MTH113TC - Introduction to Probability and Statistics

Introduction to Probability Theory

Department of Mathematical Sciences
Xi'an Jiaotong Liverpool University

Term: 2020

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Preface

This text is intended as the lecture notes for the part of Probability Theory of module MTH113TC named Introduction to Probability and Statistics (Taicang Campus), when I taught this course for 2nd Year undergraduates at Xi'an Jiaotong-Liverpool University (XJTLU). It is intended as an elementary introduction to the theory of probability for students in Data Science & Big Data Technology, Financial Mathematics, Applied Mathematics, Actuarial Science and etc.

Some part of this note is referred to the books in the bibliography, which I found very well match the syllabus of the probability part of this module during my teaching, especially the book [1] . This version is a modified version based on the original one that I wrote in Jul. 2016, where I found some typos or mistakes during my teaching until Oct. 2018. In the teaching process, the students and the teaching group members also helped a lot to correct some errors in this note.

This lecture note can be used for the purpose of internal study for students on the campus of XJTLU. It is also used as a standard and necessary outline of topics to be taught for the teaching group members of this module. Without author's permission, it is not permitted to publish/post in any external platform outside of XJTLU. If this happens, the person is responsible for any consequences caused.

Dr. Ye Yinna

27th Aug. 2020 at Suzhou, China

Contents

MTH113 - Introduction to probability and statistics

Recommended textbook:

- A first course in probability, Sheldon M. Ross, *Pearson*
- Probability and random processes, 3rd edition, Geoffery Grimmett, David Sttirzaker, *Oxford University Press* (**essentially for Probability part mentioned in this course**)
- Introduction to probability and statistics for engineers and scientists, Sheldon M. Ross, *Elsevier Academic Press*
- Statistics for business and economics, W. L. Newbold, Carlson B. Thorne, *Pearson* (**essentially for Statistical part mentioned in this course**)

Aims

- To provide a rigorous introduction to probability and mathematical statistics particularly for math majored students;
- To discuss the potential scope of the applications and illustrate typical ways of analysis;
- To provide an appropriate technical background for related higher level MTH modules.

Learning outcomes

Students completing the module successfully should be able to:

- describe statistical data;
- apply basic probability theory to solve related problems;
- provide good knowledge on typical distributions such as Bernoulli, Binomial, Geometric, Uniform, Poisson, Exponential and Normal distributions and their applications;
- understand the idea and perform simple goodness-of-fit tests.

Assessment

Coursework	15%
Final exam	85%
Total	100%

Schedule (estimated)

Weeks 1 to 2	Introduction to Probability
Week 4	Coursework due time (details to be advised)
Weeks 3 to 5	Introduction to Statistics
Week 6	Final exam revision (if time permitted)

Chapter 1

Elementary probability

1.1 Sample space and events

1.1.1 Introduction to sample space

Definition 1.1.1 (Experiment and outcome)**.** A process of observation and trial is called an **experiment**, and the result of this experiment is called **outcome**.

Example 1.1.1 (Toss of 2 dice)**.** A simultaneous throw of 2 dice is an experiment. If we sum up the numbers of dots on the face of those two dice, we may have the possible outcomes: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Definition 1.1.2 (Sample space and event)**.** The set of all possible outcomes of an experiment is called a **sample space**, which is usually denoted by Ω. We will assume from now on that $\Omega \neq \emptyset$. Any subset E of the sample space is called an **event**.

Example 1.1.2. (1) The sex of a newborn child:

$$\Omega = \{g, b\},$$

where the outcome g means that the child is a girl and b that it is a boy.
If $E = \{g\}$, then E is the event that the child is a girl.

(2) Flipping two coins:

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\};$$

where the outcome (H,H) means both coins are heads, the outcome (H,T) means the first coin is heads and the second tails, the outcome (T,H) means the first coin is tails and the second heads and the outcome (T,T) means both coins are tails.
If $E = \{(H,H),(H,T)\}$, then E is the event that a head appears on the first coin.

Figure 1.1: Left side - Head and Tail of a UK coin. Right side - Flipping a coin

(3) Tossing 2 dice:

$$\Omega = \{(i,j) : i,j = 1,2,3,4,5,6\} = \{1,2,\cdots,6\} \times \{1,2,\cdots,6\},$$

it consists of $6 \times 6 = 36$ outcomes, where (i,j) is said to occur if i appears on the leftmost die and j on the other die.
If $E = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, then E is the event that the sum of the dice equals 7.

(4) Weight (in kg) of new-born babies: the sample space is the interval

$$\Omega = (0,10].$$

Ω is a uncountable set.
If $E = \{x : 0 < x \leq 4\}$, then E is the event that the new-born baby is weighted no heavier than 4 kg.

1.1.2 Set operations

Definition 1.1.3 (Set operations)**.** Let Ω be a given sample space and $E, F \subset \Omega$ are two events. Define

$$E \cup F := \{x \in \Omega : x \in E \text{ or } x \in F\}, \tag{1.1}$$

$$E \cap F := \{x \in \Omega : x \in E \text{ and } x \in F\}, \tag{1.2}$$

$$E^c := \{x \in \Omega : x \notin E\}; \tag{1.3}$$

where the operation $\cup$ in (1.1) is called the **union** of the events E and F, the operation $\cap$ in (1.2) is called the **intersection** of the events E and F and the operation ^{c} in (1.3) is called the **complement** of the event E.

Remark 1.1.1. (1) $E \cup F$ stands for the event that either event E or event F occurs.

(2) $E \cap F$ (or EF) is the event that both E and F occur.

(3) E^c is the **negation** of E ('non E'), the event that E does not occur.

Example 1.1.3.

- In Example 1.1.2 (1), defined the events $E := \{g\}$ and $F := \{b\}$, then
$$E \cup F = \{g, b\} = \Omega.$$

- In Example 1.1.2 (2), if $E := \{(H, H), (H, T)\}$ and $F := \{(T, H)\}$, then
$$E \cup F = \{(H, H), (H, T), (T, H)\}$$
is the event that a head appear on either coin.
If $E := \{(H, H), (H, T), (T, H)\}$ is the event that at least 1 head occurs and $F := \{(H, T), (T, H), (T, T)\}$ is the event at least 1 tail occurs, then
$$E \cap F = \{(H, T), (T, H)\}$$
is the event that exactly 1 head and 1 tail occur.

- In Example 1.1.2 (3), if $E := \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ is the event that the sum of the dice is 7 and $F := \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ is the event that the sum is 6, then the event $E \cap F$ does not contain any outcomes and hence could not occur, i.e.,
$$E \cap F = \emptyset.$$
And E^c is the event that the sum of the dice does not equal 7, i.e.,
$$E^c = \{\text{The sum of the dice does not equal } 7\}$$

Definition 1.1.4. (1) **Null event**, denoted by $\emptyset$, is the event consisting of no outcomes.

(2) If $E \cap F = \emptyset$, then E and F are said to be **mutually exclusive** or **disjoint events**.

(3) More generally, for a given finite collection of events $\{C_1, C_2, \cdots, C_n\}$, if for any $i \neq j$ and $i, j = 1, 2, \ldots, n$, $C_i \cap C_j = \emptyset$, then the events are said to be **pairwise mutually exclusive**.

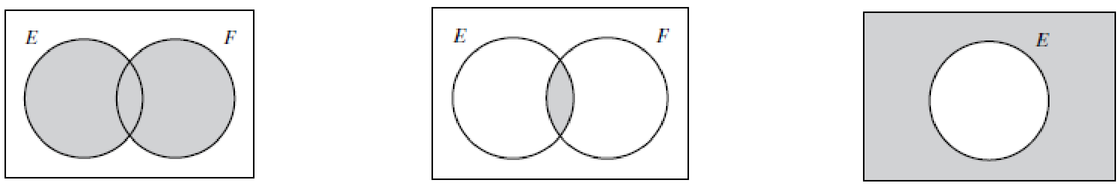
Remark 1.1.2. (1) Note that because the experiment must result in some outcome (i.e. $\Omega \neq \emptyset$), it follows that
$$\Omega^c = \emptyset.$$

(2) For any event $A \subseteq \Omega$,
$$\begin{aligned} A \cap \Omega = A, &\quad A \cup \Omega = \Omega \\ A \cap \emptyset = \emptyset, &\quad A \cup \emptyset = A. \end{aligned}$$

Definition 1.1.5 (Subsets and equal events)**.** For any two events E and F, if all of the outcomes in E are also in F, then we say that E is **contained** in F, or E is a **subset** of F, and write $E \subseteq F$. Thus, if $E \subseteq F$, then the occurrence of E implies the occurrence of F. If $E \subseteq F$ and $F \subseteq E$, then we say that E and F are **equal** and write $E = F$.

Definition 1.1.6 (Unions and intersections of more than 2 events)**.** If E_1, E_2, $\cdots$ are events, then the **union** of these events, denoted by $\bigcup_{n=1}^{\infty} E_n$, is defined to be that event which consists of all outcomes that are in E_n for at least one value of $n = 1, 2, \cdots$. Similarly, the **intersection** of the events E_n, denoted by $\bigcap_{n=1}^{\infty} E_n$, is defined to be the event consisting of those outcomes which are in all of the events E_n, $n = 1, 2, \cdots$.

The **Venn diagram** is a very useful graphical representation for illustrating logical relations among events. The sample space Ω is represented as consisting of all outcomes in a large rectangle, and the events E, F, G, $\cdots$ are represented as consisting of all the outcomes in given circles within the rectangle. Events of interest can then be indicated by shading appropriate regions of the diagram. For instance, in the three Venn diagrams shown in Figure 1.2, the shaded areas represent, respectively, the events $E \cup E$, $E \cap F$ and E^c. The Venn diagram in Figure 1.3 indicates that $E \subseteq F$.



(a) Shaded region: $E \cup F$

(b) Shaded region: EF

(c) Shaded region: E^c.

Figure 1.2: Venn diagrams

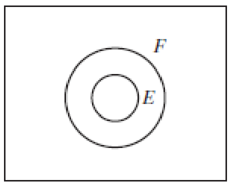


Figure 1.3: $E \subseteq F$

1.1.3 Properties of set operations

Theorem 1.1.3 (Properties of set operations)**.** Let E, F, G be any sets. Then we have the following properties of their unions and intersections:

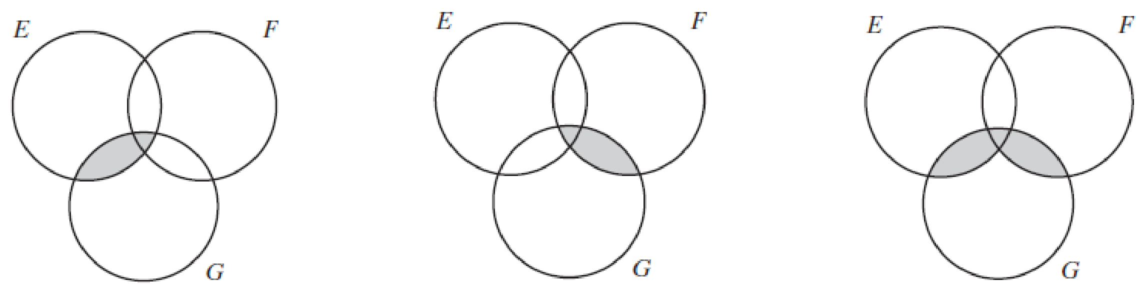
Commutative laws $E \cup F = F \cup E \qquad E \cap F = F \cap E$

Associative laws $(E \cup F) \cup G = E \cup (F \cup G) \qquad (E \cap F) \cap G = E \cap (F \cap G)$

Distributive laws $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

Idempotence $E \cup E = E$ $E \cap E = E$.

