notion of event with the idea of a set or subset. However, you've all seen Venn diagrams below. This diagrams are simply a pictorial representation of events as sets. An example Venn diagram is drawn in **Figure 1**. The diagrams shows the sample space Ω and two events A and B. The intersection $A \cap B$, the union $A \cup B$, the complement $A^c = \overline{A}$ are further events that can be pictured in the Venn diagram.

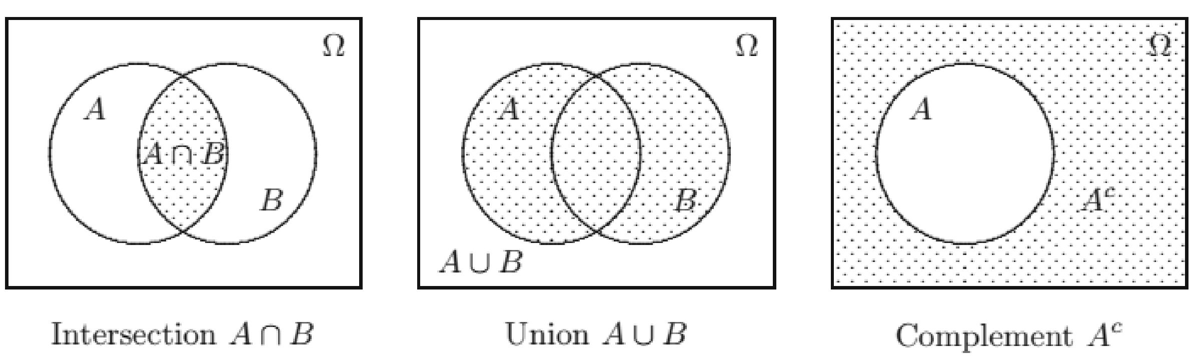


Figure 1.1: Diagrams of intersection, union, and complement

Definition 5. We call events A and B disjoint or mutually exclusive if A and B have no outcomes in common; in set terminology: $A \cap B = \emptyset$

Definition 6. The **complement** of an event $A \subset \Omega$. We denote the complement of A by the symbol A^c or $\overline{A}$.

Definition 7. The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.

Definition 8. The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

Example 3. Let the random experiment is a throw of a dice, $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $A = \{2, 4, 6\}$ and $B = \{1, 2, 3\}$. In set notation, it is obvious that $A \cap B = \{2\}$, $A \cup B = \{1, 2, 3, 4, 6\}$, $\overline{A} = \{1, 3, 5\}$ and $\overline{B} = \{4, 5, 6\}$. In event" terminology, it may be less easy to specify some of these events.

- A (the event that the outcome is even), B (the event that the outcome is 3 or less).
  - $\overline{A}$ (the event that the outcome is NOT even, i.e., odd) and $\overline{B}$ (the event that the outcome is NOT 3 or less, i.e., 4 or more) are all easy to think about.
  - However, $A \cap B$ (the event that the outcome is even AND 3 or less) or $A \cup B$ (the event that the outcome is even OR 3 or less) are more difficult to consider intuitively.

The set analogy and its pictorial representation the Venn diagram are tools that help us refine our intuition about events.

DeMorgan's laws. For any two events A and B we have

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \;\text{ and }\; \overline{A \cap B} = \overline{A} \cup \overline{B}$$

Table 1. Corresponding statements in set theory and probability

Set theory	Probability theory
Space, Ω	Sample space, sure event
empty set, $\emptyset$	Impossible event
Elements, $a, b, ...$	Sample points, $a, b, ...$
Sets, A, B	Events, A, B
A	Event A accurs
$\overline{A}$	Event A does not accur
$A \cup B$	At least one of A and B occurs
$A \cap B = AB$	Both A and B occur
$A \subset B$	A is a subset of B
$A \cap B = \emptyset$	A and B are mutually exclusive

1.1.5 Probability of an event

Probability of an event A in Ω is defined as $P(A)$ and equals to

$$P(A) = \frac{\text{number of possible points of the event}}{\text{number of all possible points in } \Omega} = \frac{n(A)}{n(\Omega)}$$

Axioms of probability

Let Ω be a random sample space and A be an event within Ω. Then

1) $0 \le P(A) \le 1$

2) $P(\Omega) = 1$

3) The sum of the probabilities of all simple events must be 1.e.g $\sum_{i \ge 1} P_i = 1$

4) For any disjoint events $E_i, i = 1, 2, ..., n$ e.g $E_i \cap E_j = \emptyset$ for $i \ne j$ where $\emptyset$ is empty set, we have

$$P(E_1 \cup E_2 \cup ... \cup E_n) = P(E_1) + P(E_2) + ... + P(E_n)$$

Example 4. A spinner has 4 equal sectors colored yellow, blue, green and red. After spinning the spinner, what is the probability of landing on each color?

The possible outcomes of this experiment are yellow, blue, green, and red.

$$P\left(yellow\right) = \frac{\text{number of ways to land on yellow}}{\text{total number of colors}} = \frac{1}{4}$$

$$P\left(blue\right) = \frac{\text{number of ways to land on blue}}{\text{total number of colors}} = \frac{1}{4}$$

Find $P(green)$ and $P(red)$?

Example 5. Let us toss a coin. The sample space is (Ω, S), where $\Omega = \{H, T\}$ and S is the all subsets of Ω. Let us define P on S as follows:

$$P\left(H\right) = \frac{1}{2}, P\left(T\right) = \frac{1}{2}$$

Then P clearly defines a probability.

Example 6. Let $\Omega = \{1, 2, 3, ...\}$ be the set of positive integers, and let S be the class of all subsets of Ω. Define P on S as follows:

$$P\left(k\right) = \frac{1}{2^k}, k = 1, 2, ...$$

Then $\sum\limits_{k=1}^{\infty} P\left(k\right) = 1$, we have P defines a probability.

Example 7. When throwing a dice. We have $\Omega = \{1, 2, 3, 4, 5, 6\}$

- The event $A = \{2.4.6\} \subset \Omega$ that the outcome is even.
- The event $B = \{1, 2, 3\}$ that the outcome is 3 or less.
- The event $C = \{4\}$ that the outcome is 4.

The probability of the events A, B and C are given by

$$P\left(A\right) = \frac{3}{6} = 0.5, P\left(B\right) = \frac{3}{6} = 0.5, P\left(C\right) = \frac{1}{6}$$

Theorem 1. P is monotone and subtractive; that is, if A, B are two events and $A \subset B$, then

$$P(A) < P(B)$$

and

$$P(B - A) = P(B) - P(A)$$

where $B - A = B \cap A^c$, A^c being the complement of the event A.

Proof. If $A \subset B$, then

$$B = (A \cap B) + (B - A) = A + (B - A),$$

and we obtain

$$P(B) = P(A) + P(B - A)$$

Remark if $P(A) = 0$ for any event A, we call A an event with zero probability or a null event. However, it does not follow that $A = \emptyset$.

Similarly, if $P(B) = 1$ for for any event, we call B a certain event, but it does not follow that $B = \Omega$.

Theorem 2 (Addition Rule). If A, B are two events, then

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

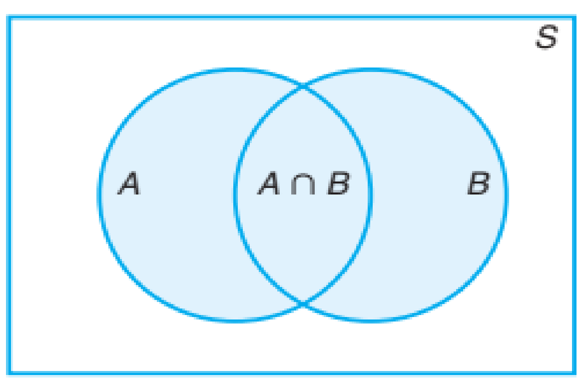


Figure 1.2: Additive rule of probability

Corollary 1. P is subadditive, that is, if A, B are two events, then

$$P(A \cup B) \leq P(A) + P(B)$$

Corollary 1 can be extended to an arbitrary number of events A_j

$$P\left(\bigcup_j A_j\right) \leq \bigcup_j P(A_j)$$

Corollary 2. If $B = A^c$, then A and B are disjoint and

$$P(A^c) = P\left(\overline{A}\right) = 1 - P(A)$$

Example 8. A die is rolled twice. Let all the elementary events in $\Omega = \{(i,j) : i,j = 1,2,...,6\}$ be assigned the same probability. Let C be the event that the first throw shows a number ≤ 2, and D be the event that the second throw shows at least 5. Then

$$C = \{(i,j) : 1 \leq i \leq 2, j = 1,2,...,6\}$$

$$D = \{(i,j) : 5 \leq j \leq 6, i = 1,2,...,6\}$$

and

$$C \cap D = \{(1,5),(1,6),(2,5),(2,6)\}$$

we have

$$P(C \cup D) = P(C) + P(D) - P(C \cap D) = \frac{12}{36} + \frac{12}{36} - \frac{4}{36} = \frac{20}{36} = 0.55$$

and

$$P\left(\overline{C}\right) = 1 - P(C) = 1 - \frac{12}{36} = \frac{24}{36}$$

where $\overline{C} = C^c = \{(i,j) : i > 2, j = 1,2,...,6\}$, $P\left(\overline{C}\right) = \frac{24}{36}$.

Theorem 3 (Boole's Inequality). For any two events A and B

$$P\left(A \cap B\right) \geq 1 - P\left(\overline{A}\right) - P\left(\overline{B}\right)$$

Corollary 3.

1) Let $\{B_j\}$, $j = 1, 2, ...$ be a countable sequence of events; then

$$P\left(\cap B_j\right) \geq 1 - \sum P\left(B_j\right)$$

2) (**Implication Rule**). If A, B and C are three events and $A, B \subset C$, then

$$P\left(\overline{C}\right) \leq P\left(\overline{A}\right) + p\left(\overline{B}\right)$$

Fundamental principal of counting

Factorial Notation. Factorial n is denoted as $n!$ and is defined as $n! = 1 \times 2 \times 3 \times ... \times n$.
Note: If some procedure can be performed in n_1 different ways and a second procedure can be performed in n_2 ways , third procedure can be performed in n_3 ways, and so forth then the number of ways the procedure can be performed in the order indicated is

$$n_1 \times n_2 \times n_3 \times ....$$

For example. The number of different signals each consists of 8 flags in a vertical line formed from a set of 4 indistinguishable red flags and 3 indistinguishable white flags and one blue flag is 8!4!3!.

Combinations

A combination of n objects taken r at a time is denoted by C_r^n or $\binom{n}{r}$ where

$$C_r^n = \binom{n}{r} = \frac{n!}{k!\left(n-k\right)!}$$

Example 9. The number of committees of 3 can be formed from 8 persons is

$$C_3^8 = \binom{8}{3} = \frac{8!}{3!\,(8-3)!} = 56$$

Permutation

An arrangement of set of n objects in a given order is called a permutation. An arrangements of any $(r \leq n)$ objects from n objects taken at a time is denoted by

$$P_r^n = A_r^n = \frac{n!}{(n-r)!}$$

Example 10. In one year, three awards (research, teaching, and service) will be given to a class of 30 graduate students in a Mathematics department. If each student can receive at most one award, how many possible selections are there?

Solution: Since the awards are distinguishable, it is a permutation problem. The total number of sample points is

$$P_3^{30} = A_3^{30} = \frac{30!}{(30-3)!} = 24360$$

Permutation with repetitions

The number of permutations of n objects of which n_1 are alike, n_2 are alike,... n_k are alike is given by

$$\frac{n!}{n_1!n_2!...n_k!}$$

Example 11. In a college football training session, the defensive coordinator needs to have 10 players standing in a row. Among these 10 players, there are 1 freshman, 2 sophomores, 4 juniors, and 3 seniors. How many different ways can they be arranged in a row if only their class level will be distinguished?

Solution: Directly using precedent relation, we find that the total number of arrangements is

$$\frac{10!}{1!2!3!4!} = 12600$$

1.1.6 Conditional probability

The probability of an event B occurring when it is known that some event A. has occurred is called a conditional probability and is denoted by $P(B|A)$. The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

Definition 9. The conditional probability of A given B, denoted by $P(A|B)$, is defined by

$$P(A|B) = \frac{P\left(A \cap B\right)}{P\left(B\right)}, P\left(A\right) > 0$$

Example 12. Assume that the population of adults in a small town who have completed the requirements for a college degree. We shall categorize them according to gender and employment status. The data are given in Table **2**.

Table 2. Categorization of the Adults in a Small Town.

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

We shall be concerned with the following events:

M: the one chosen is Male,

F: the one chosen is Female

E: the one chosen is employed.

T: the one is chosen.

Using the reduced sample space E, we find that

$$P(M|E) = \frac{P\left(M \cap E\right)}{P\left(E\right)} = \frac{460}{600} = \frac{23}{30}$$
$$P(F|E) = \frac{P\left(F \cap E\right)}{P\left(E\right)} = \frac{140}{600} = \frac{7}{30}$$

Using the reduced sample space total T, we find that

$$P(M|T) = \frac{P(M \cap T)}{P(T)} = \frac{500}{900} = \frac{5}{9}$$
$$P(F|T) = \frac{P(F \cap T)}{P(T)} = \frac{400}{600} = \frac{2}{3}$$

Independent events

Definition 10. Two events A and B are independent if and only if

$$P(B|A) = P(B) \; or \; P(A|B) = P(A)$$

,

assuming the existences of the conditional probabilities. Otherwise, A and B are **dependent.**

Definition 11. If in an experiment the events A and B can both occur, then

$$P(A \cap B) = P(A)P(B|A), P(A) > 0$$

Theorem 4. Two events A and B are independent if and only if

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

Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

Proposition 1.

Let $A, B \in \Omega$ be events with $P(A), P(B) > 0$. Then

1) $P(A \cap B) = P(A)P(B/A)$,

2) $P(A) = P(A/B)P(B) + P(A/B^c)P(B^c)$, $\left(B^c = \overline{B}\right)$.

3) $P(A^c/B) = 1 - P(A/B)$, $\left(A^c = \overline{A}\right)$.

Example 13. A small town has one fire engine and one ambulance available for emergencies. The probability that the fire engine is available when needed is 0.98, and the probability that the ambulance is available when called is 0.92. In the event of an injury resulting from a burning building, find the probability that both the ambulance and the fire engine will be available, assuming they operate independently.

Solution: Let A and B represent the respective events that the fire engine and the ambulance are available. Then

$$P(A \cap B) = P(A) \times P(B) = 0.98 \times 0.92 = 0.9016$$

Example 14. Phan wants to take either a Biology course or a Chemistry course. His adviser estimates that the probability of scoring an A in Biology is $\frac{4}{5}$, while the probability of scoring an A in Chemistry is $\frac{1}{7}$. If Phan decides randomly, by a coin toss, which course to take, what is his probability of scoring an A in Chemistry?

Solution: denote by B the event that Phan takes Biology, and by C the event that Phan takes Chemistry, and by $A =$ the event that the score is an A. Then, since $P(B) = P(C) = \frac{1}{2}$ we have

$$P(A \cap C) = P(C)\,P(A/C) = \frac{1}{2} \times \frac{1}{7} = \frac{1}{14}$$

The identity $P(A \cap B) = P(A)P(B/A)$ from Proposition **1** can be generalized to any number of events in what is sometimes called the multiplication rule.

Proposition 2 (Multiplication rule)

Let $A_1, A_2, ..., A_n \in \Omega$ be events. Then

$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2)...\ P(A_n/A_1 \cap A_2 \cap ... \cap A_{n-1})$.

Example 15. An urn has 5 blue balls and 8 red balls. Each ball that is selected is returned to the urn along with an additional ball of the same color. Suppose that 3 balls are drawn in this way.

(a) What is the probability that the 3 balls are blue?

(b) What is the probability that only 1 ball is blue?

Solution: a) In this case, we can de ne the sequence of events $B_1, B_2, B_3, ...$ where B_i is the event that the ith ball drawn is blue. Applying the multiplication rule yields

$$P(B_1 \cap B_2 \cap B_3) = P(B_1)\, P(B_2/B_1)\, P(B_3/B_1 \cap B_2) = \frac{5}{13} \times \frac{6}{14} \times \frac{7}{15} = \frac{1}{13}$$

b) denoting by R_i = the even that the ith ball drawn is red, we have

P (only 1 blue ball) $= P(B_1 \cap R_2 \cap R_3) + P(R_1 \cap B_2 \cap R_3) + P(R_1 \cap R_2 \cap B_3) = 3\frac{5\times 6\times 7}{13\times 14\times 15} = \frac{3}{13}$

1.1.7 Total probability

Theorem 5. (Bayes' Rule 1) If the events $B_1, B_2, ..., B_k$ constitute a partition of the sample space Ω such that $P(B_i) \neq 0$ for $i = 1, 2, ..., k$, then for any event A of Ω

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)\, P(A/B_i)$$

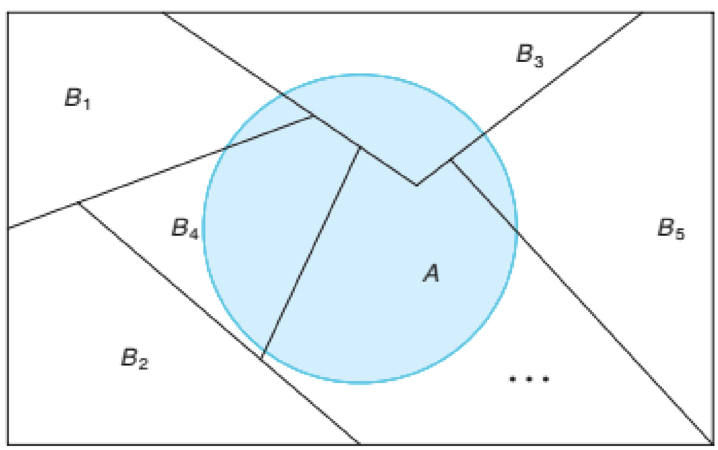


Figure 1.3: Partitioning the sample space Ω

Example 16. In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Solution: Consider the following events:

A: the product is defective,

B_1: the product is made by machine B_1,

B_2: the product is made by machine B_2,

B_3: the product is made by machine B_3.

Applying the rule of elimination, we can write

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$

Referring to the tree diagram of **Figure 1.4**, we find that the three branches give the probabilities

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

and hence

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245$$

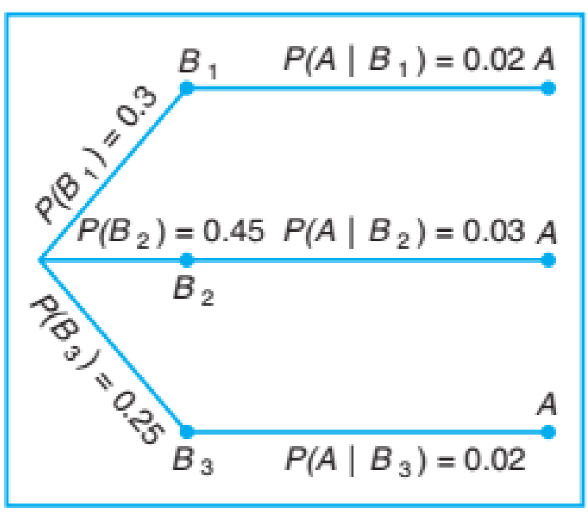


Figure 1.4: Tree diagram for Example 14.

Bayes' Rule

Theorem 6. If the events $A_1, A_2, ..., A_k$ constitute a partition of the sample space Ω such that $P(A_i) \neq 0$ for $i = 1, 2, ..., k$, then for any event B of Ω such that $P(B) \neq 0$

$$P(A_j/B) = \frac{P(A_j \cap B)}{\sum_{i=1}^{i=k} P(A_i \cap B)} = \frac{P(A_j)\, P(B/A_j)}{\sum_{i=1}^{i=k} P(A_i)\, P(B/A_i)}, j = 1, 2, ..., k$$

Example 17. With reference to Example 14, if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

Solution: Using **Bayes' rule** to write

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

and then substituting the probabilities calculated in Example 14, we have

$$P(B_3|A) = \frac{0.005}{0.006 + 0.0135 + 0.005} = \frac{0.005}{0.0245} = \frac{10}{49}$$

In view of the fact that a defective product was selected, this result suggests that it probably was not made by machine B_3.

1.2 Exercises with solutions

Exercise 1. Suppose the manufacturer's specifications for the length of a certain type of computer cable are 2000 ± 10 millimeters. In this industry, it is known that small cable is just as likely to be defective (not meeting specifications) as large cable. That is the probability of randomly producing a cable with length exceeding 2010 millimeters is equal to the probability of producing a cable with length smaller than 1990 millimeters. The probability that the production procedure meets specifications is known to be 0.99.

(a) What is the probability that a cable selected randomly is too large?

(b) What is the probability that a randomly selected cable is larger than 1990 millimeters?

Solution: Let M be the event that a cable meets specifications. Let S and L be the events that the cable is too small and too large, respectively. Then

(a) $P(M) = 0.99$ and $P(S) = P(L) = \dfrac{1-0.99}{2} = 0.005$.

(b) Denoting by X the length of a randomly selected cable, we have

$P(1990 \leq X \leq 2010) = P(M) = 0.99$.

Since $P(X \geq 2010) = P(L) = 0.005$,

$P(X \geq 1990) = P(M) + P(L) = 0.995$.

This also can be solved by using **Corollary 2.**$(P\left(\overline{A}\right) = 1 - P\left(A\right))$:

$P(X \geq 1990) + P(X < 1990) = 1$.

Thus, $P(X \geq 1990) = 1 - P(S) = 1 - 0.005 = 0.995$.

Exercise 2. Suppose that each child born is equally likely to be a boy or a girl. Consider a family with exactly **three** children.

1) List the 8 elements in the sample space whose outcomes are all possible genders of the three children.

2) Write each of the following events as a set and find its probability :

a) The event that exactly **one** child is a girl.

b) The event that at least **two** children are girls.

c) The event that no child is a girl.

Solution:

1) All possible genders are expressed as :

$\Omega = \{BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG\}$

2) a) Let A denote the event : A ={exactly one child is a girl}

$A = \{BBG, BGB, GBB\}$

$P(A) = \dfrac{3}{8}$

b) Let B denote the event that at least two children are girls.

$B = \{GGB, GBG, BGG, GGG\}$,

$P(B) = \frac{4}{8}$

c) Let C denote the event : C ={no child is a girl }

$C = \{BBB\}$

$P(C) = \frac{1}{8}$.

Exercise 3. A typical PIN (personal identification number) is a sequence of any four symbols chosen from the 26 letters in the alphabet and the 10 digits. If all PINs are equally likely, what is the probability that a randomly chosen PIN contains a repeated symbol?

Solution: A PIN is a sequence of four symbols selected from 36 (26 letters + 10 digits) symbols.

By the fundamental principle of counting, there are $36 \times 36 \times 36 \times 36 = 36^4 = 1,679,616$ PINs in all. When repetition is not allowed the multiplication rule can be applied to conclude that there are

$36 \times 35 \times 34 \times 33 = 1,413,720$ different PINs.

The number of PINs that contain at least one repeated symbol

$= 1,679,616 - 1,413,720 = 2,65,896$

Thus, the probability that a randomly chosen PIN contains a repeated symbol is

$$\frac{265,896}{1,679,616} = 0.1583$$

Exercise 4. An urn contains twenty white slips of paper numbered from 1 through 20, ten red slips of paper numbered from 1 through 10, forty yellow slips of paper numbered from 1 through 40, and ten blue slips of paper numbered from 1 through 10. If these 80 slips of paper are thoroughly shuffled so that each slip has the same probability of being drawn. Find the probabilities of drawing a slip of paper that is

(a) blue or white

(b) numbered $1, 2, 3, 4$ or 5

(c) red or yellow and numbered $1, 2, 3$ or 4

(d) numbered $5, 15, 25$, or 35

(e) white and numbered higher than 12 or yellow and numbered higher than 26.

Solution: (a) $P(BlueorWhite) = P(Blue) + P(White)(Why?) = \frac{10}{80} + \frac{20}{80} = \frac{3}{8}$

(b) $P(numbered 1, 2, 3, 4 \textit{ or } 5) = P$ (1 of any colour) $+P$ (2 of any colour)

$+P$ (3 of any colour) $+P$ (4 of any colour) $+P$ (5 of any colour)

$= \frac{4}{80} + \frac{4}{80} + \frac{4}{80} + \frac{4}{80}\frac{4}{80} = \frac{1}{4}$

(c) P (Red or yellow and numbered $1, 2, 3$ or 4) $=$

P (Red numbered $1, 2, 3$ or 4) $+P$ (yellow numbered 1, $2, 3$ or 4) $= \frac{4}{80} + \frac{4}{80} = \frac{1}{10} = 0.1$

(d) P (numbered $5, 15, 25$ or 35)$= P(5) + P(15) + P(25) + P(35)$

$= P$ (5 of White, Red, Yellow, Blue) $+P$ (15 of White, Yellow) $+P$ (25 of Yellow)

$+P$ (35 of Yellow)$=\frac{4}{80} + \frac{2}{80} + \frac{1}{80} + \frac{1}{80}$

(e) P (White and numbered higher than 12 or Yellow and numbered higher than 26)

$= P$ (White and numbered higher than 12)$+P$ (Yellow and numbered higher than 26)

$= \frac{8}{80} + \frac{14}{80} = \frac{11}{40}$.

Exercise 5. Suppose 30% of the women in a class received an A on the test and 25% of the men received an A. The class is 60% women. Given that a person chosen at random received an A, what is the probability this person is a women?

Solution: Let A be the event of receiving an A, W be the event of being a woman, and M the event of being a man. We are given $P(A/W) = 0.30$; $P(A/M) = 0.25$; $P(W) = 0.60$ and we want $P(W/A)$. From the definition

$$P(W/A) = \frac{P(W \cap A)}{P(A)}$$

We have

$$P(W \cap A) = P(A/W)P(W) = (0.30)(0.60) = 0.18$$

To find $P(A)$, we write

$$P(A) = P(W \cap A) + P(M \cap A) :$$

Since the class is 40% men,

$$P(M \cap A) = P(A/M)P(M) = (0.25)(0.40) = 0.10$$

So

$$P(A) = P(W \cap A) + P(M \cap A) = 0.18 + 0.10 = 0.28$$

Finally,

$$P(W/A) = \frac{P(W \backslash A)}{P(A)} = \frac{0.18}{0.28}$$

Exercise 6. Sarah and Bob draw 13 cards each from a standard deck of 52. Given that Sarah has exactly two aces, what is the probability that Bob has exactly one ace?

Solution: Let A = Sarah has two aces, and let B = Bob has exactly one ace. In order to compute $P(B/A)$, we need to calculate $P(A)$ and $P(A \cap B)$. On the one hand, Sarah could have any of $\binom{52}{13}$ possible hands. Of these hands, $\binom{14}{2}\binom{48}{11}$ will have exactly two aces so that

$$P(A) = \frac{\binom{4}{2}\binom{48}{11}}{\binom{52}{13}}$$

On the other hand, the number of ways in which Sarah can pick a hand and Bob another (different) is $\binom{52}{13} \times \binom{39}{13}$. The the number of ways in which A and B can simultaneously occur is $\binom{4}{2} \times \binom{48}{11} \times \binom{2}{1} \times \binom{37}{12}$ and hence

$$P(A \cap B) = \frac{\binom{4}{2} \times \binom{48}{11} \times \binom{2}{1} \times \binom{37}{12}}{\binom{52}{13}\binom{39}{13}}$$

Applying the definition of conditional probability we finally get

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\binom{4}{2} \times \binom{48}{11} \times \binom{2}{1} \times \binom{37}{12}}{\binom{52}{13}\binom{39}{13}} \times \frac{\binom{52}{13}}{\binom{4}{2}\binom{48}{11}} = \frac{2\binom{37}{12}}{\binom{39}{13}}$$

Exercise 7. A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02$$

where $P(D|P_j)$ is the probability of a defective product, given plan j.

If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

Solution: From the statement of the problem

$$P(P_1) = 0.30, P(P_2) = 0.20, P(P_3) = 0.50$$

,

we must find $P(P_j|D)$ for $j = 1, 2, 3$. Bayes' rule (Theorem 6) shows

$$\begin{aligned}
P(P_1|D) &= \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)} \\
&= \frac{(0.30)(0.01)}{(0.3)(0.01) + (0.20)(0.03) + (0.50)(0.02)} \\
&= \frac{0.003}{0.019} \\
&= 0.158
\end{aligned}$$

Similarly,

$P(P_2|D) = \dfrac{(0.03)(0.20)}{0.019} = 0.316$ and $P(P_3|D) = \dfrac{(0.02)(0.50)}{0.019} = 0.526$

The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3. Using Bayes' rule, a statistical methodology called the Bayesian approach has attracted a lot of attention in applications.