These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the right side, and vice versa. One way of showing this is by the sequence of diagrams. For instance, the distributive law may be verified by the sequence of diagrams in Figure 1.4. More generally, the properties of set



(a) Shaded region: EG. (b) Shaded region: FG. (c) Shaded region: $(E \cup F)G$.

Figure 1.4: $(E \cup F)G = EG \cup FG$

operations above are also true for finite number of sets (see Exercise 2.5).

Theorem 1.1.4 (DeMorgan's laws)**.** Let E_1, E_2, $\ldots$, E_n be a collection of n events in the sample space Ω.

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

Proof. We only prove the first equality here, the method for proving the second one is similar. Suppose first that x is a outcome of $\left(\bigcup_{i=1}^{n} E_i\right)^c$. Then x is not contained in $\bigcup_{i=1}^{n} E_i$, which means that x is not contained in any of the events E_i, $i = 1, 2, \cdots, n$, implying that x is contained in E_i^c for all $i = 1, 2, \cdots, n$ and thus is contained in $\bigcap_{i=1}^{n} E_i^c$. So we have found $\left(\bigcup_{i=1}^{n} E_i\right)^c \subseteq \bigcap_{i=1}^{n} E_i^c$.

Conversely, suppose that $x \in \bigcap_{i=1}^{n} E_i^c$. Then $x \in E_i^c$, for all $i = 1, 2, \cdots n$, which means that x is not contained in E_i for any $i = 1, 2, \cdots, n$, implying that x is not contained in $\bigcup_{i=1}^{n} E_i$, in term implying that x is contained in $\left(\bigcup_{i=1}^{n} E_i\right)^c$. So we have $\bigcap_{i=1}^{n} E_i^c \subseteq \left(\bigcup_{i=1}^{n} E_i\right)^c$.

In conclusion, $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$. $\square$

Exercise 1.1.1. Try to prove the second of DeMorgan's laws by taking complements and the first law.

1.2 Axioms of probability

Definition 1.2.1. Consider an experiment whose sample space is Ω. The probability $\mathbb{P}$ is a function which assigns a set in 2^Ω to a value in $[0,1]$. More precisely, for each event E ($\subseteq \Omega$), the **probability of the event** E, denoted by $\mathbb{P}(E)$, is defined and satisfies the following three axioms:

Axiom 1 $0 \le \mathbb{P}(E) \le 1$;

Axiom 2 $\mathbb{P}(\Omega) = 1$ and Ω is the unique set in 2^Ω such that this equality holds;

Axiom 3 For any sequence of pairwise mutually exclusive events E_1, E_2, ... (that is, events for which $E_i \cap E_j = \emptyset$ when $i \neq j$),

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

The Axioms 1-3 are called **Kolmogorov's axioms**.

Remark 1.2.1. In words, Axiom 1 states that the probability that the outcome of the experiment is an outcome in E is some number between 0 and 1. Axiom 2 states that, with probability 1, the outcome will be a point in the sample space Ω. Axiom 3 states that, for any sequence of pairwise mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

1.3 Fundamental formula for probability

Now, as an example, let's see one of the most classical probabilities defined by the following 'fundamental formula'.

Theorem 1.3.1. Let Ω be a finite sample space, the function $\mathbb{P}$ defined as follows is a probability: for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}. \tag{1.4}$$

Proof. In order to prove this theorem, we need to check the Kolmogorov's axioms hold for the given $\mathbb{P}$.

(1) Since $\emptyset \subseteq A \subseteq \Omega$, we have

$$0 = |\emptyset| \le |A| \le |\Omega|.$$

From this, we obtain

$$0 = \frac{|\emptyset|}{|\Omega|} \le \frac{|A|}{|\Omega|} = \mathbb{P}(A) \le \frac{|\Omega|}{|\Omega|} = 1.$$

The Axiom 1 is therefore verified.

(2) Let $A = \Omega$, then (1.4) gives

$$\mathbb{P}(\Omega) = \frac{|\Omega|}{|\Omega|} = 1.$$

And Ω is the only event with probability 1.

(3) Suppose that E_1, E_2, $\cdots$ is a sequence of pairwise mutually exclusive events in the sample space Ω. Since $E_i \cap E_j = \emptyset$, for any positive integers $i \neq j$,

$$\left| \bigcup_{i=1}^{\infty} E_i \right| = \sum_{i=1}^{\infty} |E_i|.$$

So (1.4) gives

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \frac{|\bigcup_{i=1}^{\infty} E_i|}{|\Omega|} = \frac{\sum_{i=1}^{\infty} |E_i|}{|\Omega|} = \sum_{i=1}^{\infty} \frac{|E_i|}{|\Omega|} = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

By now we have shown that $\mathbb{P}$ defined in (1.4) is a probability. □

Remark 1.3.2. The Fundamental formula (1.4) can be interpreted as the probability of an event which is given by the ratio of the number of outcomes favorable to the total number of possible outcomes. In practice, we shall use this formula when each of the outcome in an experiment is equally probable to occur.

Example 1.3.1. If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we could have

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 1/2.$$

Example 1.3.2. If a die is rolled and we suppose that all six sides are equally likely to appear, i.e.

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

From Axiom 3, it would thus follow that the probability of rolling an even number would equal

$$\mathbb{P}(\{2,4,6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Example 1.3.3. On a toss of two fair dice, we record the sum total of spots on the uppermost sides of the dice. What is the probability of having this number equal to 7?

Solution: In Example 1.1.2 (3), we listed all the possible outcomes,

$$\begin{aligned}\Omega =& \{(i,j) : i,j = 1,2,\cdots 6\} \\ =& \left\{\begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix}\right\}.\end{aligned}$$

This gives a total number of $|\Omega| = 6^2 = 36$ outcomes, each of which is equally likely to occur.

Denote by (ω_1, ω_2) an arbitrary outcome. The event, say A, we are interested is then represented by

$$A = \{(\omega_1, \omega_2) \in \Omega : \omega_1 + \omega_2 = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}.$$

Therefore, by Fundamental formula,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

1.4 Some useful results on probability

Theorem 1.4.1 (Rule of complement)**.** If $A \subseteq \Omega$ is an event in sample space Ω, then

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Proof. Note that $A^c \cup A = \Omega$ and $A^c \cap A = \emptyset$. By using Kolmogorov's axioms, we have

$$1 \overset{\text{Ax.2}}{=} \mathbb{P}(\Omega) = \mathbb{P}(A^c \cup A) \overset{\text{Ax.3}}{=} \mathbb{P}(A^c) + \mathbb{P}(A),$$

which yields the desired result. $\square$

Theorem 1.4.2 (Rule of total probability)**.** If $A, B \subseteq \Omega$ are events in the sample space Ω, then

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c). \tag{1.5}$$

Proof. By using distributive law we can rewrite A as

$$A = A \cap \Omega = A \cap (B \cup B^c) \overset{\text{distr.}}{=} (A \cap B) \cup (A \cap B^c). \tag{1.6}$$

This is a disjoint union (why? Exercise). Then by Kolmogorov's axiom 3, we have

$$\mathbb{P}(A) = \mathbb{P}((A \cap B) \cup (A \cap B^c)) \overset{\text{Ax.3}}{=} \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$

$\square$

Remark 1.4.3 (Generalization of the rule of total probability)**.** More generally, if a finite collection of pairwise mutually exclusive events $\{C_1, C_2, \cdots, C_n\}$ satisfy

$$C_1 \cup C_2 \cup \cdots \cup C_n = \Omega$$

(we will see this kind of events is called a **partition** of Ω in the next chapter), then for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \mathbb{P}(A \cap C_1) + \mathbb{P}(A \cap C_2) + \cdots + \mathbb{P}(A \cap C_n).$$

(As exercise, prove this generalization.)

Theorem 1.4.4 (Monotonicity)**.** If $A, B \subseteq \Omega$ are events in the sample space Ω and $A \subseteq B$, then

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

Proof. By exchanging A and B in (1.5), we have

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c).$$

Since $A \subseteq B$, $A \cap B = A$. So

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \cap A^c)}_{\geq 0 \text{ by Ax.1}},$$

that is

$$\mathbb{P}(B) - \mathbb{P}(A) = \mathbb{P}(B \cap A^c) \geq 0.$$

And hence we obtain $\mathbb{P}(B) \geq \mathbb{P}(A)$. $\square$

Theorem 1.4.5 (Rule of addition)**.** If $A, B \subseteq \Omega$ are events in the sample space Ω, then

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

Proof. Note that $A \cup B$ can be written as the union of two disjoint events A and $B \cap A^c$, that is

$$A \cup B = A \cup (B \cap A^c).$$

So by Kolmogrov's axiom 3, we have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \cap A^c)) \overset{\text{Ax.3}}{=} \mathbb{P}(A) + \mathbb{P}(B \cap A^c). \tag{1.7}$$

Using rule of total probability (1.5), we have

$$\mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Substituting $\mathbb{P}(B \cap A^c)$ into (1.7), we thus obtain the desired result. $\square$

Exercise 1.4.1. Try to use the Venn diagram to prove the Rule of addition.

EXERCISES

1.1. A box contains 3 marbles: 1 red, 1 green, and 1 blue. Consider an experiment that consists of taking 1 marble from the box and then replacing it in the box and repeat this action for 3 times. Describe the sample space. Repeat when the second marble is drawn without replacing the first marble and the third one is drawn without replacing the second.

1.2. Identify the sample spaces of the following experiments:

(a) If the outcome of an experiment consists in the determination the sex of a newborn child (B for boy and G for girl);

(b) If the outcome of an experiment is the order of finishing in a race among 7 horses having post positions 1, 2, 3, 4, 5, 6, 7;

(c) If the experiment consists of flipping two coins;

(d) If the experiment consists of measuring (in hours) the lifetime of a transistor;

(e) If the experiment consists of counting how many people will pop into one shop during some specific time interval;

(f) If the experiment consists of counting how many phone calls have to be made by the service center during one day.

1.3. Two dice are thrown. Let E be the event that the sum of the dice is odd, let F be the event that at least one of the dice lands on 1, and let G be the event that the sum is 5. Describe the events EF, $E \cup F$, FG, EF^c, and EFG.