Exercice 8. One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

Solution: Let B_1, B_2, and W_1 represent, respectively, the drawing of a black ball from bag

1, a black ball from bag 2, and a white ball from bag 1. We are interested in the union of the mutually exclusive events $B_1 \cap B_2$ and $W_1 \cap B_2$. The various possibilities and their probabilities are illustrated in **Figure 1.5**. Now

$$
\begin{aligned}
P[(B_1 \cap B_2) \cup (W_1 \cap B_2)] &= P(B_1 \cap B_2) + P(W_1 \cap B_2) \\
&= P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\
&= 3769 + 4759 \\
&= 3863
\end{aligned}
$$

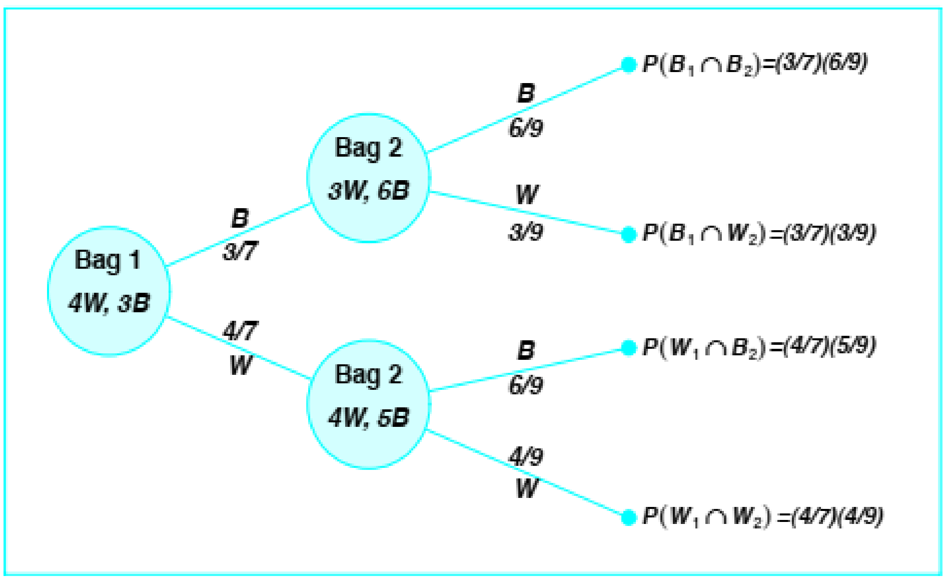


Figure 1.5: Tree diagram for Exercise 8.

Exercise 9. Suppose that 60% of UConn students will be at random exposed to the flu. If you are exposed and did not get a flu shot, then the probability that you will get the flu (after being exposed) is 80%. If you did get a flu shot, then the probability that you will get the flu (after being exposed) is only 15%.

(a) What is the probability that a person who got a flu shot will get the flu?

(b) What is the probability that a person who did not get a flu shot will get the flu?

Solution: a) Suppose we look at students who have gotten the flu shot.

Denote by

$E = \{$the event that a student is exposed to the flu$\}$, and by $F = \{$the event that a student gets the flu$\}$. We know that $P(E) = 0.6$ and $P(F/E) = 0.15$. This means that

$P(E \cap F) = (0.6)(0.15) = 0.09$, and it is clear that $P(E^c \cap F) = 0$. Since

$P(F) = P(E \cap F) + P(E^c \cap F)$,

we see that $P(F) = 0.09$.

b) Suppose we look at students who have not gotten the flu shot.

Let

$E = \{$the event that a student is exposed to the flu$\}$, and $F = \{$the event that a student gets the flu$\}$. We know that $P(E) = 0.6$ and $P(F/E) = 0.15$. This means that

$P(E \cap F) = (0.6)(0.8) = 0.48$, and it is clear that $P(E^c \cap F) = 0$. Since

$P(F) = P(E \cap F) + P(E^c \cap F)$,

we see that $P(F) = 0.48$.

Exercise 10. Two factories supply light bulbs to the market. Bulbs from factory X work for over 5000 hours in 99% of cases, whereas bulbs from factory Y work for over 5000 hours in 95% of cases. It is known that factory X supplies 60% of the total bulbs available in the market.

1) What is the probability that a purchased bulb will work for longer than 5000 hours?

2) Given that a light bulb works for more than 5000 hours, what is the probability that it was supplied by factory Y ?

3) Given that a light bulb work does not work for more than 5000 hours, what is the probability that it was supplied by factory X?

Solution: 1) let H be the event a bulb works over 5000 hours, X be the event that a bulb comes from factory X, and Y be the event a bulb comes from factory Y .

Then by the law of total probability

$$\begin{aligned} P(H) &= P(H/X)\,P(X) + P(H/Y)\,P(Y) \\ &= (0.99)(0.6) + (0.95)(0.4) \\ &= 0.974 \end{aligned}$$

2) By Part (1) we have

$$\begin{aligned} P(Y/H) &= \frac{P(H/Y)\,P(Y)}{P(H)} \\ &= \frac{0.95 \times 0.4}{0.974} \\ &= 0.39 \end{aligned}$$

3) We again use the result from Part (1)

$$\begin{aligned} P\left(X/\overline{H}\right) &= \frac{P\left(\overline{H}/X\right)P(X)}{P\left(\overline{H}\right)} \\ &= \frac{(1 - P(H/X))\,P(X)}{1 - P(H)} \\ &= \frac{(1 - 0.99)\,0.6}{1 - 0.974} \\ &= 0.23 \end{aligned}$$

Exercise 11. Two events A and B are such that $P(A) = 0.5$, $P(B) = 0.3$ and $P(A \cap B) = 0.1$. Calculate

(a) $P(A|B)$;

(b) $P(B|A)$;

(c) $P(A|A \cup B)$;

(d) $P(A|A \cap B)$;

(e) $P(A \cap B|A \cup B)$.

Solution: **(a)** $P(A|B) = P(A \cap B)/P(B) = \frac{1}{3}$

(b) $P(B|A) = P(A \cap B)/P(A) = \frac{1}{5}$

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7$, and the event $A \cap (A \cup B) = A$, so $P(A|A \cup B) = P(A)/P(A \cup B) = \frac{5}{7}$

(d) $P(A|A \cap B) = P(A \cap B)/P(A \cap B) = 1$, since $A \cap (A \cap B) = A \cap B$.

(e) $P(A \cap B|A \cup B) = P(A \cap B)/P(A \cup B) = \frac{1}{7}$, since $A \cap B \cap (A \cup B) = A \cap B$.

Chapter 2

Discrete Random Variables

2.1 Random variables

Random variables are the most useful objects of probability theory. Intuitively, random variables are numerical variables whose value is determined by underlying randomness. However, the formal definition of a random variable follows from the axiomatic approach in the previous section as follows.

Definition 1. A random variable is a scalar-valued function of the outcomes of a random experiment, i.e., a function that assigns elements in Ω to numbers

$$X : \Omega \longrightarrow numbers$$

We will use upper case letters, e.g., X, to denote random variables. Some may find this formal definition hard to reconcile with their intuition.In applications of probability theory, we often handle a large number of random variables. The temptation would be to view them each as the outcomes of individual random experiments. However, probablity theory, as stated in the previous section, is unable to deal with distinct random experiments. Viewing all random variables across space and time as functions of an underlying common random experiment has the advantage that all variables remain comparable within the framework of probability theory.

2.1.1 Discrete and continuous random variable

Definition 2. A random variable is called a **discrete random** variable if its set of possible outcomes is **countable**. In this case, random variable takes on discrete values, and it is possible to enumerate all the values it may assume.

Example 1. Statisticians use sampling plans to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values $0, 1, 2, ..., 9, 10$.

Definition 3. A random variable whose set of possible values is an entire interval of numbers is not discrete. When a random variable can take on values on a continuous scale, it is called a **continuous random** variable.

Example 2. Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

2.2 Discrete probability distributions

Definition 4. If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

The behavior of a random variable is characterized by its probability distribution, that is, by the way probabilities are distributed over the values it assumes.

A probability distribution function and a probability mass function are two ways to characterize this distribution for a discrete random variable. They are equivalent in the sense that the knowledge of either one completely specifies the random variable. The corresponding functions for a continuous random variable are the probability distribution function, defined in the same way as in the case of a discrete random variable, and the probability density function. The definitions of these functions now follow.

2.2.1 Probability distribution and cumulative probability function

For any random experiment with sample space Ω, conditions on random variables such as

- is X equal to 1?
- is Y smaller or equal to 1?

define events, in that the condition can be stated as (the set of outcomes for which X is 1) or (the
set of outcomes for which Y is 1 or less). Hence, we can rightfully use our probability measure to denote the probability of such events, e.g., $p(X = 1)$ and $p(Y \leq 1)$. We will use the notation

$$P_X(x) = P(X = x)$$

for the probability that the random variable X takes on the value x. P_X is called the **probability distribution** or the probability mass function (PMF) of X. Note that we use upper case letters to denote the random variable and lower case letters to denote possible values.
We will also use the notation

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

and call F_X the **cumulative probability function or distribution function** of X .
The event corresponding to this statement is the intersection of the two events and we use the notation

$$P_{XY}(x; y) = p(X = x \cap Y = y)$$

Example 3. Consider the throw of a fair dice, $\Omega = \{1, 2, 3, 4, 5, 6\}, n(\Omega) = |\Omega| = 6$.
• let X be a random variable that takes the value 0 if the outcome is even and 1 if the outcome is odd.

• let Y be a random variable that takes on the value 0 if the outcome is in $\{1,4\}$; 1 if the outcome is in $\{2,5\}$, and 2 if the outcome is in $\{3,6\}$.

The values of X are $\{0,1\}$, and the values of Y are $\{0,1,2\}$ i.e

$$X = \begin{cases} 0 & \text{if the outcome is even} \\ 1 & \text{if the outcome is odd} \end{cases}$$

and

$$Y = \begin{cases} 0 & \text{if the outcome in } \{1,4\} \\ 1 & \text{if the outcome in } \{2,5\} \\ 2 & \text{if the outcome in } \{3,6\} \end{cases}$$

The probability distribution of X and Y

x_i	0	1
$P(X = x_i)$	$\frac{3}{6}$	$\frac{3}{6}$

and

y_i	0	1	2
$P(X = y_i)$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$

We have

$$F_X(1) = P(x \leq 1) = P(X = 0) + P(X = 1) = \frac{3}{6} + \frac{3}{6} = 1$$
$$F_X(0) = P(x \leq 0) = P(X = 0) = \frac{3}{6}$$

Hence, the distribution function of the random variable X

$$F_X(x) = \begin{cases} 1 & \text{for } x \geq 1 \\ \frac{3}{6} & \text{for } 0 \leq x < 1 \\ 0 & \text{for } x < 0 \end{cases}$$

Venn diagrams for events indicate the subset of the sample space Ω corresponding to the event, i.e., the event is "true" for outcomes of the random experiment within the subset and "false" for outcomes outside the subset. Figure **2.1** shows what the equivalent for a random variable would be, though this representation is not comonly called a Venn diagram. The random variable partitions the sample space Ω into subsets corresponding to its possible values.

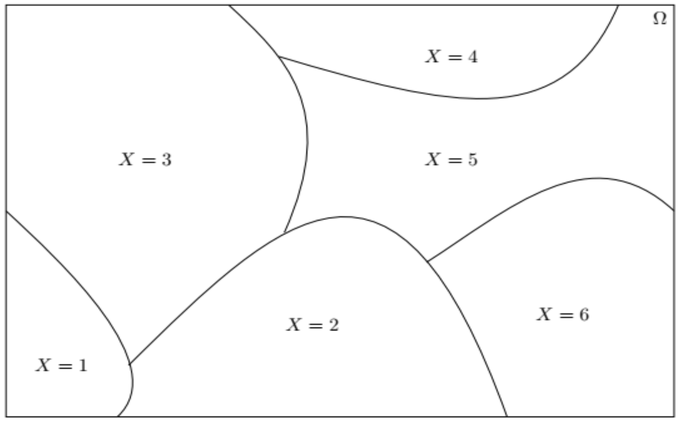


Figure 2.1: The "Venn Diagram" of a Random Variable.

Now that we've defined random variables and distributions, there are a number of derived measures of a distribution that are useful.

Mean

The first is the expectation of a random variable, also called the mean or average, defined as

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

where X is random variable and $x_1, x_2, ..., x_N$ are values of X which corresponding the probability measures $P(X = x_i), i = 1, 2, ..., N$.

Remark: We can also take the expectation of a function of a random variable

$$E(f(X)) = \sum_{i=1}^{N} f(x_i) P(X = x_i)$$

Properties of $E(X)$

1) If c is a constant then $E(c) = c$

2) If X and Y are random variables such that $X \leq Y$ then $E(X) \leq E(Y)$

3) $E(X + Y) = E(X) + E(Y)$

4) $E(X + c) = E(X) + c$

5) $E(aX) = aE(X)$

Variance of a distribution

There are a few expectations of particular interest:

$$E\left(X^2\right) = \sum_{i=1}^{N} x_i^2 P\left(X = x_i\right)$$

is called the **second moment** of a distribution. The expectation or mean $E\left(X\right)$ is also called the first moment. Another quantity of interest is called the central second moment or variance and is defined as

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

If a distribution is tightly concentrated around its mean, its variance will be small. Using linearity, we can re-write the **variance** as

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

Properties of $Var(X)$

1) $V(X + k) = V(X)$,

2) $Var(aX) = a^2 V(X)$

Standard deviation

It is defined as the square root of variance i.e,

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

Example 4. Let X be a random variable taking on the value of the outcome of a throw of a fair dice. Then

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

The probability distribution of X is given by the table below

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The mean of X is

$$E(X) = \sum_{i=1}^{6} x_i P(X = x_i) = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6}$$

Also we can calculate these quantities

$$E\left(X^2\right) = \sum_{i=1}^{6} x_i^2 P(X = x_i) = 1^2.\frac{1}{6} + 2^2.\frac{1}{6} + 3^2.\frac{1}{6} + 4^2.\frac{1}{6} + 5^2.\frac{1}{6} + 6^2.\frac{1}{6} = \frac{91}{6}$$

$$E(2X) = \sum_{i=1}^{6} 2x_i P(X = x_i) = 2\left(1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6}\right) = \frac{21}{3}$$

The variance and Standard deviation of X

$$Var(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = 2.9167$$

$$\sigma = \sqrt{Var(X)} = \sqrt{2.9157} = 1.7075$$

Covariance

If X and Y are two random variables then the extent to which two random variables vary together (co-vary) is measured by an indicator known as Covariance and it is denoted by $Cov(X, Y)$ and given by

$$\begin{aligned} Cov(X, Y) &= E([X - E(X)][Y - E(Y)]) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

where

$$E(XY) = \sum_{i=1}^{N} x_i y_i P_{XY}(x_i; y_i) = \sum_{i=1}^{N} x_i y_i p(X = x_i \cap Y = y_i)$$

Note:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Properties of the Covariance

If X, Y, W, and V are real-valued random variables and a, b, c, d are constant ("constant" in this context means non-random), then the following can be proved easily.

1) $Cov(X, a) = 0$,

2) $Cov(X, Y) = Cov(Y, X)$

3) $Cov(aX, bY) = abCov(X, Y)$

4) $Cov(X + a, Y + b) = Cov(X, Y)$

5) $Cov(aX + bY, cW + dV) = acCov(X, W) + adCov(X, Y) + bcCov(Y, W) + bdCov(Y, V)$

Correlation coefficient

The correlation coefficient ρ_{xy} between two random variables X and Y with expected values $\mu_X\,(E\,(X))$ and $\mu_Y\,(E\,(Y))$ and standard deviations σ_X and σ_Y is defined as

$$\begin{aligned}\rho_{xy} &= \frac{E\,(x-\mu_x)\,E\,(y-\mu_Y)}{\sigma_X \sigma_Y} = Corr\,(X, y) \\ &= \frac{Cov\,(X, Y)}{\sigma_X \sigma_Y}\end{aligned}$$

Example 5. A pair of fair dice is tossed. Let $X = max(a, b)$ and $Y = a + b$ where (a, b) is any ordered pair belongs to Ω. For calculate the covariance between X and Y using the table below.

$$p(X = 3 \cap Y = 5) = 2/36$$

		Y											Sum
		2	3	4	5	6	7	8	9	10	11	12	
X	1	1/36	0	0	0	0	0	0	0	0	0	0	**1/36**
	2	0	2/36	1/36	0	0	0	0	0	0	0	0	**3/36**
	3	0	0	2/36	2/36	1/36	0	0	0	0	0	0	**5/36**
	4	0	0	0	2/36	2/36	2/36	1/36	0	0	0	0	**7/36**
	5	0	0	0	0	2/36	2/36	2/36	2/36	1/36	0	0	**9/36**
	6	0	0	0	0	0	2/36	2/36	2/36	2/36	2/36	1/36	**11/36**
Sum		1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	

$$
\begin{aligned}
E(XY) &= 1 \times 2 \times \frac{1}{36} + 2 \times 3 \times \frac{2}{36} + 2 \times 4 \times \frac{1}{36} + ... + 6 \times 12 \times \frac{1}{36} \\
&= \frac{1232}{36} = 34.2 \\
E\left(X\right) &= 4.47, \sigma_X = 1.4 \\
E\left(Y\right) &= 7, \sigma_Y = 2.4
\end{aligned}
$$

Thus the covariance $Cov\left(X, Y\right)$ and the correlation coefficient ρ_{xy}

$$
\begin{aligned}
Cov\left(X, Y\right) &= 34.2 - 4.47 \times 7 = 2.91 \\
\rho_{xy} &= \frac{Cov\left(X, Y\right)}{\sigma_X \sigma_Y} = \frac{2.91}{2.4 \times 1.4} = 0.866
\end{aligned}
$$

2.3 Special discrete random variables

2.3.1 The Bernoulli distribution

We begin with a simple binary distribution. A binary random variable X with a probability distribution $P_X(1) = p$ and $P_X(0) = 1 - p = q$ is said to have a Bernoulli distribution with parameter p, denoted by

$$X \sim Ber\left(p\right)$$

We've already established that

$$E\left(X\right) = \sum_{i=1}^{2} p_i x_i = 1 \times p + 0 \times (1 - p) = p$$

for any binary random variable in the previous chapter. The other expectations of interest are the second moment

$$E\left(X^2\right) = \sum_{i=1}^{2} p_i x_i^2 = 1^2 \times p + 0^2 \times (1-p) = p$$

and the variance,

$$Var\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = p\left(1-p\right) = pq$$

Example 6. One dice was thrown. Let the random variable X be the number 6 shown by the dice face. Find

1) The probability distribution and the distribution function of X .

2) Calculate $E(X)$ and $Var(X)$.

3) Compute the probabilities $P(-3 \leq X < 0), P(-2 \leq X < 1), P(0 \leq X < 2)$ and $P(X \geq 3)$.

Solution:

If the number 6 appears, then $X = 1$ (success).

If the number $1, 2, 3, 4,$ or 5 appear, then $X = 0$ (failure).

Hence, $X \sim Ber(p)$ with $p = \frac{1}{6}$ and $q = 1 - p = \frac{5}{6}$

1) The probability distribution

$$P_X\left(k\right) = P\left(X = k\right) = \begin{cases} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{1-k}, k = 0, 1 \\ 0 \qquad\qquad\qquad \text{elsewhere} \end{cases}$$

The distribution function of X

$$F_x\left(x\right) = P\left(X \leq x\right) = \begin{cases} 0 & x < 0 \\ \frac{5}{6} & 0 \leq x < 1 \\ 1 & x > 1 \end{cases}$$

2) $E(X) = p = \frac{1}{6}$ and

$Var(X) = E(X^2) - [E(X)]^2 = p{\cdot}q = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$.

3)

- $P(-3 \leq X < 0) = 0$.
- $P(0 \leq X < 2) = p(X = 0) + p(X = 1) = \frac{5}{6} + \frac{1}{6} = 1$.
- $P(-2 \leq X < 1) = p(X = -2) + p(X = -1) + p(X = 0) = 0 + 0 + \frac{5}{6} = \frac{5}{6}$
- $P(X \geq 3) = 0$.

2.3.2 The Binomial distribution

Consider n independent $Ber(p)$ distributed random variables $X_1, X_2, ..., X_n$. The random variable $Y = \sum_{k=1}^{n} X_k$ is said to follow a binomial distribution with parameters n and p, denoted

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

We've now defined an expression for the binomial distribution

$$P_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} = C_k^n p^k (1-p)^{n-k}$$

where $\binom{n}{k} = C_k^n = \dfrac{n!}{k!\,(n-k)!}$ and $k = 0, 1, ..., n$.

Conditions For Applicability of Binomial Distributions:

1. Number of trials must be finite (n is finite)
2. The trails are independent
3. There are only two possible outcomes in any event i.e., success and failure.
4. Probability of success in each trail remains constant.

Examples:

1. Tossing a coin n times.
2. Throwing a die.
3. Number of defective items in the box.

For the binomial distribution with $n = 10$ and $p = 0.2$ are plotted in Figure **2.2**.

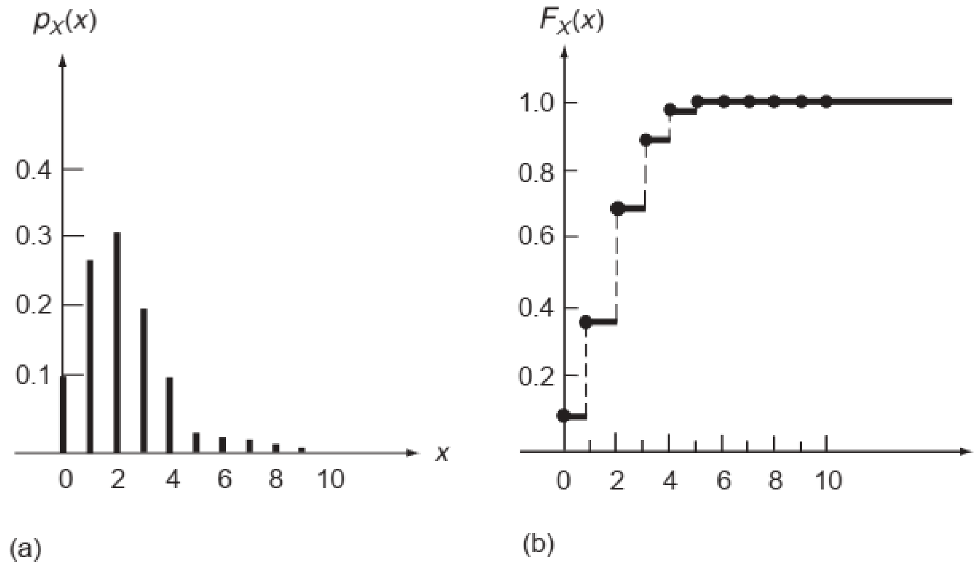


Figure 2.2: (a) Probability mass function, $P(X = x)$, and (b) probability distribution function, $F_X(x)$, for the discrete random variable X.

Mean of the Binomial Distribution

The expected value of the binomial distribution can be computed tediously using the definition of the expectation

$$E(Y) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

or, alternatively, we can use the linearity properties developed in the previous section to note that,

if $Y = X_1 + X_2 + ... + X_n$ where each X_k is $Ber(p)$, then

$$E(Y) = E(X_1) + E(X_2) + ... + E(X_n) = np$$

Variance of the Binomial Distribution

We can compute the second moment of a binomial distribution using the expectation product rule for independent random variables

$$E\left(Y^2\right) = E[(X_1 + X_2 + ... + X_n)^2] = nE\left(X_1^2\right) + 2\binom{n}{2} E[X_1]^2 = np + n(n-1)p^2$$

and the variance follows as

$$Var\left(Y\right)=E\left(Y^{2}\right)-\left(E\left(Y\right)\right)^{2}=np+n\left(n-1\right)p^{2}+np=np\left(1-p\right)$$

Example 7. A hypothetical student "C." has a probability of $p=\frac{1}{3}$ of failing to wake up in time for lectures on any given morning. What is the probability that "C." attends of the 6 IB Paper 7 "probability and statistics" lectures?.

The random variable Y counting the number of lectures attended is binomial $B(6,\frac{1}{3})$. Figure **2.3** illustrates the distribution graphically. Observe that this particular distributon has two modes (largest probabilities) at 1 and 2 lectures. In general, the mode of a random variable $X\sim B(n,p)$ is the largest integer smaller than np, except when np is an integer as is the case here, when you get two modes at np and $np-1$. We will compute the mean and standard deviation of binomial distributions below.

We have

$$X\sim B(n,p)=B\left(6,\frac{1}{3}\right)$$

yields

$$\begin{aligned}P_{X}\left(1\right)&=C_{1}^{6}\left(\frac{1}{3}\right)^{1}\left(1-\frac{1}{3}\right)^{6-1}=0.263\,37,k=1\\P_{X}\left(2\right)&=C_{2}^{6}\left(\frac{1}{3}\right)^{2}\left(1-\frac{1}{3}\right)^{6-2}=0.329\,22,k=2\end{aligned}$$

and

$$\begin{aligned}E\left(X\right)&=np=6\times\frac{1}{3}=2\\Var\left(X\right)&=np\left(1-p\right)=6\times\frac{1}{3}\left(1-\frac{1}{3}\right)=\frac{4}{3}\end{aligned}$$

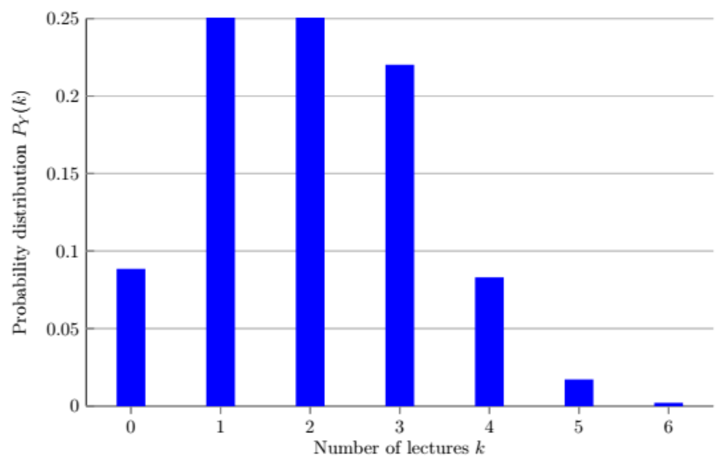


Figure 2.3: Binomial distribution $B(6, \frac{1}{3})$

2.3.3 The geometric distribution

The next two distributions we will study concern random variables taking on values in the countably infinite set of integers $1, 2, 3, ...$ as opposed to the Bernoulli and binomial distributions that both had finite sets of possible values. The geometric distribution can also be derived from a collection of independent Bernoulli random variable as the distribution of the index of the first 1 in the sequence. Hence, Y is a geometric distributed random variable

$$Y \sim Geom(p)$$

derived from an infinite collection $X_1, X_2, ...$ of independent $Ber(p)$ random variables, and therefore

$$P_Y(k) = p(1-p)^{k-1}, k = 1, 2, ..$$

i.e., the probability that the first k - 1 variables are zero and that the k-th variable is one. We leave it as an exercise to verify that

$$E(Y) = \frac{1}{p}$$

The variance can also be obtained by the relation

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

Exemple 8. For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the 5 item inspected is the first defective item found?

Solution: Using the geometric distribution with $k = 5$ and $p = 0.01$, we have

$$P_X(5) = 0.01\,(1 - 0.01)^{5-1} = 0.01 \times 0.99^4 = 0.0096$$

and

$$\begin{aligned} E(X) &= \frac{1}{p} = \frac{1}{0.01} = 100 \\ Var(X) &= \frac{1-p}{p^2} = \frac{1-0.01}{0.01^2} = 9900 \end{aligned}$$

2.3.4 The Poisson distribution

The Poisson distribution, named after the French scientist Siméon Poisson, is used to model the probability of the number of incidents in a time interval when incidents happen independently at a given rate of λ incidents per time interval. The incidents in this context are assumed to be of zero duration.

Rather than postulate an expression for the Poisson distribution, we will derive it from first principles. Let Y be the random variable counting the number of incidents in the time interval of interest.

$$Y \sim Poisson(\lambda)$$

Hence, we can define indicator random variables $X_1, X_2, ..., X_n$ that are 1 if an incident occurs in the corresponding sub-interval and 0 otherwise. This is therefore well approximated by a sequence of independent Bernoulli random variables with parameter $\lambda = n$, i.e., $X_k \sim Ber(\frac{\lambda}{n})$ for $k = 1, 2, .., n$, and the approximation is exact in the limit as n goes to infinity. Since $Y = X_1 + X_2 + ... + Xn$, Y follows a binomial distribution $Y \sim B(n; \frac{\lambda}{n})$. The **Poisson distribution** is the limit of a binomial distribution $B(n; \frac{\lambda}{n})$ as n goes to infinity. We need to evaluate

$$P_Y(k) = \lim_{n \longrightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

This limit gives

$$P_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, ...$$

Since we've derived the **Poisson distribution** as an approximation to a binomial, it is clear that the Poisson distribution can also be used in reverse to approximate a binomial distribution. This works for small p and large n and is useful when the binomial distribution is difficult to compute numerically due to the difficulty of evaluating $\binom{n}{k}$. Hence, for large n and small p,

$$B(n, p) = \binom{n}{k} (p)^k (1 - p)^{n-k} \simeq \frac{(pn)^k}{k!} e^{-pn} = Poisson(np)$$

i.e., a Poisson distribution with parameter $\lambda = pn$.