1.4. A cafeteria offers a 3-course meal consisting of an entree, a starch, and a dessert. The possible choices are given in the following table:

Course	**Choices**
Entree	Chicken or roast beef
Starch	Pasta or rice or potatoes
Dessert	Ice cream or Jello or apple pie or a peach

A person is to choose one course from each category (i.e. Entree, Starch and Dessert).

(a) What is the expression for the sample space Ω, if we regard the choice of a 3-course meal as an experiment?

(b) How many outcomes are there in the sample space Ω?

(c) Let A be the event that ice cream is chosen. How many outcomes are there in A?

(d) Let B be the event that chicken is chosen. How many outcomes are there in B?

(e) List all the outcomes in the event AB.

(f) Let C be the event that rice is chosen. How many outcomes are there in C?

(g) List all the outcomes in the event ABC.

1.5. Let E, F, and G be three events. Find expressions for the following events by using the set operations such as $\cap$, $\cup$ and ^{c},

(a) only E occurs;

(b) both E and G, but not F, occur;

(c) at least one of the events occurs;

(d) at least two of the events occur;

(e) all three events occur;

(f) none of the events occurs;

(g) at most one of the events occurs;

(h) at most two of the events occur;

(i) exactly two of the events occur;

(j) at most three of the events occur.

1.6. Suppose E and F be two sets. The **set difference** is defined by

$$E \setminus F := E \cap F^c = \{x \in \Omega : x \in E \text{ and } x \notin F\}.$$

Let Ω be a given sample space, and $A, B, C \subseteq \Omega$ are three events. Use the definition of set difference and DeMorgan's law to show the following identities:

(a) $A \setminus B = B^c \setminus A^c$;

(b) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$;

(c) $(A \setminus B) \setminus C = A \setminus (B \cup C)$.

1.7. Suppose that an experiment is performed n times. For any event E of the sample space, let $n(E)$ denote the number of times that event E occurs and define $f(E) = n(E)/n$. Show that $f(\cdot)$ satisfies Axioms 1, 2, and 3.

1.8. There is a fake die such that the probability of the side i equals ki, where k is a constant. (Note that $\mathbb{P}(\Omega) = 1$, where Ω is the sample space).

(a) Determine the constant k.

Suppose one is tossing the fake die. Calculate the following probabilities of number that appears on the upper-side:

(b) an even number;

(c) an odd number;

(d) a prime number (1 is not a prime number);

(d) a prime and odd number.

1.9. (Let's make a deal) Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

(This is a problem that the inviters confronted in a television game show called "Let's make a deal" created and hosted by Monty Hall, American TV host and producer, in the early 1960s.)

1.10. Let A and B be two events such that $\mathbb{P}(A) = \dfrac{3}{8}$, $\mathbb{P}(B) = \dfrac{1}{2}$ and $\mathbb{P}(A \cap B) = \frac{1}{4}$. Calculate $\mathbb{P}(A \cup B)$, $\mathbb{P}(A^c)$, $\mathbb{P}(B^c)$, $\mathbb{P}(A^c B^c)$, $\mathbb{P}(A^c \cup B^c)$, $\mathbb{P}(AB^c)$ and $\mathbb{P}(BA^c)$.

Chapter 2

Conditional probability and independence

One of the most important concepts in the theory of probability is based on the question: *How do we modify the probability of an event in light of the fact that something new is known?* What is the chance that we will win the game now that we have taken the first point? What is the chance that I am a carrier of a genetic disease now that my first child does not have the genetic condition? What is the chance that a child smokes if the household has two parents who smoke? This question leads us to the concept of **conditional probability**.

2.1 Conditional probability

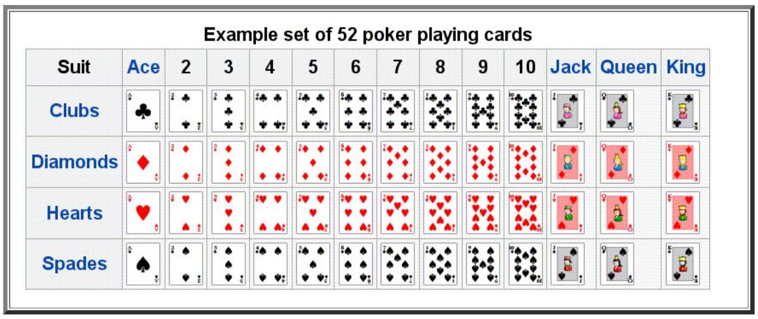


Figure 2.1: Deck of 52 cards

Example 2.1.1. Suppose you play bridge with three of your friends. After the cards are shuffled, the deck of 52 cards (see Fig 2.1) is dealt so that each player receives a hand of 13 cards.
Before you look at your deal, your judgement about the probability that your partner has the ace of hearts is $1/4$. Suppose you look at your deal and discover the ace of hearts among your cards. The probability that your partner has it is now 0. If you do not have it in your deal, then the probability that your partner has it goes up to $1/3$.
Your judgement of the probabilities differs because it is conditioned by the extra information whether you do or do not have the ace of hearts.

Example 2.1.2. Consider a degree programme in a mathematical department located in China with the following student numbers:

	1st year	2nd year	total
Overseas	10	18	28
China	15	20	35
total	25	38	63

By using the Fundamental formula of probability

$$\mathbb{P}(\text{overseas student}) = \frac{|\{\text{overseas students}\}|}{|\{\text{students}\}|} = \frac{28}{63} \approx 0.44.$$

$$\mathbb{P}(\text{Chinese student}) = \frac{|\{\text{Chinese students}\}|}{|\{\text{students}\}|} = \frac{35}{63} \approx 0.56.$$

Definition 2.1.1 (Conditional probability)**.** Suppose A and B are two events in the sample space Ω and $\mathbb{P}(B) \neq 0$. The probability given by

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.1}$$

is called the **conditional probability of event A given (or condition on) event B.**

Example 2.1.3 (Revisited Example 2.1.2)**.** If we select a second year student at random, what is the chance that the student is from China?

Solution:
The required probability gives:

$$\mathbb{P}(\text{China}|\text{ 2nd year student}) := \frac{|\{\text{2nd year Chinese student}\}|}{|\{\text{2nd year students}\}|} = \frac{20}{38} \approx 0.53.$$

Example 2.1.4 (Revisit Example 2.1.1)**.** Now we can reinterpret Example 2.1.1 by the definition of conditional probability,

$$\mathbb{P}(\text{your partner has ace of hearts}|\text{ before you look at your cards}) = \frac{1}{4};$$

$$\mathbb{P}(\text{your partner has ace of hearts}|\text{ after you notice you have it}) = 0;$$

$$\mathbb{P}(\text{your partner has ace of hearts}|\text{ after you notice you do not have it}) = \frac{1}{3}.$$

Theorem 2.1.1. For fixed event $B \subseteq \Omega$, under the conditions of Definition 2.1.1, we have $\widetilde{\mathbb{P}}(A) := \mathbb{P}(A|B)$ is a probability defined on the sample space $\widetilde{\Omega} = B$, i.e.,

(1) For any event E ($\subseteq \widetilde{\Omega}$), $0 \leq \widetilde{\mathbb{P}}(A) \leq 1$;

(2) $\widetilde{\mathbb{P}}(\widetilde{\Omega}) = 1$ and $\widetilde{\Omega}$ is the unique event which satisfies this identity;

(3) For any sequence of pairwise mutually exclusive events E_1, E_2, ... (that is, events for which $E_i \cap E_j = \emptyset$ when $i \neq j$) in $\widetilde{\Omega}$,

$$\widetilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \widetilde{\mathbb{P}}(E_i).$$

Proof. (1) Non-negativity of $\widetilde{\mathbb{P}}(E)$ follows both of the numerator and the denominator are non-negative. Since $E \cap B \subseteq B$, the monotonic property, implies $\mathbb{P}(E \cap B) \leq \mathbb{P}(B)$. Thus,

$$0 \leq \widetilde{\mathbb{P}}(E) := \mathbb{P}(E|B) = \frac{\mathbb{P}(E \cap B)}{\mathbb{P}(B)} \leq \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

(2) By the definition, we have

$$\widetilde{\mathbb{P}}(\widetilde{\Omega}) = \frac{\mathbb{P}(B \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1,$$

where $\widetilde{\Omega} = B$ is the only event which satisfies the identity above.

(3) Let E_1, E_2, ... be a sequence of pairwise mutually exclusive events, i.e., $E_i \cap E_j = \emptyset$ if $i \neq j$. Then the sequence of events $E_1 \cap B$, $E_2 \cap B$, ... is also pairwise mutually exclusive, i.e., $(E_i \cap B) \cap (E_j \cap B) = \emptyset$, if $i \neq j$. Then

$$\begin{aligned}
\widetilde{\mathbb{P}}\left(\bigcup_{i=1}^{\infty} E_i\right) &\overset{\text{def}}{=} \frac{\mathbb{P}\left[\left(\bigcup_{i=1}^{\infty} E_i\right) \cap B\right]}{\mathbb{P}(B)} \overset{\text{distr.}}{=} \frac{\mathbb{P}\left[\bigcup_{i=1}^{\infty}(E_i \cap B)\right]}{\mathbb{P}(B)} \\
&\overset{\text{Ax.3}}{=} \frac{\mathbb{P}(E_1 \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(E_2 \cap B)}{\mathbb{P}(B)} + \cdots \\
&= \sum_{i=1}^{\infty} \mathbb{P}(E_i|B) = \sum_{i=1}^{\infty} \widetilde{\mathbb{P}}(E_i)
\end{aligned}$$

□

2.2 The multiplication principle

Corollary 2.2.1 (Multiplication principle)**.** The defining formula (2.1) for conditional probability can be rewritten to obtain the **the multiplication principle**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B).$$

Remark 2.2.2. We can continue this process as a **chain rule**:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A|B \cap C)\,\mathbb{P}(B \cap C) = \mathbb{P}(A|B \cap C)\,\mathbb{P}(B|C)\,\mathbb{P}(C).$$

Example 2.2.1. Suppose that two balls are to be randomly drawn, one after another, from a container holding 4 **red** balls and 2 **green** balls. See Figure below for illustration. Under the scenario of sampling without replacement, calculate the probability of the events

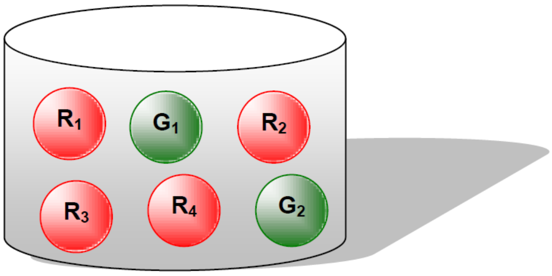


Figure 2.2: urn model

(1) $A = \{\text{first ball is } \textbf{red}\}$,

(2) $B = \{\text{second ball is } \textbf{red}\}$,

(3) $A \cap B = \{\text{first ball is } \textbf{red} \text{ AND second ball is } \textbf{red}\}$.

Solution:
This type of problem - known as an "urn model" - can be solved with the use of a **tree diagram**, where each branch of the "tree" represents a specific event, conditioned on a preceding event. The product of the probabilities of all such events along a particular sequence of branches is equal to the corresponding intersection probability, by the multiplication principle. In this example, we obtain the following values $\mathbb{P}(A)$ and $\mathbb{P}(A \cap B)$ as shown in Figure below.

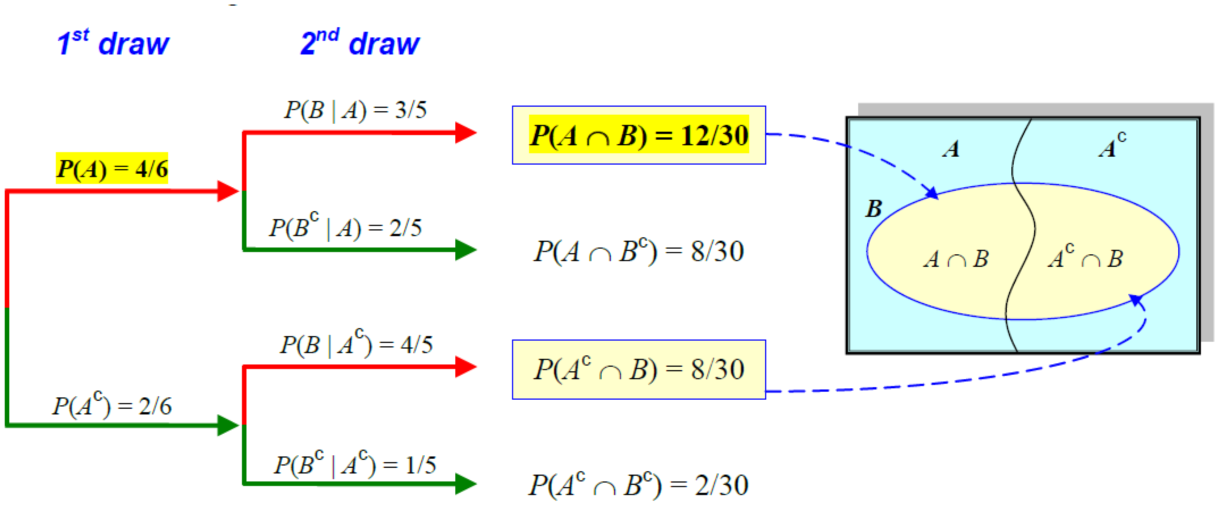


Figure 2.3: illustration of the solution

For the probability $\mathbb{P}(B)$, we can calculate it by the rule of total probability (see Section 1.4) and adding the two "boxed" values in the figure, i.e.

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = 12/30 + 8/30 = 20/30 \quad \text{or } \mathbb{P}(B) = 2/3.$$

This last formula, which can be written as $\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)$, can be extended to more general situations, where it is known as the **law of total probability**, and is a useful tool in **Bayes' formula** (see next sections).

(As an exercise, list the $6 \times 5 = 30$ outcomes in the sample space of this experiment, and use "brute force" to solve this problem.)

2.3 The law of total probability

Definition 2.3.1. A **partition** of the sample space Ω is a finite collection of events C_1, C_2, $\cdots$, C_n in Ω such that

(1) $C_i \cap C_j = \emptyset$, for any $i \neq j$ and $i, j = 1, 2, \cdots n$.

(2) $C_1 \cup C_2 \cup \cdots \cup C_n = \Omega$.

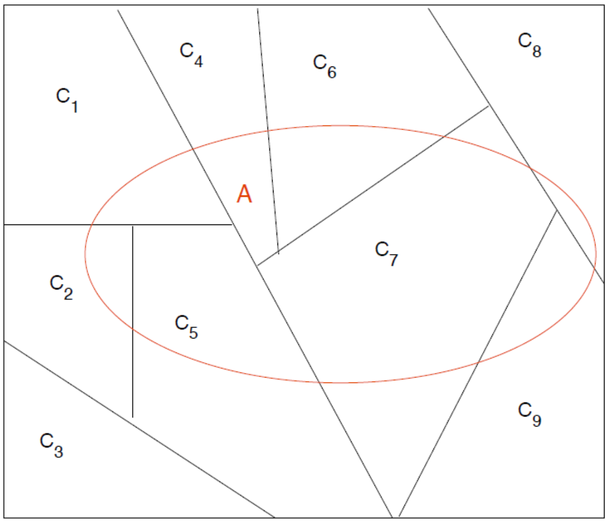


Figure 2.4: A partition $\{C_1, \cdots, C_9\}$ of the sample space Ω (large rectangle).

Remark 2.3.1. (1) We can refer to the figure above for illustration of the partition. The events $\{C_1, \cdots, C_9\}$ is a partition of the sample space and every outcome $\omega\,(\in \Omega)$ belongs to *exactly* one of the C_i.

(2) In the figure above, the event A can be written as the union $(A \cap C_1) \cup \cdots \cup (A \cap C_9)$ of mutually exclusive events.

Theorem 2.3.2 (Law of total probability)**.** Let $\mathbb{P}$ be a probability on Ω and let $\{C_1, C_2, \cdots, C_n\}$ be a partition of Ω chose so that $\mathbb{P}(C_i) \neq 0$ for all $i = 1, 2, \ldots, n$. Then for any event $A \subseteq \Omega$,

$$\mathbb{P}(A) = \mathbb{P}(A|C_1)\,\mathbb{P}(C_1) + \mathbb{P}(A|C_2)\,\mathbb{P}(C_2) + \cdots + \mathbb{P}(A|C_n)\,\mathbb{P}(C_n) = \sum_{i=1}^{n} \mathbb{P}(A|C_i)\,\mathbb{P}(C_i). \quad (2.2)$$

Sketch of proof: Because $\{C_1, C_2, \cdots, C_n\}$ is a partition,

$$\{(A \cap C_1), (A \cap C_2), \ldots, (A \cap C_n)\}$$

is a group of pairwise mutually exclusive events. By the distributive laws of events, their union is the event A (for instance, see Figure above). Using the generalization of the rule of total probability (see Remark 1.4.3), we can write

$$\mathbb{P}(A) = \mathbb{P}(A \cap C_1) + \mathbb{P}(A \cap C_2) + \cdots + \mathbb{P}(A \cap C_n). \tag{2.3}$$

Finish by using the multiplication principle (see Corollary 2.2.1),

$$\mathbb{P}(A \cap C_i) = \mathbb{P}(A|C_i)\,\mathbb{P}(C_i), \quad i = 1, 2 \ldots, n$$

and substituting into (2.3) to obtain the identity (2.2).

Remark 2.3.3. The most frequent use of the law of total probability comes in the case of a partition of the sample space into 2 events, say $\{B, B^c\}$. In this case, the law of total probability becomes the identity

$$\mathbb{P}(A) = \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c).$$

See the figure above for illustration.

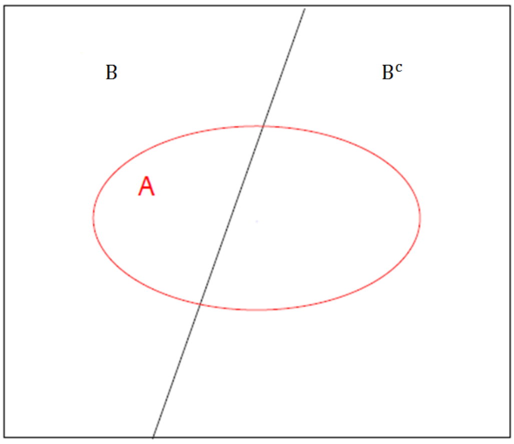


Figure 2.5: A partition into 2 events B and B^c.

Example 2.3.1. There are two urns, each containing coloured balls. In Urn I there are 2 white and 3 blue balls, in Urn II there are 3 white and 4 blue ones. A ball is drawn at random from Urn I and put into Urn II, and then a ball is drawn at random from Urn II and examined. What is the probability this ball is blue? We may suppose that picking out a ball of any colour from any urn is equally likely.
Solution: Let $E := \{\text{2nd ball is blue}\}$, $F := \{\text{1st ball is blue}\}$. We have

$$\begin{aligned} F &= \{\text{the number of balls in Urn II is increased by one blue ball}\} \\ &= \{\text{the Urn II contains 3 white and 5 blue balls}\} \end{aligned}$$

and

$$\begin{aligned} F^c &= \{\text{the number of balls in Urn II is increased by one white ball}\} \\ &= \{\text{the Urn II contains 4 white and 4 blue balls}\}. \end{aligned}$$

Hence we have

$$\mathbb{P}(E|F) = \mathbb{P}(E|\{\text{the Urn II contains 3 white and 5 blue balls}\}) = \frac{5}{8},$$

$$\mathbb{P}(E|F^c) = \mathbb{P}(E|\{\text{the Urn II contains 4 white and 4 blue balls}\}) = \frac{1}{2}.$$

Moreover, $\mathbb{P}(F) = 3/5$ and $\mathbb{P}(F^c) = 1 - 3/5 = 2/5$. By the law of total probability,

$$\mathbb{P}(E) = \mathbb{P}(E|F)\,\mathbb{P}(F) + \mathbb{P}(E|F^c)\,\mathbb{P}(F^c) = \frac{5}{8} \times \frac{3}{5} + \frac{1}{2} \times \frac{2}{5} = \frac{23}{40}.$$

2.4 Bayes' formula

Theorem 2.4.1 (Bayes' formula). Let $\{C_1, C_2, \ldots, C_n\}$ be a partition of Ω and $A \subseteq \Omega$ is an event such that $\mathbb{P}(A|C_i) \neq 0$ for $i = 1, 2, \ldots, n$. Then

$$\mathbb{P}(C_i|\, A) = \frac{\mathbb{P}(A|\, C_i)\,\mathbb{P}(C_i)}{\sum_{j=1}^{n} \mathbb{P}(A|\, C_j)\mathbb{P}(C_j)}.$$

Proof. By using the multiplication principle (Theorem 2.2.1) and the law of total probability (2.2), we have

$$\mathbb{P}(C_i|\, A) = \frac{\mathbb{P}(C_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|\, C_i)\,\mathbb{P}(C_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|\, C_i)\,\mathbb{P}(C_i)}{\sum_{j=1}^{n} \mathbb{P}(A|\, C_j)\,\mathbb{P}(C_j)}.$$

□

Example 2.4.1 (CFA, Reading #8, 2015). You are forecasting the sales of a building materials supplier by assessing the expansion plans of its largest customer, a homebuilder. You estimate the probability that the customer will increase its orders for building materials to 25%. If the customer does increase its orders from this supplier, you estimate the probability that the homebuilder will start a new development at 70%. If the customer does not increase its orders from the supplier, you estimate only a 20% chance that it will start the new development. Later, you find out that the homebuilder will start the new development. In light of this new information, what is your new (updated) probability that the builder will increase its orders from this supplier?