Conditions For Poisson Distribution

1. The number of trials are very large (infinite)
2. The probability of occurrence of an event is very small $(\lambda = np)$
3. $\lambda = np =$ finite

Examples:

1. The number of printing mistakes per page in a large text
2. The number of telephone calls per minute at a switch board
3. The number of defective items manufactured by a company.

For the Poisson distribution with some values of $\lambda\,(\lambda = 1, 4, 10)$ are plotted in Figure **2.4**.

The Poisson distribution unusally has identical expectation and variance (which we won't derive but is again a fun exercise if you enjoy juggling around with algebra and derivatives)

$$E(Y) = Var(Y) = \lambda$$

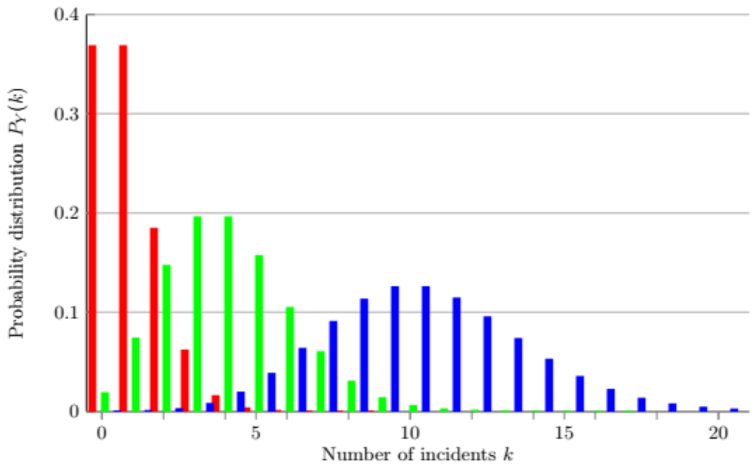


Figure 2.4: Poisson distributions for $\lambda = 1$ (red), $\lambda = 4$ (green) and $\lambda = 10$ (blue)

Example 9. During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

Solution: Using the Poisson distribution (X) with $k = 6$ and $\lambda = 4$, we have

$$P_X(k) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{4^6}{6!}e^{-4} = 0.1042$$

and

$$E(X) = Var(X) = \lambda = 4$$

Example 10. The number of errors in an essay ismodelled by a poisson distribution with 3.8 errors per page on overage. Find the probability that there are

1) no error on the next page.

2) Fewer than 2 errors on the 5 page.

3) more than 7 errorson the last 2 page.

Solution:

X represents the number of errors made on a page on an essay

$$X \sim P_0 (3.8)$$

1) $P(X = 0) = e^{-3.8} = 2.2371 \times 10^{-2}$.

2) $P(X < 2) = P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-3.8} + 3.8e^{-3.8} = 0.10738$.

Y represents the number of errors made on two page on an essay

$$Y \sim P_0 (7.6)$$

3) $P(Y > 7) = 1 - P(Y < 7) = 1 - 0.51 = 0.49$.

2.4 Exercices with solutions

Exercise 1. Suppose that a coin is tossed three times and the sequence of heads and tails is noted.Now let the random variable X be the number of heads in three coin tosses.

1) Determine the sample space Ω.

2) Determine the probability distribution of X.

3) Calculate the probability mesures $P(X \leq 2), P(X > 2), P(0 \leq X < 2)$.

4) Determine mean, variance and Standard deviation of random variable X.

Solution:

1) The sample space for this experiment evaluates to:

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ and $n(\Omega) = 8$

2) The random variable X assigns each outcome in Ω a number from the set $x_i = \{0, 1, 2, 3\}$.

The table below lists the eight outcomes of Ω and the corresponding values of X

Outcome	HHH	HHT	HTH	THH	HTT	TTH	THT	TTT
X	3	2	2	2	1	1	1	0

The table below shows the probability distribution of X

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

3) We have $P(X \leq 2) = P(X = 2) + P(X = 1) + P(X = 0) = \frac{2}{8} + \frac{3}{8} + \frac{1}{8} = \frac{3}{4}$

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

$P(0 \leq X < 2) = P(X = 1) + P(X = 0) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}$

4) The mean of X

$$E(X) = \sum_{i=1}^{4} x_i P(X = x_i) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{2}{8} + 3 \times \frac{1}{8} = \frac{10}{8}$$

The second moment of X

$$E\left(X^2\right) = \sum_{i=1}^{4} x_i^2 P(X = x_i) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{3}{8} + 2^2 \times \frac{2}{8} + 3^2 \times \frac{1}{8} = \frac{20}{8}$$

The variance of X

$$Var(X) = E\left(X^2\right) - (E(X))^2 = \frac{20}{8} - \left(\frac{10}{8}\right)^2 = 0.937\,5$$

Exercise 2. A random variable X has the following probability distribution

x_i	-3	6	9
$P(X = x_i) = p_i$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

1) Find the mean $E(X)$,

2) Determine $E(X^2), Var(X)$

3) Evaluate $E\left((2X+1)^2\right)$

Solution:

1) The mean

$$E(X) = \sum_{i=1}^{3} x_i p_i = -3 \times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3} = \frac{11}{2}$$

2) we calculate $E(X^2), Var(X)$

The second moment of X

$$E\left(X^2\right)=\sum_{i=1}^{4}x_i^2p_i=(-3)^2\times\frac{1}{6}+6^2\times\frac{1}{2}+9^2\times\frac{1}{3}=\frac{93}{2}$$

The variance of X

$$Var\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=\frac{93}{2}-\left(\frac{11}{2}\right)^2=16.25$$

3) $E\left((2X+1)^2\right)=E\left(4X^2+4X+1\right)=4E\left(X^2\right)+4E\left(X\right)+1=4\times\frac{93}{2}+4\times\frac{11}{2}+1=209$

Exercise 3. Let X be a random variable of **maximum** of two numbers in throwing two fair dice simultaneously.

1) Find the probability distribution of X

2) Determine the mean

3) Determine the variance

4) Compute $P(1<x<4)$

5) Compute $P(2\leq X\leq 4)$

Solution: 1) Sample space of throwing two dices

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

$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)$

$(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)$

$(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)$

$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)$

$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$

$n\left(\Omega\right)=36.$

Let $X=$ Maximum of two numbers in throwing two dice $=\{1,2,3,4,5,6\}$

X	Favorable cases	Number of Favorable cases	p_i
1	$(1,1)$	1	$\frac{1}{36}$
2	$(2,1),)(1,2),(2,2)$	3	$\frac{3}{36}$
3	$(3,1),(1,3),(2,3)(3,3),(3,2)$	5	$\frac{5}{36}$
4	$(1,4),(4,1),(4,2),(2,4)(4,3),(3,4),(4,4)$	7	$\frac{7}{36}$
5	$(1,5),(5,1),(2,5),(5,2)(3,5),(5,3),(5,4),(4,5),(5,5)$	9	$\frac{9}{36}$
6	$(1,6)(6,1),(6,2),(2,6),(6,3),(3,6),(4,6),(6,4),(6,5)(5,6),(6,6)$	11	$\frac{11}{36}$

Clearly ,$p(x_i) > 0$ and $\sum_{i=1}^{6} p(x_i) = 1$

Probability distribution of X is given by

x_i	1	2	3	4	5	6
p_i	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

2) The mean of X

$$E\left(X\right) = \sum_{i=1}^{6} x_i p_i = 1\left(\frac{1}{36}\right) + ... + 6\left(\frac{11}{36}\right) = 4.47$$

The variance of X

$$Var\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = 1^2\left(\frac{1}{36}\right) + ... + 6^2\left(\frac{11}{36}\right) - \left(4.47\right)^2 \simeq 2$$

4) $P(1 < x < 4) = p_2 + p_3 = \frac{3}{36} + \frac{5}{36} = \frac{2}{9}$

5) $P(2 \leq X \leq 4) = p_2 + p_3 + p_4 = \frac{3}{36} + \frac{5}{36} + \frac{7}{36} = \frac{15}{36}$

Exercise 4. A random variable X has the following probability function

x_i	-3	-2	-1	0	1	2	3
p_i	k	0.1	k	0.2	$2k$	0.4	$2k$

1) Find the number k,

2) Calculate the mean and the variance of X.

Solution: 1) We know that $\sum_{i=1}^{7} p_i = 1$ i.e $k + 0.1 + k + 0.2 + 2k + 0.4 + 2k = 1$

i.e $6k + 0.7 = 1$, we get $k = 0.05$

2) The mean $E(X) = \sum_{i=1}^{7} x_i p_i = k(-3) + 0.1(-2) + k(-1) + 2k(1) + 2(0.4) + 3(2k) = 0.8$

Variance $Var(X) = E(X^2) - (E(X))^2 = k(-3)^2 + 0.1(-2)^2 + k(-1)^2 + 2k(1)^2 + 4(0.4) + 9(2k) - 0.8^2 = 2.86$.

Exercise 5. In tossing a coin 10 times simultaneously. Find the probability of getting

a) at least 7 heads

b) almost 3 heads

c) exactly 6 heads

Solution: Given $n = 10$

Probability of getting a head in tossing a coin $p = \frac{1}{2}$

Probability of getting no head $q = 1 - p = \frac{1}{2}$.

The random variable X to follow a binomial distribution with parameters $n = 10$ and $p = \frac{1}{2}$

i.e $X \sim B\left(10, \frac{1}{2}\right)$ and $P(X = k) = \binom{10}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10-k}$, $k = 0, 1, ..., 10$.

1) Probability of getting at least seven heads is given by

$$\begin{aligned}
P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\
&= \binom{10}{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7} + \binom{10}{8}\left(\frac{1}{2}\right)^{8}\left(\frac{1}{2}\right)^{10-8} + \binom{10}{9}\left(\frac{1}{2}\right)^{9}\left(\frac{1}{2}\right)^{10-9} + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right. \\
&= \frac{1}{2^{10}}\left(C_7^{10} + C_8^{10} + C_9^{10} + C_{10}^{10}\right) = \frac{1}{2^{10}}(120 + 45 + 10 + 1) = \frac{176}{1024} \simeq 0.1719
\end{aligned}$$

2) Probability of getting at most 3 heads is given by

$$\begin{aligned}
P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\
&= \binom{10}{0}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{10-0} + \binom{10}{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{10-1} + \binom{10}{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{10-2} + \binom{10}{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right. \\
&= \frac{1}{2^{10}}\left(C_7^{10} + C_8^{10} + C_9^{10} + C_{10}^{10}\right) = \frac{1}{2^{10}}(120 + 45 + 10 + 1) = \frac{176}{1024} \simeq 0.1719
\end{aligned}$$

3) Probability of getting exactly 6 heads is given by

$$P(X=6)=\binom{10}{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{10-6}=0.205$$

Exercise 6. A hospital switch board receives an average of 4 emergency calls in a 10−minute interval. What is the probability that

1) there are at most 2 emergency calls in a 10 minute interval

2) there are exactly 3 emergency calls in a 10−minute interval.

Solution: Mean, $\lambda=$ (4 calls/10 minutes) = 4 calls, then the random variable X to follow a Poisson distribution with parameter $\lambda = 4$ i.e

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}=\frac{4^k}{k!}e^{-4}$$

1) P (at most 2 calls) $= P(X \leq 2)$

$= P(X=0)+P(X=1)+P(X=2)$

$= \frac{4^0}{0!}e^{-4}+\frac{4^1}{1!}e^{-4}+\frac{4^2}{2!}e^{-4}=0.2381$

2) P (exactly 3 calls) $= P(X=3)=\frac{4^3}{3!}e^{-4}=0.1954.$

Exercise 7. You play a game of chance that you can either win or lose (there are no other possibilities) until you lose. Your probability of losing is $p = 0.57$. What is the probability that it takes 5 games until you lose? Let $X =$ the number of games you play until you lose (includes the losing game).

Calculate $E(X)$ and $Var(X)$.

Solution: Let $X =$ the number of games you play until you lose (includes the losing game). Then X takes on the values $1, 2, 3, ...$ (could go on indefinitely). The probability question is $P(X = 5)$.

Using the geometric distribution with $k = 5$ and $p = 0.57$, we have

$$P_X(5)=p(1-p)^{k-1}=0.57(1-0.57)^{5-1}=0.57\times 0.43^4=0.01948$$

and

$$E(X) = \frac{1}{p} = \frac{1}{0.57} = 1.7544$$
$$Var(X) = \frac{1-p}{p^2} = \frac{1-0.57}{0.57^2} = 1.3235$$

Exercise 8. A manufacturer of television set known that on an average 5% of their product is defective. They sell television sets in consignment of 100 and guarantees that not more than 2 set will be defective. What is the probability that the TV set will fail to meet the guaranteed quality?

Solution: The probability of TV Set to be defective $p = 5\% = 0.05$

Total number of TV sets , $n = 100$

Mean, $\lambda = np = 100(0.05) = 5$

We have by Poisson distribution $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{5^k}{k!}e^{-5}$

P (a TV set will fail to meet the guarantee) $= P(X > 2) = 1 - P(X \leq 2)$

$= 1 - (P(X = 0) + P(X = 1) + P(X = 2)) = 1 - \left(\frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5} + \frac{5^2}{2!}e^{-5}\right) = 0.8753.$

Exercise 9. The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, we assume that X follow the binomiale distribution. What is the probability that

(a) at least 10 survive,

(b) from 3 to 8 survive,

(c) exactly 5 survive?

Solution: Let X be the number of people who survive which $X \sim B(15, 0.4)$.

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\,(n = 15, p = 0.4)$

(a) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{k=0}^{9} P(X = k) = 1 - 0.9662 = 0.0338$

(b) $P(3 \leq X \leq 8) = \sum_{k=3}^{8} P(X = k) = 0.9050 - 0.0271 = 0.8779$

(c) $P(X = 5) = \binom{15}{5} 0.4^5 (1 - 0.4)^{15-5} = 0.1859$

Chapter 3

Continuous Random Variables

3.1 Fundamentals of continuous random variables

Definition 1. A random variable whose set of possible values is an entire interval of numbers is not discrete. When a random variable can take on values on a continuous scale, it is called a **continuous random** variable.

Example 1. Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

3.1.1 Cumulative probability function

We will in general refrain from considering events with a finite number of outcomes and hence not consider probability distributions $P_X(x) = p(X = x)$ as we did for discrete random variables. This distribution would in general be zero everywhere and hence be of little use. However, we can still consider events corresponding to intervals, and in particular, use the **cumulative probability function** defined in Chapter II.