Solution: Set the following events:

$$I = \{\text{homebuilder increases order from the supplier}\},$$

$$N = \{\text{homebuilder starts new development}\}.$$

Then

$$I^c = \{\text{homebuilder does not increase order from the supplier}\}$$

and

$$N^c = \{\text{homebuilder does not start new development}\}.$$

$$\mathbb{P}(I) = 0.25 \Longrightarrow \mathbb{P}(I^c) = 1 - \mathbb{P}(I) = 1 - 0.25 = 0.75;$$

Moreover, it is given that $\mathbb{P}(N|\, I) = 0.7$ and $\mathbb{P}(N|\, I^c) = 0.2$.

We need to find the value of $\mathbb{P}(I|N)$. Using Bayes' formula,

$$\begin{aligned}\mathbb{P}(I|\, N) &= \frac{\mathbb{P}(N|\, I)\,\mathbb{P}(I)}{\mathbb{P}(N|\, I)\,\mathbb{P}(I) + \mathbb{P}(N|\, I^c)\,\mathbb{P}(I^c)}\\ &= \frac{0.7 \times 0.25}{0.7 \times 0.25 + 0.2 \times 0.75}\\ &= \frac{0.175}{0.175 + 0.15}\\ &= \frac{0.175}{0.325} \approx 0.5385.\end{aligned}$$

Example 2.4.2. Let A be the event that an individual tests positive for some disease and C be the event that the person actually has the disease. We can perform clinical trials to estimate the probability that a randomly chosen individual tests positive given that they have the disease,

$$\mathbb{P}(\{\text{tests positive}|\, \text{has the disease}\}) = \mathbb{P}(A|\, C),$$

by taking the individuals with the disease and playing the test.

The Public Health Department gives us the following information:

- A test for the disease yields a positive result 90% of the time when the disease is present.
- A test for the disease yields a positive result 1% of the time when the disease is not present.
- One person in 1,000 has the disease.

However, we would like to use the test as a method of *diagnosis* of the disease. Thus, we would like to be able to give the test and assert the chance that the person has the disease. That is we want to know the probability with the reverse conditioning

$$\mathbb{P}(\{\text{has the disease}|\, \text{tests positive}\}) = \mathbb{P}(C|\, A).$$

Solution: From the information given by the Public Health Department, we have $\mathbb{P}(A|\, C) = 90\%$, $\mathbb{P}(A|\, C^c) = 1\%$ and $\mathbb{P}(C) = 1/1000$. We want to evaluate the probability $\mathbb{P}(C|\, A)$. For this, we can apply Bayes' formula

$$\begin{aligned}\mathbb{P}(C|\, A) &= \frac{\mathbb{P}(A|\, C)\,\mathbb{P}(C)}{\mathbb{P}(A|\, C)\,\mathbb{P}(C) + \mathbb{P}(A|\, C^c)\,\mathbb{P}(C^c)}\\ &= \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.01 \times (1 - 0.001)}\\ &= \frac{0.0009}{0.01089} \approx 0.0826.\end{aligned}$$

2.5 Independent events

Definition 2.5.1. Two events A and B are said to be **independent** if either:

$$\mathbb{P}(A|\,B) = \mathbb{P}(A), \quad \text{i.e. } \mathbb{P}(B|A) = \mathbb{P}(B),$$

or equivalently,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B) \tag{2.4}$$

More generally, a finite collection of events $\{A_1, A_2, \ldots, A_n\}$ are called **mutually independent**, if

$$\mathbb{P}(\bigcap_{k \in I} A_k) = \prod_{k \in I} \mathbb{P}(A_k),$$

for every subset $I \subseteq \{1, 2, \ldots, n\}$.

Remark 2.5.1.

A, B **disjoint** $\Leftrightarrow$ If either event occurs, then the other <u>cannot</u> occur: $\mathbb{P}(A \cap B) = 0$.

A, B **independent** $\Leftrightarrow$ If either event occurs, this gives <u>no information</u> about the other: $\mathbb{P}(A|\,B) =$

Example 2.5.1. Let A and B be two independent events. Show that the following pairs of events are independent.

(1) A^c and B.

(2) A^c and B^c.

<u>Solution</u>:

(1) By the rule of total probability and definition of independent events, we have

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &\overset{\text{def.}}{=} \mathbb{P}(B) - \mathbb{P}(A)\,\mathbb{P}(B) \\ &= (1 - \mathbb{P}(A))\,\mathbb{P}(B) \\ &= \mathbb{P}(A^c)\,\mathbb{P}(B). \end{aligned}$$

(2) The second claim can be done in the same fashion.

$$\begin{aligned} \mathbb{P}(A^c \cap B^c) &= \mathbb{P}(A^c) - \mathbb{P}(A^c \cap B) \\ &\overset{\text{def.}}{=} \mathbb{P}(A^c) - \mathbb{P}(A^c)\,\mathbb{P}(B) \\ &= \mathbb{P}(A^c)\,(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A^c)\,\mathbb{P}(B^c). \end{aligned}$$

Example 2.5.2. On average, a faulty switch works on every 12 trials. Assume that trials are made independently. What is the probability that the switch works on the third trial (but not before)?

<u>Solution</u>: The probability that the switch works is $1/12$ and that it does not is $1 - 1/12 = 11/12$. Since the successive attempts to use the switch are independent, then

$$
\begin{aligned}
&\mathbb{P}(\{\text{works on the 3rd trial for the first time}\})\\
=&\mathbb{P}\left(\{\text{does not work on the 1st}\} \cap \{\text{does not work on the 2nd}\} \cap \{\text{works on the 3rd}\}\right)\\
=&\mathbb{P}\left(\{\text{does not work on the 1st}\}\right) \times \mathbb{P}\left(\{\text{does not work on the 2nd}\}\right) \times \mathbb{P}\left(\{\text{works on the 3rd}\}\right)\\
=&\frac{11}{12} \times \frac{11}{12} \times \frac{1}{12}\\
=&\frac{11^2}{12^3} \approx 0.07.
\end{aligned}
$$

EXERCISES

2.1. Six balls are to be randomly chosen from an urn containing 8 red, 10 green, and 12 blue balls.

(a) What is the probability at least one red ball is chosen?

(b) Given that no red balls are chosen, what is the conditional probability that there are exactly 2 green balls among the 6 chosen?

2.2. Three sorts of milk (A_1, A_2, A_3) are sterilized by either ash pasteurization (B_1) or cold pasteurization (B_2). A total of 2000 bottles of milk have been pasteurized by using these two procedures split according to the following table:

	A_1	A_2	A_3	total
B_1	260	350	200	810
B_2	440	450	300	1190
total	700	800	500	2000

Determine

(a) $\mathbb{P}(A_2)$ and $\mathbb{P}(B_1)$;

(b) $\mathbb{P}(A_2 \cap B_1)$ and $\mathbb{P}(A_2 \cup B_1)$;

(c) $\mathbb{P}(A_3|B_2)$, $\mathbb{P}(A_1|B_1^c)$, $\mathbb{P}(A_1|B_1 \cup B_2)$, $\mathbb{P}(A_1^c|B_2)$, $\mathbb{P}(B_1^c|A_3^c)$ and $\mathbb{P}(A_2^c|B_1 \cap B_2)$.

2.3. Let Ω be a sample space, A, $B \subseteq \Omega$ be two events and $\mathbb{P}$ be a probability defined on Ω. Given $\mathbb{P}(A) = 1/4$, $\mathbb{P}(A|B) = 1/4$ and $\mathbb{P}(B|A) = 1/2$, compute $\mathbb{P}(A \cup B)$ and $\mathbb{P}(A^c|B^c)$.

2.4. 52% of the students at a certain college are females. Five percent of the students in this college are majoring in computer science. Two percent of the students are women majoring in computer science. If a student is selected at random, find the conditional probability that

(a) this student is female, given that the student is majoring in computer science;

(b) this student is majoring in computer science, given that the student is female.

2.5. A total of 500 married working couples were polled about their annual salaries, with the following information resulting.

	Husband with less than 25,000	Husband with more than 25,000
Wife with less than 25,000	212	198
Wife with more than 25,000	36	54

Thus, for instance, in 36 of the couples the wife earned more and the husband earned less than 25,000. If one of the couples is randomly chosen, what is

(a) the probability that the husband earn less than 25,000;

(b) the conditional probability that the wife earns more than 25,000 given the husband earns more than this amount;

(c) the conditional probability that the wife earns more than 25,000 given that the husband earns less than this amount?

2.6. Celine is undecided as to whether to take a French course or a chemistry course. She estimates that her probability of receiving an A grade would be 1/2 in a French course and 2/3 in a chemistry course. If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?

2.7. Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement.

(a) If we assume that at each draw each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red?

(b) Now suppose that the balls have different weights, with each red ball having weight r and each white ball having weight w. Suppose that the probability that a given ball in the urn is the next one selected is its weight divided by the sum of the weights of all balls currently in the urn. Now what is the probability that both balls are red?

2.8. At a company a central computer is used in half of the working time in accountancy related to income, in 2/5 of the time in accountancy related to expenditure, and in the remaining time in accountancy related to taxes. The computer is idle in 15% of the time allocated to processing income, 9% of the time allocated to processing expenditure, and 18% of the time allocated to processing taxes.

(a) What is the probability that at a randomly chosen moment the computer is idle?

(b) What is the probability that the time is allocated to processing taxes given that the computer is idle?

(c) What is the probability that given that the computer is not idle, it is processing either income or expenditure?

2.9. There are two local factories that produce radios. Each radio produced at factory A is defective with probability 0.05, whereas each one produced at factory B is defective with probability 0.01. Suppose you purchase two radio that were produced at the same factory, which is equally likely to have been either factory A or factory B. If the first radio that you check is defective, what it the conditional probability that the other is also defective?

2.10. Suppose that an insurance company classifies people into one of three classes - good risks, average risks, and bad risks. Their records indicate that the probabilities that good, average, and bad risk persons will be involved in an accident over a 1-year span are, respectively, 0.05, 0.15 and 0.3. If 20 percent of the population are "good risks", 50 percent are "average risks", and 30 percent are "bad risks", what proportion of people have accidents in a fixed year? If policy holder A had no accidents in 1987, what is the probability that he or she is a good (average) risk?

2.11. You ask your neighbour to water a sickly plant while you are on vacation. Without water it will die with probability 0.8; with water it will die with probability 0.15. You are 90 percent certain that your neighbour will remember to water the plant.

(a) What is the probability that the plant will be alive when you return?

(b) If it is dead, what is the probability your neighbour forgot to water it?

2.12. Of three cards, one is painted red on both sides; one is painted black on both sides; and one is painted red on one side and black on the other. A card is randomly chosen and placed on a table. If the side facing up is red, what it the probability that the other side is also red?

2.13. A total of 600 of the 1,000 people in a retirement community classify themselves as Republicans, while the others classify themselves as Democrats. In a local election in which everyone voted, 60 Republicans voted for the Democratic candidate, and 50 Democrats voted for the Republican candidate. If a randomly chosen community member voted for the Republican, what is the probability that she or he is a Democrat?