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

This is still well defined and satisfies the following properties

$$\begin{cases} F_X\left(x\right) \geq 0 \\ \lim\limits_{x \longrightarrow +\infty} F_X\left(x\right) = 1 \\ \lim\limits_{x - \infty} F_X\left(x\right) = 0 \\ F_X(x) \text{ is non-decreasing in } x \end{cases}$$

Example 2. The probability that the car speed recorded by speed camera A is zero or less (assuming that the camera records the magnitude of the speed only) is 0, hence $F_X(x) = 0$ for all $x \leq 0$. The probability that the speed is less than 1000 mph is 1 as there are no cars that can drive that fast, hence $F_X(1000) = \lim\limits_{x \longrightarrow +\infty} F_X\left(x\right) = 1$.

Properties: For $a, b \in \mathbb{R}^+$ and let X is continuous random variable

1) $P\left(X \geq a\right) = 1 - P\left(X \leq a\right) = 1 - F_X\left(a\right)$

2) $P\left(X \leq -a\right) = 1 - P\left(X \leq a\right) = 1 - F_X\left(a\right)$

3) $P\left(a \leq X \leq b\right) = P\left(X \leq b\right) - P\left(X \leq a\right) = F_X\left(b\right) - F_X\left(a\right)$

Joint cumulative probability functions are defined similarly to joint distributions for discrete random variables

$$F_{XY}(x; y) = p(X \leq x \cap Y \leq y)$$

and independence of continuous random variables is defined as the independence of all events associated with the random variables, e.g.,

$$F_{XY}(x; y) = p(X \leq x \cap Y \leq y) = p(X \leq x)p(Y \leq y) = F_X(x)F_Y(y)$$

for all $(x, y) \in X \times\ Y$, where the condition "for all" specifies an infinity of conditions in the continuous case.

3.1.2 The probability density function

For a discrete random variable X taking values over a set $S = \{x_1, x_2, ..., x_n\}$ such that

$x_1 \leq x_2 \leq ... \leq x_n$, we have

$$P_X(x_k) = F_X(x_k) - F_X(x_{k-1}), \text{ for } k = 2, 3, ..., n$$

Can we apply a similar approach to continuous random variables in order to obtain something along the lines of a probability distribution from the well defined cumulative probability function?

If we try to apply the expression above to a continuous random variable for an infinitesimal interval, we obtain in general

$$\lim_{\Delta x \longrightarrow 0} \left(F_X(x + \Delta x) - F_X(x)\right) = 0$$

Definition 5. The probability density function noted by $f_X(x)$ is the derivative of the cumulative probability function $F_X(x)$ defined by

$$f_X(x) = \frac{dF_X}{dx} = F'_X(x)$$

Conditions of the probability density function $f_X(x)$

1) $f_X(x) \geq 0$ and f is continuous for all x

2) $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$

3) $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx$

The probability density function is a very useful quantity that often takes precedence over the cumulative probability function in people's perception because it acts like a probability distribution in helping us visualise the probabilistic behaviour of random variables. Figure **3.1** shows the cumulative probability distribution and the probability density function for two random variables.

Joint probability density functions are obtained from the cumulative density function through a multiple differentiation, e.g.,

$$f_{XY}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y)$$

and independent random variables satisfy

$$f_{XY}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y) = f_X(x)\, f_Y(y)$$

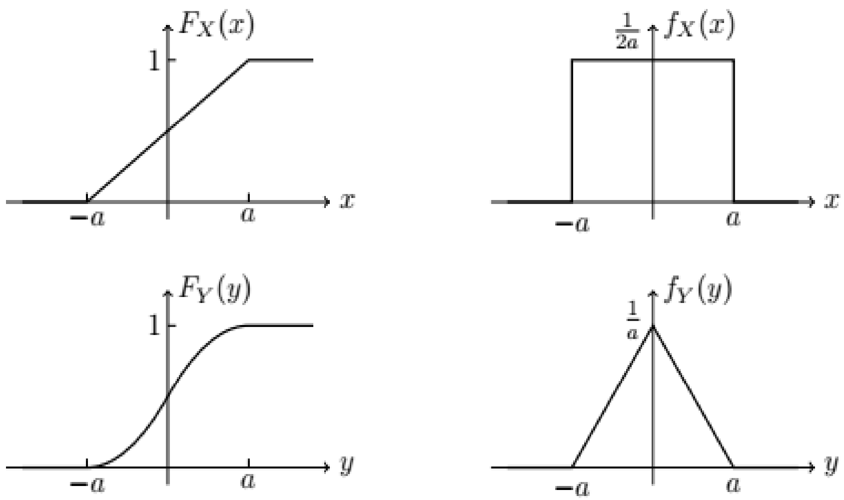


Figure 3.1: Cumulative probability function (left) and probability density function (right) for two random variables, X uniform and Y triangular.

Mean of the random variable

The probability density function can also be used to compute expectations

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

for any function $g(.)$. In particular, we will be interested in the **mean** (also called "first moment")

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

the "second moment"

$$E\left(X^2\right)=\int_{-\infty}^{+\infty}x^2 f_X\left(x\right)dx$$

and the $n-th$ moment

$$E\left(X^n\right)=\int_{-\infty}^{+\infty}x^n f_X\left(x\right)dx$$

Variance of random variable

The variance is defined by the same relation

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

Example 3. The probability density function of X is denoted by $f\left(x\right)$, and is given by

$$f\left(x\right)=\begin{cases} kx & 0\leq x\leq 12 \\ 0 & \text{elsewhere}\end{cases}$$

1) Show that $k=\frac{1}{72}$.

2) Determine $P(X>5)$.

3) Show by calculation that $E\left(X\right)=Var\left(X\right)$.

Solution:

1) By definition

$$\int_0^{12}kxdx=1$$
$$\left[k\frac{x^2}{2}\right]_0^{12}=1$$

yields $k=\frac{1}{72}$.

2) $P(X>5)=\int_5^{12}\frac{1}{72}xdx=\frac{119}{144}$.

3) $E\left(X\right)=\int_0^{12}x\frac{1}{72}xdx=\int_0^{12}\frac{1}{72}x^2dx=8$ and $E\left(X^2\right)=\int_0^{12}x^2\frac{1}{72}xdx=\int_0^{12}\frac{1}{72}x^3dx=72$

we have $Var\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=72-8^2=8$ we get $E\left(X\right)=Var\left(X\right)$.

3.2 Special continuous random variables

3.2.1 Continuous uniform distribution

One of the simplest continuous distributions in all of statistics is the continuous uniform distribution. This distribution is characterized by a density function that is "flat," and thus the probability is uniform in a closed interval, say $[a, b]$. Although applications of the continuous uniform distribution are not as abundant as those for other distributions discussed in this chapter, it is appropriate for the novice to begin this introduction to continuous distributions with the uniform distribution.

Definition 3. The density function of the continuous uniform random variable X on the interval $[a, b]$ is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$

The density function forms a rectangle with base $b - a$ and constant height $\frac{1}{b-a}$. As a result, the uniform distribution is often called the rectangular distribution. Note, however, that the interval may not always be closed: $[a, b]$. It can be (a, b) as well. The density function for a uniform random variable on the interval $[1, 3]$ is shown in Figure **3.2**. Probabilities are simple to calculate for the uniform distribution because of the simple nature of the density function. However, note that the application of this distribution is based on the assumption that the probability of falling in an interval of fixed length within $[a, b]$ is constant.

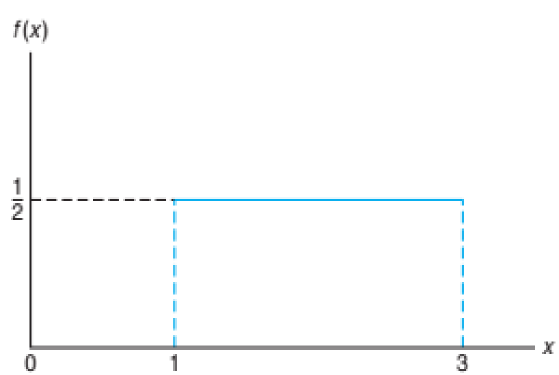


Figure 3.2: The density function for a random variable on the interval $[1, 3]$.

Mean and variance of uniform distribution

The mean and variance of the uniform distribution are

$$E\left(X\right)=\frac{a+b}{2},Var\left(X\right)=\frac{\left(a-b\right)^{2}}{12}$$

Example 4. Suppose that a large conference room at a certain company can be reserved for no more than 5 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval $[0,5]$.

1) Determine the probability density function?

2) What is the probability that any given conference lasts at least 3 hours?

3) Calculate mean and variance of this distribution.

Solution: 1) The appropriate density function for the uniformly distributed random variable X in this case is

$$f\left(x\right)=\begin{cases}\frac{1}{5}, & 0\leq x\leq 5\\ 0 & \text{elsewhere}\end{cases}$$

2) The probability that any given conference lasts at least 3 hours

$$P\left(X\geq 3\right)=\int_{3}^{5}f\left(x\right)dx=\left[\frac{1}{5}x\right]_{3}^{5}=1-\frac{3}{5}=\frac{2}{5}$$

3) We have

$$E\left(X\right)=\frac{a+b}{2}=\frac{0+5}{2}=2.5$$
$$Var\left(X\right)=\frac{\left(a-b\right)^{2}}{12}=\frac{\left(0-5\right)^{2}}{12}=\frac{25}{12}$$

3.2.2 The exponential density

The exponential density can be derived from the Poisson distribution we studied in the previous lecture. The exponential density for the continuous random variable X is defined by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, \lambda \geq 0, x \geq 0 \\ 0 \ \text{ elsewhere} \end{cases}$$

and hence

$$F_X(x) = 1 - e^{-\lambda x}, \lambda \geq 0, x \geq 0$$

The exponential random variable is noted by

$$X \sim Exp(\lambda)$$

Remark: The Poisson distribution is

$$P_Y(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Note that the λ in our original expression for the Poisson distribution has been replaced by λt in this expression simply because we are considering an interval of length t and λ is the rate of arrival per unit of time, it was the rate of arrival for the time interval considered.
We express the cumulative probability function of the inter-arrrival time as

$$F_X(x) = P(X \leq t) = 1 - P(X > t)$$

X can only be larger than t if no arrivals occur in the interval, implying

$$P(X > t) = P_Y(0) = e^{-\lambda t}$$

and hence

$$F_X(x) = 1 - e^{-\lambda t}$$

We now derive the probability density function of X above.
As mentioned above, the exponential density is used to model the time intervals in a Poisson process with independent arrival times. Figure **3.3** shows the probability density functions

for exponentially distributed random variables for various values of λ.

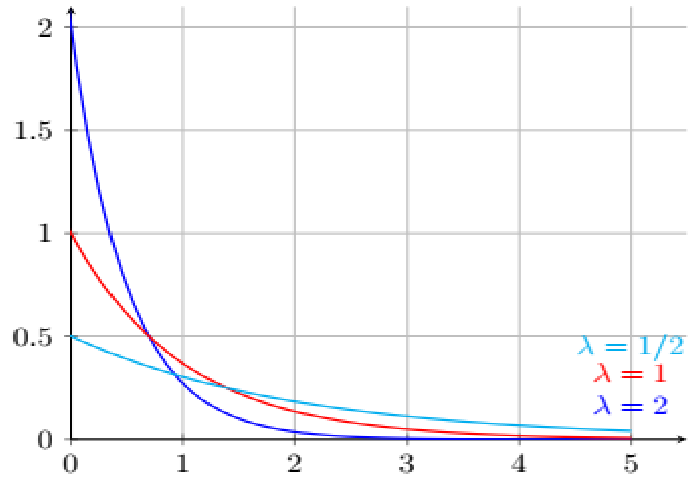


Figure 3.3: Exponential probability density functions for $\lambda = 2, 1$ and $\frac{1}{2}$.

The mean of exponential variable

The mean of an exponential density function can be calculated easily using integration by parts

$$E\left(X\right) = \int_0^\infty t f_X\left(t\right) dt = \frac{1}{\lambda}$$

and the second moment

$$E\left(X^2\right) = \int_0^\infty t^2 f_X\left(t\right) dt = \frac{2E\left(X\right)}{\lambda} = \frac{2}{\lambda^2}$$

The variance of exponential variable

By using the same relation, we get

$$Var\left(X\right) = E\left(X^2\right) - \left(E\left(X\right)\right)^2 = \frac{1}{\lambda^2}$$

Example 5. Suppose that on a certain stretch of highway, cars pass at an average rate of 5 cars per minute. Assume that the duration of time between successive cars follows the exponential distribution.

1) On average, how many seconds elapse between two successive cars?

2) After a car passes by, how long on average will it take for another 7 cars to pass by?

3) Find the probability that after a car passes by, the next car will pass within the next 20 seconds.

4) Find the probability that after a car passes by, the next car will not pass for at least another 15 seconds.

Solution:

1) At a rate of five cars per minute, we expect $\frac{60}{5} = 12$ seconds to pass between successive cars on average.

2) Using the answer from part a, we see that it takes $12 \times 7 = 84$ seconds for the next seven cars to pass by.

3) Let $X-$the time (in seconds) between successive cars.

The mean of X is 12 seconds, so the decay parameter is $\frac{1}{12}$and $X \sim \exp\left(\frac{1}{12}\right)$. The cumulative distribution function of X is $P\left(X < x\right) = 1 - e^{-\frac{1}{12}x}$. Then $P\left(X < 20\right) = 1 - e^{-\frac{1}{12}20} = 0.811\,12$.

4) $P\left(X > 15\right) = 1 - P\left(X < 15\right) = 1 - \left(1 - e^{-\frac{1}{12}15}\right) = e^{-\frac{15}{12}} = 0.286\,5$.

3.2.3 The Gaussian density

The Gaussian density, named after the German mathematician Carl Friedrich Gauss (1777 − 1855), is of central importance in the study of probability It arises whenever a quantity is the sum of many smaller effects. Theremore, the most important continuous probability distribution in the entire field of statistics is the **normal distribution**. Its graph, called the **normal curve**, is the bell-shaped curve of Figure **3.4**, which approximately describes many phenomena that occur in nature, industry, and research. It is a good model (sometimes provably accurate) for many quantities, such as

- the velocities of particles in an ideal gas,
- noise added to a signal by a variety of parasitic effects in a communications receiver,
- measurement noise,
- quantities in biology (e.g. height or weight of humans);
- quantities in finance (price indices),
- quantities in social studies (IQ test results).

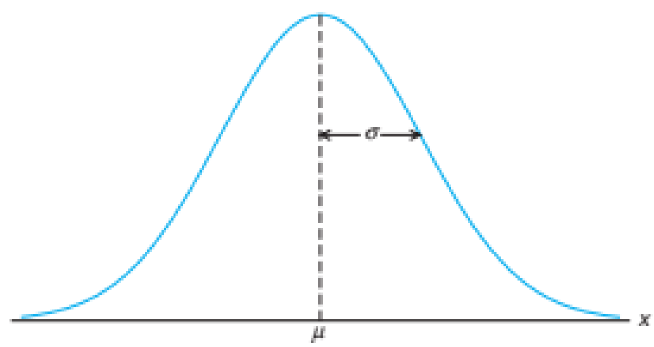


Figure 3.4: The normal curve.