2.14. Prostate cancer is the most common type of cancer found in males. As an indicator of whether a male has prostate cancer, doctors often perform a test that measures the level of the PSA protein (prostate specific antigen) that is produced only by the prostate gland. Although higher PSA levels are indicative of cancer, the test is notoriously unreliable. Indeed, the probability that a noncancerous man will have an elevated PSA level is approximately 0.135, with this probability increasing to approximately 0.268 if the man does have cancer. If based on other fact, a physician is 70 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that

(a) the test indicates an elevated PSA level;

(b) the test does not indicate an elevated PSA level?

Repeat the preceding, this time assuming that the physician initially believes there is a 30 percent chance the man has prostate cancer.

2.15. Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6, E_2 denote the event that the sum of the dice equals 7 and F the event that the first die equals 4.

(a) Is E_1 independent of F? Justify your answer.

(b) Is E_2 independent of F? Justify your answer.

2.16. Suppose a couple is waiting for a baby to be born, both of them has the combination of ab, where a and b are given genes. Assume that each of the couple will transmit independently either the genes a or b to the baby, with equal probability 1/2. The baby has a character $\mathcal{C}$ if and only if its parents transmit one of the combinations aa or bb to the baby. We consider the following events:

$$A = \{\text{the father transmits the gene } a\},$$
$$B = \{\text{the mother transmits the gene } b\},$$
$$C = \{\text{the baby shows the character } \mathcal{C}\}.$$

Determine the following probabilities

(a) $\mathbb{P}(A)$;

(b) $\mathbb{P}(B)$;

(c) $\mathbb{P}(C)$;

(d) $\mathbb{P}(A \cap B)$;

(e) $\mathbb{P}(A \cap C)$;

(f) $\mathbb{P}(B \cap C)$;

(g) $\mathbb{P}(A \cap B \cap C)$;

(h) Are the events A, B, C mutually independent?

2.17. A sequence of independent trials is to be performed. Each trial results in a success with probability p and a failure with probability $1 - p$. What is the probability that

(a) at least 1 success occurs in the first n trials;

(b) exactly k successes occur in the first n trials;

Chapter 3

Random variables and their distributions

Frequently, when an experiment is performed, we are interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, or $(6, 1)$. Also, in flipping a coin, we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results. These quantities of interests, or more formally, these real-valued functions defined on the sample space, are known as "random variables".

Example 3.0.3. A fair coin is flipped twice: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that

$$X((H, H)) = 2, \quad X((H, T)) = X((T, H)) = 1, \quad X((T, T)) = 0.$$

The function $X = X(\omega)$, $\omega \in \Omega$, is a random variable which represents the number of heads in this experiment.
Because the value of the random variable X is determined by the outcomes of the experiment, we may assign probabilities to the possible values of X:

$$\mathbb{P}(X = 0) = \mathbb{P}(\{(T, T)\}) = \frac{1}{4},$$

$$\mathbb{P}(X = 1) = \mathbb{P}(\{(H, T), (T, H)\}) = \frac{1}{2},$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{(H, H)\}) = \frac{1}{4}.$$

Since X must take on one of the values 0, 1, 2, we must have

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=0}^{2}\{X = i\}\right) = \sum_{i=0}^{2}\mathbb{P}(X = i).$$

Example 3.0.4. Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total number of n flips is made. If we let $X = X(\omega)$, $\omega \in \Omega$, denote the number of times the coin is flipped, then X is a random variable taking on one of the values $1, 2, 3, \ldots, n$ with respective probabilities

$$\mathbb{P}(X = 1) = \mathbb{P}(\{(H)\}) = p,$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\{(T, H)\}) = (1-p)p,$$

$$\mathbb{P}(X = 3) = \mathbb{P}(\{(T, T, H)\}) = (1-p)^2 p,$$

$$\vdots$$

$$\mathbb{P}(X = n-1) = \mathbb{P}(\{(\underbrace{T, T, \ldots, T,}_{(n-2)\text{ times}} H)\}) = (1-p)^{n-2} p,$$

$$\begin{aligned}\mathbb{P}(X = n) =& \mathbb{P}(\{(\underbrace{T, T, \ldots, T,}_{(n-1)\text{ times}} T)\} \text{ or } \{(\underbrace{T, T, \ldots, T,}_{(n-1)\text{ times}} H)\}) \\ =&(1-p)^n + (1-p)^{n-1} p \\ =&(1-p)^{n-1}(1-p+p) \\ =&(1-p)^{n-1}\end{aligned}$$

As a check, note that

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} \{X = i\}\right) &= \sum_{i=1}^{n} \mathbb{P}(X = i) \\ &= \sum_{i=1}^{n-1} p(1-p)^{i-1} + (1-p)^{n-1} \\ &= p\left[\frac{1-(1-p)^{n-1}}{1-(1-p)}\right] + (1-p)^{n-1} \\ &= 1 - (1-p)^{n-1} + (1-p)^{n-1} \\ &= 1.\end{aligned}$$

There are two important class of random variables, namely **discrete random variables** and **continuous random variables**, which will be discussed respectively in the next sections.

3.1 Discrete random variables

Definition 3.1.1. A function $X : \Omega \to M$ is called a **random variable**. If the codomain M is a finite or countable set, then X is called a **discrete random variable**.

Remark 3.1.1. The random variable X's defined in Example 3.0.3 and Example 3.0.4 are all discrete random variables.

Definition 3.1.2 (Probability mass function)**.** Let $\mathbb{P}$ be a probability defined on the sample space Ω, and $X : \Omega \to M$ be a discrete random variable. The function $f_X : \mathbb{R} \to [0,1]$ defined by

$$f_X(x) := \begin{cases} \mathbb{P}(X = x) & , \quad \text{if } x \in M, \\ 0 & , \quad \text{if } x \notin M; \end{cases}$$

is called the **probability mass function (p.m.f) of the random variable** X.

Example 3.1.1 (Revisit Example 3.0.3)**.** In a coin twice flips experiment, let X denotes number of heads. All the possible values that X can take is in $M = \{0, 1, 2\}$. And so the p.m.f of X is

$$f_X(x) = \begin{cases} \frac{1}{4}, & \text{if } x = 0, 2 \\ \frac{1}{2}, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 3.1.2 (Revisit Example 3.0.4)**.** In the experiment described in Example 3.0.4, all the possible values that X can take is in $M = \{1, 2, \ldots, n\}$. And so the p.m.f. of X is

$$f_X(k) = \begin{cases} (1-p)^{k-1}p, & \text{for } k = 1, 2 \ldots, n-1 \\ (1-p)^{k-1}, & \text{for } k = n \\ 0, & \text{otherwise} \end{cases}$$

Theorem 3.1.2 (Properties of probability mass functions)**.** Let $X : \Omega \to M$ be a discrete random variable with probability mass function f_X. Then the following hold:

(1) Positivity: $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$.

(2) Normalization: $\sum_{x \in M} f_X(x) = 1$.

Proof. (1) Since $0 \leq \mathbb{P}(X = x) \leq 1$, clearly $f_X(x)$ is nonnegative.

(2) Suppose $M := \{x_1, x_2, \cdots, x_m\}$, where $m = |M|$. Note that the events $\{X = x_j\}$ form a partition of Ω. Hence

$$\begin{aligned} \sum_{x \in M} f_X(x) &\overset{\text{def.}}{=} \sum_{x \in M} \mathbb{P}(X = x) \\ &= \mathbb{P}(X = x_1) + \mathbb{P}(X = x_2) + \cdots + \mathbb{P}(X = x_m) \\ &\overset{\text{Ax. 3}}{=} \mathbb{P}\left(\{X = x_1\} \cup \{X = x_2\} \cup \ldots \cup \{X = x_m\}\right) \\ &= \mathbb{P}(\Omega) \overset{\text{Ax. 1}}{=} 1. \end{aligned}$$

$\square$

Definition 3.1.3 (Cumulative distribution function)**.** Let $X : \Omega \to M$ be a discrete random variable with probability mass function f_X. Then the function $F_X : \mathbb{R} \to [0,1]$, defined by

$$F_X(x) := \mathbb{P}(X \leq x)$$

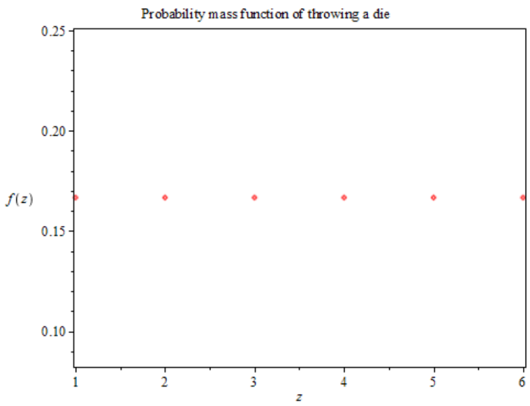
is called **cumulative distribution function (c.d.f) of the random variable** X**.**

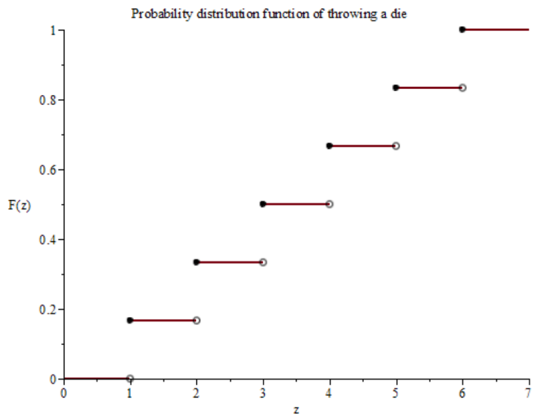
Remark 3.1.3. The relationship between f_X and F_X of a discrete random variable X is

$$F_X(x) = \sum_{\substack{y \in M \\ y \le x}} f_X(y).$$

Example 3.1.3. The p.m.f of rolling a fair die with Z denoting the face number is

$$f_Z(z) := \begin{cases} 1/6, & \text{if } z = 1, \\ 1/6, & \text{if } z = 2, \\ 1/6, & \text{if } z = 3, \\ 1/6, & \text{if } z = 4, \\ 1/6, & \text{if } z = 5, \\ 1/6, & \text{if } z = 6, \\ 0, & \text{otherwise.} \end{cases}$$





The c.d.f $F_Z(z) = \mathbb{P}(Z \le z)$ of rolling a fair die with Z denoting the face number is

$$F_Z(z) := \begin{cases} 0, & \text{if } z < 1, \\ f_Z(1) = 1/6, & \text{if } 1 \le z < 2, \\ f_Z(1) + f_Z(2) = 1/3, & \text{if } 2 \le z < 3, \\ f_Z(1) + f_Z(2) + f_Z(3) = 1/2, & \text{if } 3 \le z < 4, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) = 2/3, & \text{if } 4 \le z < 5, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) + f_Z(5) = 5/6, & \text{if } 5 \le z < 6, \\ f_Z(1) + f_Z(2) + f_Z(3) + f_Z(4) + f_Z(5) + f_Z(6) = 1, & \text{if } z \ge 6. \end{cases}$$

3.2 Statistic characteristic of discrete random variables

Definition 3.2.1 (Expectation/Mean)**.** Let X be a discrete random variable with values in M and probability mass function f_X . The number

$$\mathbb{E}[X] := \sum_{x \in M} x f_X(x) = \sum_{x \in M} x \mathbb{P}(X = x) \tag{3.1}$$

is called the **expectation (or expected value, mean)** of the random variable X.