Definition 5. The density of the normal random variable X, with mean μ and variance σ^2, is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x \in \mathbb{R}$$

noted by

$$X \sim N(\mu, \sigma)$$

where $\pi = 3.14159..$, $e = 2.71828$.

Once μ and σ are specified, the normal curve is completely determined. For example, if $\mu = 50$ and $\sigma = 5$, then the ordinates $X \sim N(50, 5)$ can be computed for various values of x and the curve drawn. In Figure **3.5**, we have sketched two normal curves having the same standard deviation but different means. The two curves are identical in form but are centered at different positions along the horizontal axis. In Figure **3.6**, we have sketched two normal curves with the same mean but different standard deviations. This time we see that the two curves are centered at exactly the same position on the horizontal axis.

Figure **3.7** shows two normal curves having different means and different standard deviations. Clearly, they are centered at different positions on the horizontal axis and their shapes reflect the two different values of σ.

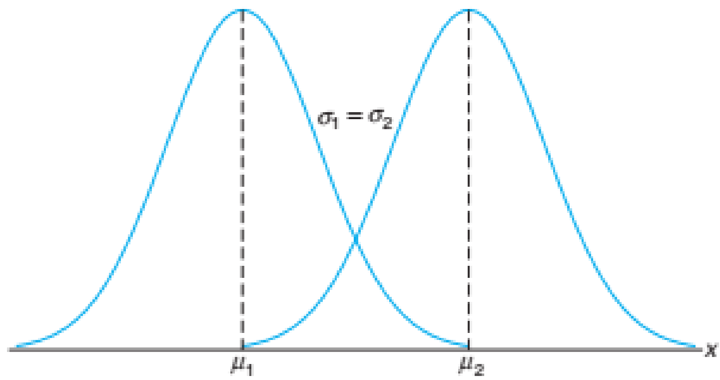


Figure 3.5: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 = \sigma_2$.

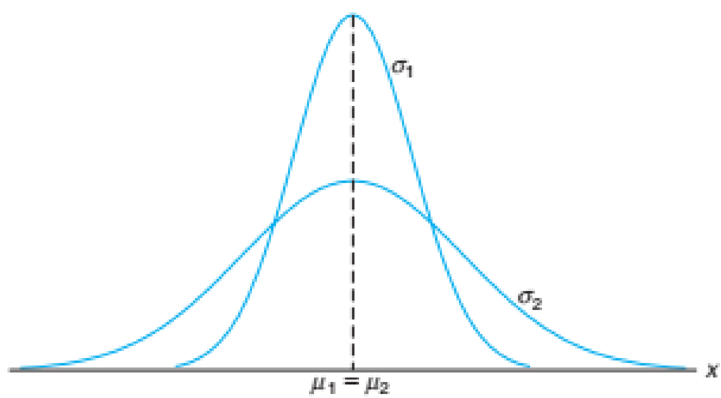


Figure 3.6: Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$

Based on inspection of Figures **3.4** through **3.7** and examination of the first and second derivatives of $N(\mu, \sigma)$, we list the following **properties** of the normal curve:

1. The mode, which is the point on the horizontal axis where the curve is a maximum, occurs at $x = \mu$.

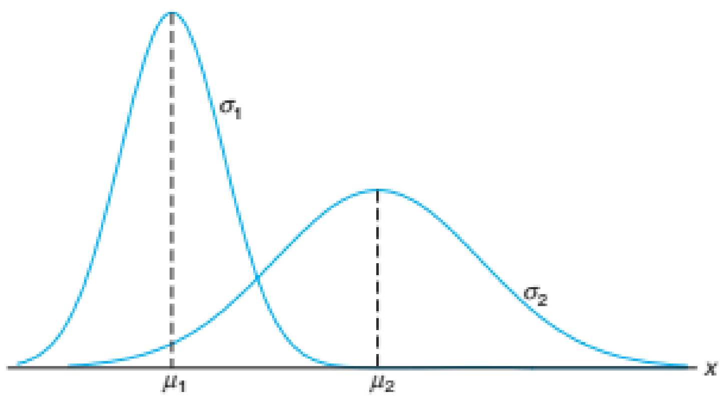


Figure 3.7: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$

2. The curve is symmetric about a vertical axis through the mean μ.

3. The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.

4. The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.

5. The total area under the curve and above the horizontal axis is equal to 1.

Theorem 1. The mean and variance of $N(\mu, \sigma)$ are μ and σ^2, respectively. Hence, the standard deviation is σ.

Standard normal distribution

Definition 5. The distribution of a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a standard normal distribution i.e

$$X \sim N(0, 1)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}}$$

The original and transformed distributions are illustrated in Figure **3.8**. Since all the values of X falling between x_1 and x_2 have corresponding z values between z_1 and z_2, the area under the $X-$curve between the ordinates $x = x_1$ and $x = x_2$ in Figure **3.8** equals the area under the $Z-$curve between the transformed ordinates $z = z_1$ and $z = z_2$.

We have now reduced the required number of **tables of normal-curve (see Table 2 below)** areas to one, that of the standard normal distribution. Table **2**. indicates the area under the standard normal curve corresponding to $P(Z < z)$ for values of z ranging from 0 to 3.49.

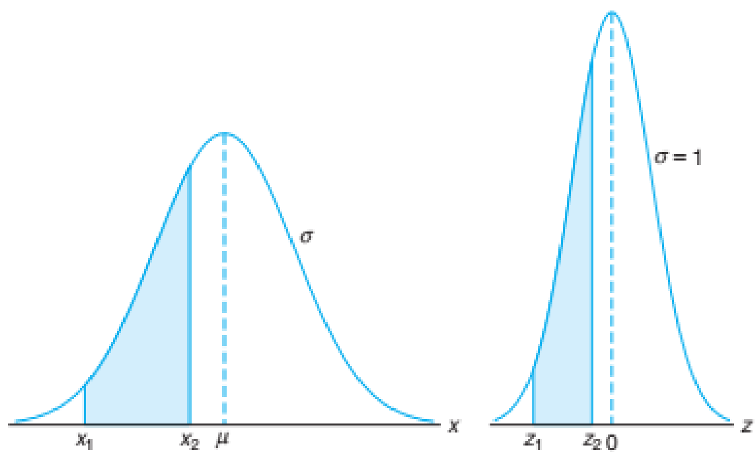


Figure 3.8: The original and transformed normal distributions.

Exemple 6. Given a standard normal distribution $X \sim N(0,1)$, Find by using table 2 the probability that X assumes a value

1) greater than 1.53

2) less than -2.45

3) between 0.86 and 1.2

4) Calculate $P(|X| < 1.25), P(|X| > 2.75)$

Solution: 1) $P(X \geq 1.53) = 1 - P(X \leq 1.53) = 1 - 0.9370 = 0.063$

2) $P(X \leq -2.45) = 1 - P(X \leq 2.45) = 1 - 0.9929 = 0.0071$

3) $P(0.86 \leq X \leq 1.2) = P(X \leq 1.2) - P(X \leq 0.86) = 0.8849 - 0.8051 = 0.0798$

4) We have

$$\begin{aligned}P(|X| < 1.25) &= P(-1.25 < X < 1.25) \\ &= P(X \leq 1.25) - P(X \leq -1.25) \\ &= 2P(X \leq 1.25) - 1 \\ &= 2 \times 0.8944 - 1 \\ &= 0.788\,8\end{aligned}$$

and

$$\begin{aligned}P(|X| > 2.75) &= P(X > 2.75 \text{ or } X < -2.75) \\ &= P((X > 2.75) \cup (X < -2.75)) \\ &= P(X > 2.75) + P(X < -2.75) \\ &= 2 - 2 \times P(X < 2.75) \\ &= 2 - 2 \times 0.9599 \\ &= 0.080\,2\end{aligned}$$

Transformation between standard normal distribution and normal distribution

We are able to transform all the observations of any normal random variable Y $(Y \sim N(\mu, \sigma))$ into a new set of observations of a normal random variable X with mean $\mu = 0$ and variance $\sigma^2 = 1$ $(X \sim N(0,1))$. This can be done by means of the transformation

$$X = \frac{Y - \mu}{\sigma}$$

we have

$$P(Y \leq A) = P\left(\frac{Y-\mu}{\sigma} \leq \frac{A-\mu}{\sigma}\right) = P\left(X \leq \frac{A-\mu}{\sigma}\right)$$

we can be find $P\left(X \leq \frac{A-\mu}{\sigma}\right)$ by using tables below **2** and **3**.

Exemple 7. Let Y has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that Y assumes a value greater than 362.

Solution: The normal probability distribution with the desired area shaded is shown in Figure **3.9**. To find $P(Y > 362)$, we need to evaluate the area under the normal curve to the right of $y = 362$. Thus

$$\begin{aligned}
P(Y > 362) &= P\left(\frac{Y-300}{50} > \frac{362-300}{50}\right) \\
&= P(X > 1.24) \\
&= 1 - P(X \leq 1.24) \\
&= 1 - 0.8925 \\
&= 0.1075
\end{aligned}$$

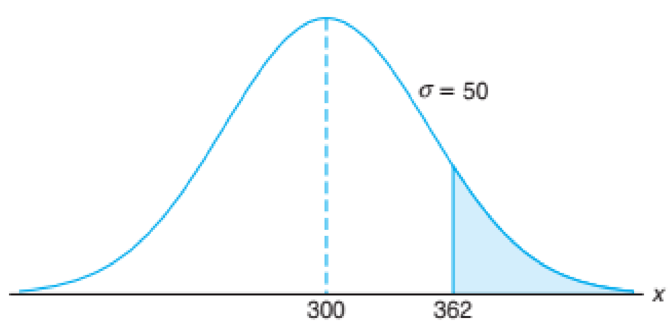


Figure 3.9: Area for Example 5.

Theorem 7. If $X_1, X_2, ..., X_n$ are independent random variables having normal distributions with means $\mu_1, \mu_2, ..., \mu_n$ and variances $\sigma_1, \sigma_2, ..., \sigma_n$, respectively, then the random variable

$$Y = a_1X_1 + a_2X_2 + ... + a_nX_n$$

has a normal distribution with mean

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n$$

and variance

$$\sigma_Y^2 = a_1^2\sigma_1 + a_2^2\sigma_2 + ... + a_n^2\sigma_n$$

It is now evident that the Poisson distribution and the normal distribution possess a reproductive property in that the sum of independent random variables having either of these distributions is a random variable that also has the same type of distribution.

3.2.4 Lognormal distribution

The lognormal distribution is used for a wide variety of applications. The distribution applies in cases where a natural log transformation results in a normal distribution.

The continuous random variable X has a lognormal distribution if the random variable $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ. The resulting density function of X is

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

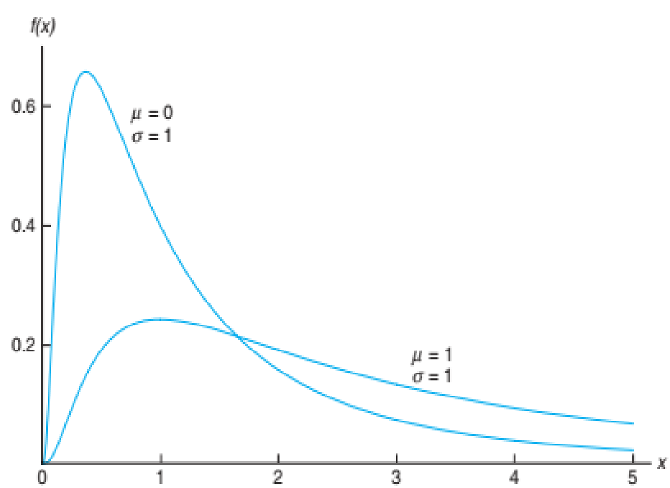


Figure 3.10: Lognormal distributions.

The graphs of the lognormal distributions are illustrated in Figure **3.10**.

Theorem 6. The mean and variance of the lognormal distribution are

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}, Var(X) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

The cumulative distribution function is quite simple due to its relationship to the normal distribution. The use of the distribution function is illustrated by the following example.

Example 8. Concentrations of pollutants produced by chemical plants historically are known to exhibit behavior that resembles a lognormal distribution. This is important when one considers issues regarding compliance with government regulations. Suppose it is assumed that the concentration of a certain pollutant, in parts per million, has a lognormal distribution with parameters $\mu = 3.2$ and $\sigma = 1$. What is the
probability that the concentration exceeds 8 parts per million?

Solution: Let the random variable X be pollutant concentration. Then

$$P(X > 8) = 1 - P(X \leq 8)$$

Since $\ln(X)$ has a normal distribution with mean $\mu = 3.2$ and standard deviation $\sigma = 1$,

.

$$P(X \leq 8) = \Phi\left(\frac{\ln(8) - 3.2}{1}\right) = \Phi(-1.12) = 0.1314$$

Here, we use Φ to denote the cumulative distribution function of the standard normal distribution. As a result, the probability that the pollutant concentration exceeds 8 parts per million is 0.1314.

Example 9. The life, in thousands of miles, of a certain type of electronic control for locomotives has an approximately lognormal distribution with $\mu = 5.149$ and $\sigma = 0.737$. Find the $5th$ percentile of the life of such an electronic control.

Solution: From Table **3**, we know that $P(Z < -1.645) = 0.05$. Denote by X the life of such an electronic control. Since $\ln(X)$ has a normal distribution with mean $\mu = 5.149$ and $\sigma = 0.737$, the $5th$ percentile of X can be calculated as
$\ln(x) = 5.149 + (0.737)(-1.645) = 3.937.$

Hence, $x = 51.265$. This means that only 5% of the controls will have lifetimes less than $51,265$ miles.

3.3 Exercises with solutions

Exercise 1. Let the function f defined by

$$f(x) = \begin{cases} \frac{1}{8}x, & x \in [0,4] \\ 0 & \text{elsewhere} \end{cases}$$

1) Show that the function f is the probability density on $[0,4]$.

2) Let X is a random variable of probability density f, calculate $P(1 \leq X \leq 3), P(X \geq 2)$.

Solution:1) f is continuous on $[0,4]$ and $f(x) \geq 0$ and

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_0^4 \frac{1}{8} x dx = \left[\frac{1}{8}\frac{1}{2}x^2\right]_0^4 = 1$$

show that f is a probability density.

2) $P(1 \leq X \leq 3) = \int_1^3 f(x)\, dx = \int_1^3 \frac{1}{8} x dx = \left[\frac{1}{8}\frac{1}{2}x^2\right]_1^3 = 0.5$

$P(X \geq 2) = \int_2^4 f(x)\, dx = \int_2^4 \frac{1}{8} x dx = \left[\frac{1}{8}\frac{1}{2}x^2\right]_2^4 = 0.75$

Exercise 2. Since it is more economical to limit long-distance telephone calls to three minutes or less, the probability density function (PDF) of X – the duration in minutes of long-distance calls – may be of the form

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\frac{x}{3}}, & \text{for } 0 \leq x < 3 \\ 1 - \dfrac{e^{-\frac{x}{3}}}{2}, & \text{for } x \geq 3 \end{cases}$$

Determine by two methods the probability that X is

(a) more than two minutes and,

(b) between two and six minutes.

Solution: The Probability distribution function, $F_X(x)$, of X is plotted in Figure **3.11**, showing that X has a mixedtype distribution. The desired probabilities can be found from the probability density function (PDF) as before.

(a) The probability that X is more than two minutes

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - F_X(2) \\ &= 1 - \left(1 - e^{-\frac{2}{3}}\right) \\ &= e^{-\frac{2}{3}} \end{aligned}$$

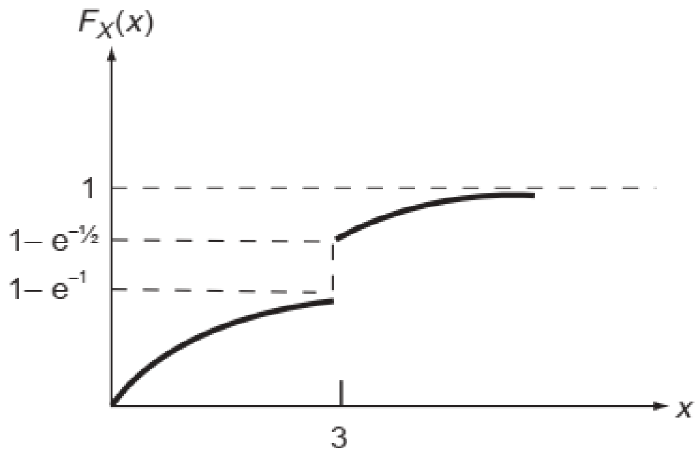


Figure 3.11: Probability distribution function, $F_X(x)$, of X, as described in Exercise 1.

(b) The probability that X is between two and six minutes

$$\begin{aligned} P(2 \leq X \leq 6) &= F_X(6) - F_X(2) \\ &= 1 - \frac{e^{-\frac{6}{3}}}{2} - \left(1 - e^{-\frac{2}{3}}\right) \\ &= e^{-\frac{2}{3}} - \frac{e^{-2}}{2} \end{aligned}$$

Figure **3.12** shows $p_X(x)$ for the discrete portion and $f_X(x)$ for the continuous portion of X. They are given by

$$P_X(x) = \begin{cases} \frac{1}{2e} & \text{at } x = 3 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$f_X(x) = F'_X(x) \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{3}e^{-\frac{x}{3}}, & \text{for } 0 \le x < 3 \\ \frac{1}{6}e^{-\frac{x}{3}}, & \text{for } x \ge 3 \end{cases}$$

Note again that the area under $f_X(x)$ is no longer one but is

$$1 - P_X(3) = 1 - \frac{1}{2e}$$

To obtain $P(X > 2)$ and $P(2 < X \le 6)$, both the discrete and continuous portions come into play, and we have, for part **(a)**,

$$\begin{aligned} P(X > 2) &= \int_2^\infty f_X(x)\,dx + P_X(3) \\ &= \int_2^3 \frac{1}{3}e^{-\frac{x}{3}}dx + \int_3^\infty \frac{1}{6}e^{-\frac{x}{3}}dx + \frac{1}{2e} \\ &= e^{-\frac{2}{3}} \end{aligned}$$

and, for part **(b)**,

$$\begin{aligned} P(2 < X \le 6) &= \int_2^6 f_X(x)\,dx + P_X(3) \\ &\int_2^3 \frac{1}{3}e^{-\frac{x}{3}}dx + \int_3^6 \frac{1}{6}e^{-\frac{x}{3}}dx + \frac{1}{2e} \\ &= e^{-\frac{2}{3}} - \frac{e^{-2}}{2} \end{aligned}$$

These results are, of course, the same as those obtained earlier using the PDF.

Exercise 3. Given a standard normal distribution $X \sim N(0,1)$, find the value of k such that

1) $P(X > k) = 0.3015$

2) $P(k < X < -0.18) = 0.4197$.

Solution: 1) if $P(X > k) = 1 - P(X \le k) = 0.3015$ then $P(X \le k) = 1 - 0.3015 = 0.698\,5$

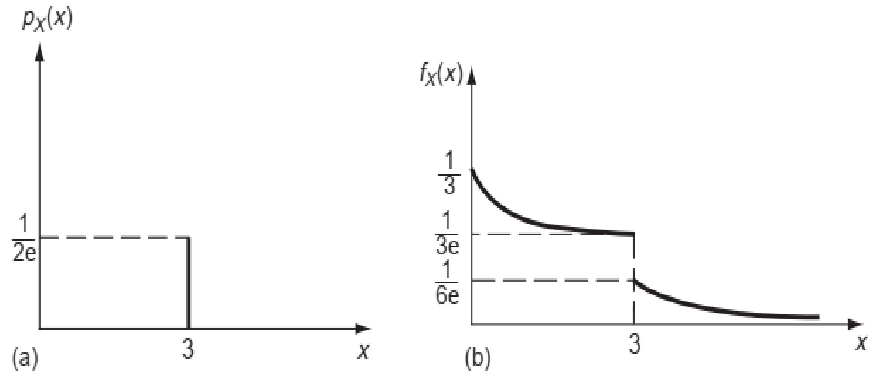


Figure 3.12: **(a)** Partial probability mass function, $p_X(x)$, and **(b)** partial probability density function, $f_X(x)$, of X, as described in Exercise 1.

yields (see table 2)

$$P(X \leq k) = 0.698\,5$$

$$k = 0.52$$

2) We have $P(k < X < -0.18) = P(X < -0.18) - P(X < k) = 0.4197$ implies that $P(X < k) = 1 - P(X < 0.18) - 0.4197$.

we get $P(X < k) = 1 - 0.5714 - 0.4197 = 0.008\,9$ i.e $k = -2.37$ (see table 3).

Exercise 4. Let a random variable Y having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the following probabilities

1) $P(Y < 55)$

2) $P(45 \leq Y \leq 62)$.

2) $P(Y \geq 30)$

Solution: 1) We have

$$P(Y < 55) = P\left(\frac{Y - 50}{10} < \frac{55 - 50}{10}\right)$$

$$= P(X < 0.5)$$

$$= 0.6915$$

2) Thus

$$\begin{aligned}
P(45 \leq Y \leq 62) &= P\left(\frac{45-50}{10} \leq \frac{Y-50}{10} \leq \frac{62-50}{10}\right) \\
&= P(-0.5 \leq X \leq 1.2) \\
&= P(X \leq 1.2) - P(X \leq -0.5) \\
&= P(X \leq 1.2) - 1 + P(X \leq 0.5) \\
&= 0.8849 - 1 + 0.6915 \\
&= 0.576\,4
\end{aligned}$$

3) We have

$$\begin{aligned}
P(Y > 30) &= P\left(\frac{Y-50}{10} > \frac{30-50}{10}\right) \\
&= P(X > -2) \\
&= 1 - P(X < -2) \\
&= P(X < 2) \\
&= 09772
\end{aligned}$$

Exercise 5. In a normal distribution 31% of the items are under 45 and 8% of the items are over 64. Find mean and variance of the distribution.

Solution: Given $P(X < 45) = 31\% = 0.31$

And $P(X > 64) = 8\% = 0.08$

Let Mean and variances of the normal distributions are μ, σ^2.

Standard normal variate for X is

$$z = \frac{x-\mu}{\sigma}$$

Standard normal variate for $X_1 = 45$ is

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{45-\mu}{\sigma} \longrightarrow \mu + \sigma z_1 = 45 \ldots\ldots (1)$$

Standard normal variate for $X_2 = 64$ is

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{64 - \mu}{\sigma} \longrightarrow \mu + \sigma z_2 = 64 \ldots\ldots (2)$$

From normal tables, we have $P\left(-z_1 \leq z \leq 0\right) = 0.19$

yields $z_1 = -0.5$.

and $\mathrm{P}(0 \leq z \leq z_2) = 0.42 \longrightarrow z_2 = 1.41$

substituting the values of z_1, z_2 in (1) and (2), we get $\mu = 50, \sigma = 98$.

Exercise 6. Let X is a random variable wich it follow the probability density function

$$f\left(x\right) = Ke^{-|x|}, -\infty < x < +\infty$$

1) Find the value of K,

2) Find mean and variance of X

Solution: 1) Since total area under the probability curve is 1 i. e $\int_{-\infty}^{+\infty} f\left(x\right) dx = 1$

$$\int_{-\infty}^{+\infty} Ke^{-|x|} dx = 1$$
$$2\int_{0}^{+\infty} Ke^{-|x|} dx = 1$$
$$2\left[-Ke^{-|x|}\right]_{0}^{+\infty} = 1$$

we obtain $K = \frac{1}{2}$.

2) Mean: $\mu = E\left(X\right) = \int_{-\infty}^{+\infty} xf\left(x\right) dx = \int_{-\infty}^{+\infty} \frac{x}{2} e^{-|x|} dx = 0$ since $xe^{-|x|}$ is the odd function.

Variance: by using the integration by parts, we get

$$\begin{aligned} Var\left(X\right) &= \int_{-\infty}^{+\infty} x^2 f\left(x\right) dx - \mu^2 \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} x^2 e^{-|x|} dx \\ &= \frac{1}{2}\int_{0}^{+\infty} 2x^2 e^{-x} dx \\ &= \left[-x^2 e^{-x} - 2xe^{-x} - 2e^{-x}\right]_0^{+\infty} \\ &= 2 \end{aligned}$$

Exercise 7. The diameter of ban electric cable assumed to be a continuous random variable X with probability density function

$$f\left(x\right) = kx\left(1-x\right), 0 \leq x \leq 1$$

1) Find k

2) Determine b such that $P(x < b) = P(x > b)$.

Solution: Given probability density function of a random variable X is

$$f\left(x\right) = kx\left(1-x\right), 0 \leq x \leq 1$$

1) Since total probability of the distribution is unity i.e, $\int_{-\infty}^{+\infty} f\left(x\right) dx = 1$

$$\begin{aligned} \int_0^1 f\left(x\right) dx &= 1 \\ \int_0^1 kx\left(1-x\right) dx &= 1 \\ \left[k\frac{x^2}{2} - k\frac{x^3}{3}\right]_0^1 &= 1 \end{aligned}$$

we get $k = \frac{1}{6}$

2) Given that $P(x < b) = P(x > b)$

$$\begin{aligned}
\int_0^b f(x)\,dx &= \int_b^1 f(x)\,dx \\
\int_0^b \frac{1}{6}x(1-x)\,dx &= \int_b^1 \frac{1}{6}x(1-x)\,dx \\
\left[\frac{x^2}{2}-\frac{x^3}{3}\right]_0^b &= \left[\frac{x^2}{2}-\frac{x^3}{3}\right]_b^1 \\
\frac{b^2}{2}-\frac{b^3}{3} &= \left(\frac{1}{2}-\frac{1}{3}\right)-\frac{b^2}{2}+\frac{b^3}{3}
\end{aligned}$$

yields the equation of three degree $6b^2 - 2b^3 - 1 = 0$

$6b^2 - 2b^3 - 1 = 0$, Solution is: $\{[b = 2.9422], [b = -0.38437], [b = 0.44213]\}$

Solving above equation, we get $b = 0.44$ (by neglecting other roots which do not belong to $(0, 1)$).

Exercise 8. Suppose that the error in the reaction temperature, in ■C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

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

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

2) Find $P(0 < X \le 1)$.

3) Compute $E(X)$ and $Var(X)$.

4) Find $F(x)$ is the probability distribution, and use it to evaluate $P(0 < X \le 1)$.

5) Find the expected value of $g(X) = 4X + 3$ i.e $E(g(X))$

Solution: 1) We have

a) $f(x) \ge 0$ and f is continuous on $[-1, 2]$

b) $\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-1}^{2} \frac{x^2}{3}dx = \left[\frac{x^3}{9}\right]_{-1}^{2} = 1$

2) $P(0 < X \le 1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{x^2}{3}dx = \left[\frac{x^3}{9}\right]_0^1 = \frac{1}{9}$.

3) $E(X) = \mu = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_{-1}^{2} \frac{x^3}{3}dx = \left[\frac{x^4}{12}\right]_{-1}^{2} = \frac{15}{12}$

$Var(X) = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \mu^2 = \int_{-1}^{2} \frac{x^4}{3}dx - \left(\frac{15}{12}\right)^2 = \left[\frac{x^5}{15}\right]_{-1}^{2} - \left(\frac{15}{12}\right)^2 = \frac{33}{15} - \left(\frac{15}{12}\right)^2 = 0.6375$

4) The probability distribution is

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

Also we obtain $P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$.

5) $E(g(X)) = E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3} dx = \frac{1}{3}\int_{-1}^{2} (4x^3 + 3x^2)\,dx = 8$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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

Table 2. The areas under the standard normal distribution curve.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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. The areas under the standard normal distribution curve.

Bibliography

[1] A. B. Clarke and Ralph L. Disney , Probability and Random Processes: A First Course with Applications, 2nd Edition (Wiley Series in Probability and Statistics), New York , 1991.

[2] F.M. Dekking and C. Kraaikamp, A Modern Introduction to Probability and Statistics: Understanding Why and How, Spinger, 2005.

[3] E. Parzen, Modern Probability Theory and Its Applications, John Wiley & sons Inc., New York, 1960.

[4] E.S. Pearson and H.O. Hartley, Biometrika Tables for Statisticians, Volume 1. Cambridge University Press, Cambridge, 1954.

[5] Sheldon M. Ross, Introduction to probability and statistics for engineers and scientists,Fifth Edition, Elsevier, 2014.

[6] Vijay K. Rohatgi, A. K. Md. Ehsanes Saleh, An introduction to probability and statistics, John Wiley & Sons, Inc, 2001.

[7] T.T. Soong, Fundamentals of probability and statistics for engineers, John Wiley & Sons, Inc, 2004.

[8] T. Veerarajan, Probability, statistics and random processes, Third edition, Tata McGraw-Hill, 2008.

[9] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Probability & Statistics for Engineers & Scientists, ninth edition, Walpole, Ronald E, 2012.