Example 3.2.1 (Revisit Example 3.1.3)**.** Note that $M = \{1, 2, 3, 4, 5, 6\}$

$$\begin{aligned}
\mathbb{E}(Z) &= \sum_{z \in M} z \cdot f_Z(z) \\
&= 1 \cdot f_Z(1) + 2 \cdot f_Z(2) + 3 \cdot f_Z(3) + 4 \cdot f_Z(4) + 5 \cdot f_Z(5) + 6 \cdot f_Z(6) \\
&= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\
&= \frac{7}{2}.
\end{aligned}$$

Example 3.2.2. We say that I is an indicator variable for the event A if

$$I = \begin{cases} 1, & \text{if } A \text{ occurs;} \\ 0, & \text{if } A^c \text{ occurs.} \end{cases}$$

Find $\mathbb{E}(I)$.

Solution: Since $f_I(1) = \mathbb{P}(I = 1) = \mathbb{P}(A)$ and $f_I(0) = \mathbb{P}(I = 0) = \mathbb{P}(A^c)$, we have

$$\mathbb{E}[I] = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

That is the expectation of the indicator variable for the event A is equal to probability that A occurs.

Remark 3.2.1. (1) When each value x has equal probability, i.e., $f_X(x) = \frac{1}{|M|}$, $\forall x \in M$, then by using the Fundamental Formula the expectation reduces to the common **arithmetic mean**:

$$\mathbb{E}[X] := \sum_{x \in M} x f_X(x) = \frac{1}{|M|} \sum_{x \in M} x.$$

(2) Note that the expectation of a random variable may be finite or infinite. If M is a finite subset of $\mathbb{R}$, the sum at the most right side of (4.1) is finite and in this case we call the mean exists. If M is countably infinite, then the series maybe divergent and thus the mean becomes infinite and we say it does not exist.

(3) More generally, given a discrete random variable X and a function $h : M \to \mathbb{R}$. The **expectation of** $h(X)$ is defined as

$$\mathbb{E}[h(X)] := \sum_{x \in M} h(x) f_X(x).$$

(4) For instance, if we choose $h(X) = aX + b$, where $a, b \in \mathbb{R}$ are constants. Then we have

$$\mathbb{E}(aX+b) \overset{\text{def.}}{=} \sum_{x \in M} (ax+b) f_X(x) = a \sum_{x \in M} x f_X(x) + b \underbrace{\sum_{x \in M} f_X(x)}_{=1 \text{ by Normal.}} = a\mathbb{E}[X] + b \cdot 1 = a\mathbb{E}(X) + b.$$

Or more generally, we have the **linear property** of taking expectations, i.e.,

$$\mathbb{E}[ah(X)+bg(Y)] = a\mathbb{E}[h(X)] + b\mathbb{E}[g(Y)],$$

where h, g are two given functions and X, Y are two discrete random variables.

Definition 3.2.2 (Variance and Standard deviation)**.** Let X be a discrete random variable with values in M and probability mass function f_X . The number

$$\mathrm{Var}(X) := \mathbb{E}[(X-\mathbb{E}[X])^2] = \sum_{x\in M}(x-\mathbb{E}[X])^2 f_X(x) \tag{3.2}$$

is called the **variance of** X, subject to the RHS of (3.2) is finite. Since the variance is non-negative, $\sqrt{\mathrm{Var}(X)}$ exists and is called the **standard deviation of** X.

Remark 3.2.2. (1) The variance and standard deviation measure the spread of the random variable about its mean.

(2) An alternative formula for $\mathrm{Var}(X)$ is derived as follows: let $\mu = \mathbb{E}(X)$,

$$\begin{aligned}
\mathrm{Var(X)} &\overset{\text{def.}}{=} \mathbb{E}[(X-\mu)^2] \\
&= \sum_{x\in M}(x-\mu)^2 f_X(x) \\
&= \sum_{x\in M}(x^2-2\mu x+\mu^2) f_X(x) \\
&= \underbrace{\sum_{x\in M} x^2 f_X(x)}_{=\mathbb{E}[X^2]} - 2\mu \underbrace{\sum_{x\in M} x f_X(x)}_{=\mu} + \mu^2 \underbrace{\sum_{x\in M} f_X(x)}_{=1 \text{ by Normal.}} \\
&= \mathbb{E}[X^2] - \mu^2.
\end{aligned}$$

That is,

$$\mathrm{Var(X)} = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{3.3}$$

Example 3.2.3 (Revisit Example 3.1.3)**.** Calculate $\mathrm{Var}(Z)$ if Z represents the outcome when a fair die is rolled.

<u>Solution</u>: It is shown in Example 3.2.1 that $\mathbb{E}[Z] = \frac{7}{2}$. Also,

$$\begin{aligned}
\mathbb{E}[Z^2] &= 1^2\cdot\frac{1}{6} + 2^2\cdot\frac{1}{6} + 3^2\cdot\frac{1}{6} + 4^2\cdot\frac{1}{6} + 5^2\cdot\frac{1}{6} + 6^2\cdot\frac{1}{6} \\
&= \frac{91}{6}.
\end{aligned}$$

Hence by (3.3),

$$\mathrm{Var}(Z) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Theorem 3.2.3. Let X and Y be two discrete random variables. Then

(1) For any constants $a, b \in \mathbb{R}$,

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

(2) Two random variables X and Y are said to be independent if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. If X and Y are independent, then for any constants $a, b \in \mathbb{R}$,

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y).$$

Proof. (1) By definition of variance, we have

$$\begin{aligned} \mathrm{Var}(aX+b) &= \sum_{x \in M} (ax + b - \mathbb{E}[aX+b])^2 \cdot f_X(x) \\ &\overset{\text{Linearity}}{=} \sum_{x \in M} (ax + b - a\mathbb{E}(X) - b)^2 \cdot f_X(x) \\ &= a^2 \sum_{x \in M} (x - \mathbb{E}[X])^2 \cdot f_X(x) \\ &= a^2 \mathrm{Var}(X). \end{aligned}$$

(2) The proof is left as exercise (See also Exercise 3.13).
Hint: Two random variables X and Y are said to be independent if

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

□

3.3 Specific types of discrete random variables

3.3.1 Bernoulli and Binomial random variables

Definition 3.3.1 (Bernoulli random variables)**.** A random variable X take values 1 and 0 with probabilities $p \in (0,1)$ and q $(:= 1-p)$, respectively, is called a **Bernoulli random variable** with parameter p. Sometimes we think of these values as representing the "success rate" or the "failure rate" of a trial. The p.m.f. is

$$f_X(0) = 1 - p, \quad f_X(1) = p.$$

Experiments involving a Bernoulli variable are called **Bernoulli trials**.

Example 3.3.1 (Bernoulli trials)**.** (1) Tossing a fair coin with $p = 1/2$.

(2) A medical treatment maybe effective with a probability p and ineffective with probability $(1-p)$.

Theorem 3.3.1. For a Bernoulli random variable X with parameter p,

$$\mathbb{E}[X] = p, \quad \mathrm{Var}(X) = p(1-p).$$

Proof. We have

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1-p) = p$$

and

$$\mathbb{E}(X^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p,$$

from which we have

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = p - p^2 = p(1-p).$$

□

Definition 3.3.2. When perform n independent Bernoulli trials, i.e. $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d) Bernoulli random variables with parameter p and count the total number of successes $Y := X_1 + X_2 + \cdots + X_n$. Then Y is called a **Binomial random variable with parameters** (n, p), where $n \in \mathbb{N}$ and $p \in (0, 1)$. Equivalently, Y is called a Binomial random variable with parameters (n, p), if the p.m.f of Y is given by

$$f_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{3.4}$$

The experiment generating a binomial variable is called a **binomial experiment**.

Remark 3.3.2. (1) The definition of Binomial random variable (3.4) may be verified by first noting that the probability of the particular sequence of n outcomes containing k success and $(n-k)$ failures is $p^k(1-p)^{n-k}$ by the assumed independence of trials. Equation (3.4) then follows, since there are $\binom{n}{k}$ different sequence of the n outcomes leading to k successes and $(n-k)$ failures. This can most easily be seen by noting that there are $\binom{n}{k}$ different choices of the k trials that result in sucess.

(2) If Y is a binomial random variable with parameters (n, p), then we use the shorthand notation:

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

(3) The normalization property of the p.m.f f_Y given by (3.4) can be verified as follows:

$$\sum_{k=0}^{n} f_Y(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \overset{\text{Bino. formu.}}{=} [p + (1-p)]^n = 1.$$

Theorem 3.3.3. Let Y be a binomial random variable with parameters (n, p). Then we have

$$\mathbb{E}[Y] = np, \quad \mathrm{Var}(Y) = np(1-p).$$

Proof. By linear property of expectation, we have

$$\mathbb{E}[Y] = \mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1] = np.$$

By Theorem 3.5.3, we obtain

$$\begin{aligned}\mathrm{Var}(Y) = \mathrm{Var}(X_1 + X_2 + \cdots + X_n) =& 1^2 \cdot \mathrm{Var}(X_1) + 1^2 \cdot \mathrm{Var}(X_2) + \cdots + 1^2 \cdot \mathrm{Var}(X_n) \\ =& n \cdot \mathrm{Var}(X_1) = np\,(1-p).\end{aligned}$$

□

3.3.2 Geometric distribution

Definition 3.3.3. A **Geometric distribution with parameter** $p \in (0,1)$ is a random varibale with the geometric mass function:

$$f(k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots \tag{3.5}$$

Remark 3.3.4. Geometric distribution arises in the following way. Suppose that independent Bernoulli trials (parameter p) are performed at times $1, 2, \dots$ Let W be the time which elapses until the first success. Then the event

$$\{W = k\} = \{\text{At the } k\text{th trial, success occurs for the first time}\}.$$

So by the the independence of the trials,

$$f(k) = \mathbb{P}(W = k) = (1-p)^{k-1}p.$$

Example 3.3.2. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that

(a) exactly n draws are needed?

(b) at least k draws are needed?

Solution: If we let X denote the number of draws needed to select a black ball, then X satisfies Equation (3.5) with $p = M/(M+N)$. Hence,

(a)

$$\mathbb{P}(X = n) = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}.$$

(b)

$$\begin{aligned}
\mathbb{P}(X \geq k) &= \sum_{n=k}^{\infty} \mathbb{P}(X = n) \\
&= \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1} \\
&= \frac{\dfrac{M}{M+N}\left(\dfrac{N}{M+N}\right)^{k-1}}{1 - \dfrac{N}{M+N}} \\
&= \left(\frac{N}{M+N}\right)^{k-1}.
\end{aligned}$$

Exercise 3.3.1. (See also Exercise 3.14)
Prove that if X follows Geometric distribution with parameter p, then

$$\mathbb{E}[X] = \frac{1}{p}, \quad \operatorname{Var}(X) = \frac{1-p}{p^2}.$$

3.3.3 Poisson random variable

Definition 3.3.4. A random variable X that takes one of the values $0, 1, 2, \dots$ is said to be a **Poisson random variable with parameter** λ, if for some $\lambda > 0$ the p.m.f. is given by

$$f_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2 \dots \tag{3.6}$$

Remark 3.3.5. (1) The equation (3.6) defines a probability mass function, since $f_X(k) \geq 0$ for $k = 0, 1, \dots$ and

$$\sum_{k=0}^{\infty} f_X(k) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

(2) If X is a Poisson random variable with parameter λ, then we use the shorthand notation:

$$X \sim Poi(\lambda).$$

(3) Poisson distribution focuses on the number of discrete events or occurrences over a specified interval or continuum (time, length, distance, etc. For instance, see the following example).

Example 3.3.3. Suppose that the number of typographical errors on a single page of this paper has a Poisson distribution with parameter $\lambda = 1/2$. Calculate the probability that there is at least one error on this page.

Solution: Letting X denote the number of errors on this page, we have

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-1/2} \simeq 0.393.$$

Theorem 3.3.6. If $X \sim Poi(\lambda)$, then

$$\mathbb{E}(X) = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

Proof. The mean can be obtained through the following calculation:

$$\begin{aligned}
\mathbb{E}[X] :=& \sum_{i=0}^{\infty} i\mathbb{P}(X = i)\\
=& \sum_{i=0}^{\infty}\frac{ie^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=1}^{\infty}\frac{i\lambda^i}{i!}\\
=& \lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}, \quad \text{by letting } j = i-1\\
=& \lambda, \quad \text{since } \sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = e^{\lambda}.
\end{aligned}$$

To determine its variance, we first compute $\mathbb{E}[X^2]$.

$$
\begin{aligned}
\mathbb{E}[X^2] :=& \sum_{i=0}^{\infty} i^2 \mathbb{P}(X=i) = \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda}\lambda^i}{i!} \\
=& \sum_{i=1}^{\infty} \frac{i^2 e^{-\lambda}\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{i e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\
=& \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda}\lambda^j}{j!}, \quad \text{by letting } j = i-1 \\
=& \lambda \left[\sum_{j=0}^{\infty} \frac{j e^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \right] = \lambda(\lambda+1),
\end{aligned}
$$

where the final equality follows because the first sum is the expected value of a Poisson random variable with parameter λ and the second is the sum of the probabilities of this random variable. Therefore, since we have shown that $\mathbb{E}[X] = \lambda$, we obtain

$$
\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda.
$$

□

Hence, the expected value and variance of a Poisson random variable are both equal to its parameter λ.

3.4 Continuous random variables

In the sections above, we considered discrete random variables - that is, random variables whose set of possible values is either finite or countably infinite. However, there also exist random variables whose set of possible values is uncountable.

Definition 3.4.1. Let $M \subseteq \mathbb{R}$ be a interval or a union of intervals and Ω be a sample space. A function $X : \Omega \to M$ is called a **continuous random variable** if there exists an integrable function $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$
\begin{aligned}
& f_X(x) \geq 0, \forall x \in M, \\
& \int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1; \qquad (3.7)
\end{aligned}
$$

for which we have

$$
F_X(x) := \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(y)\mathrm{d}y.
$$

The function f_X is called the **probability density function (p.d.f) of** X and F_X is called the **cumulative distribution function (c.d.f) of** X.

Proposition 3.4.1. Let f_X be a p.d.f of a continuous random variable X. Then $\forall a < b \in M$.

$$
\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\mathrm{d}x.
$$

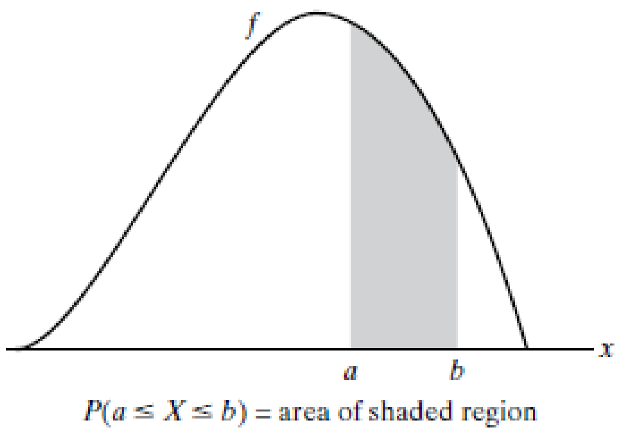


Figure 3.1: Probability density function f.

In particular, $\mathbb{P}(X = a) = 0$ for all $a \in M$, and

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b).$$

Proof.

$$\begin{aligned}
\mathbb{P}(a \leq X \leq b) &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) \\
&= F_X(b) - F_X(a) \\
&= \int_{-\infty}^{b} f_X(x)\mathrm{dx} - \int_{-\infty}^{\mathrm{a}} \mathrm{f_X(x)dx} \\
&= \int_{a}^{b} f_X(x)\mathrm{dx}.
\end{aligned}$$

Clearly, the event $\{X = a\}$ is the same as $a \leq X \leq a$. Choosing $a = b$ above, we obtain

$$\mathbb{P}(X = a) = \int_{a}^{a} f_X(x)\mathrm{dx} = 0.$$

So $\mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) - \mathbb{P}(X = b) = \mathbb{P}(a \leq X \leq b)$ etc. $\square$

Remark 3.4.1. From the proposition above, whenever the c.d.f F_X is differentiable, we have the following relation

$$f_X(x) = \frac{\mathrm{d}\, F_X(x)}{\mathrm{d}\, x} = F'_X(x).$$

3.5 Statistic characteristic of continuous random variables

Definition 3.5.1 (Expectation/Mean)**.** Let X be a continuous random variable with p.d.f f_X. The real number

$$\mathbb{E}[X] := \int_{-\infty}^{+\infty} x f_X(x)\mathrm{dx}$$

is called the **expectation (or expected value, mean)** of X whenever the integral exists.

Remark 3.5.1. (1) More generally, given a continuous random variable X and a function $h : M \to \mathbb{R}$. The expectation of $h(X)$ is defined as

$$\mathbb{E}[h(X)] := \int_{-\infty}^{+\infty} h(x) f_X(x) \mathrm{dx}.$$

(2) For instance, if we choose $h(X) = aX + b$, where $a, b \in \mathbb{R}$ are constants. Then we have

$$\mathbb{E}(aX+b) \overset{\text{def.}}{=} \int_{-\infty}^{+\infty} (ax+b) f_X(x) \mathrm{dx} = \mathrm{a} \int_{-\infty}^{+\infty} \mathrm{xf_X(x)dx} + \mathrm{b} \underbrace{\int_{-\infty}^{+\infty} f_X(x) \mathrm{dx}}_{=1 \text{ by } (3.7)} = \mathrm{a}\mathbb{E}[\mathrm{X}] + \mathrm{b} \cdot 1 = \mathrm{a}\mathbb{E}(\mathrm{X}) +$$

Or more generally, we have the **linear property** of taking expectations, i.e.,

$$\mathbb{E}[ah(X) + bg(Y)] = a\mathbb{E}[h(X)] + b\mathbb{E}[g(Y)],$$

where h, g are two given functions and X, Y are two continuous random variables.

Definition 3.5.2 (Variance and Standard deviation)**.** Let X be a continuous random variable with p.d.f f_X. The positive real number

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x) \mathrm{dx}$$

is called the **variance** of X whenever the integral exists. The square root of the variance $\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X.

Remark 3.5.2. Similarly as for a discrete random variable, an alternative formula for $\mathrm{Var}(X)$ in continuous case is as follows:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{+\infty} x^2 f_X(x) \mathrm{dx} - \left(\int_{-\infty}^{+\infty} x f_X(x) \mathrm{dx} \right)^2.$$

As for a discrete random variable, the variance of a continuous random variable also has the following properties (prove them, as exercises):

Theorem 3.5.3. Let X and Y be two continuous random variables. Then

(1) For any constants $a, b \in \mathbb{R}$,

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X).$$

(2) Two random variables X and Y are said to be independent if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. If X and Y are independent, then for any constants $a, b \in \mathbb{R}$,

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y).$$

Example 3.5.1. Find the constant C for the following p.d.f with $k > 0$,

$$f(x) := \begin{cases} Ce^{-kx}, & \text{if } x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

and the expectation and variance.

Solution: We should have

$$1 = \int_{-\infty}^{+\infty} f(x)\mathrm{dx} = C\int_0^{+\infty} e^{-kx}\mathrm{dx} = C\left[\frac{-e^{-kx}}{k}\right]_0^{+\infty} = \frac{C}{k},$$

which gives $C = k$. For expectation we have by integration by parts

$$\begin{aligned}
\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)\mathrm{dx} = \int_0^{+\infty} kxe^{-kx}\mathrm{dx} = -\int_0^{+\infty} x\mathrm{d}(e^{-\mathrm{kx}})\\
=\left[(-x)e^{-kx}\right]_0^{+\infty} - \frac{1}{k}\int_0^{+\infty} e^{-kx}\mathrm{d}(-\mathrm{kx})\\
=\left[(-x)e^{-kx}\right]_0^{+\infty} - \frac{1}{k}\left[e^{-kx}\right]_0^{+\infty}\\
=0 - \left(-\frac{1}{k}\right) = \frac{1}{k}.
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2f(x)\mathrm{dx} = \int_0^{+\infty} kx^2e^{-kx}\mathrm{dx} = -\int_0^{+\infty} x^2\mathrm{d}(e^{-\mathrm{kx}})\\
=\left[(-x^2)e^{-kx}\right]_0^{+\infty} - \int_0^{+\infty} e^{-kx}\mathrm{d}(-\mathrm{x}^2),\ \text{using integration by parts}\\
=\left[(-x^2)e^{-kx}\right]_0^{+\infty} + 2\int_0^{+\infty} xe^{-kx}\mathrm{dx}\\
=\left[(-x^2)e^{-kx}\right]_0^{+\infty} + \frac{2}{k}\int_0^{+\infty} kxe^{-kx}\mathrm{dx}\\
=\left[(-x^2)e^{-kx}\right]_0^{+\infty} + \frac{2}{k}\mathbb{E}[X] = 0 + \frac{2}{k}\cdot\frac{1}{k} = \frac{2}{k^2}.
\end{aligned}$$

Thus,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{k^2} - \frac{1}{k^2} = \frac{1}{k^2}.$$

3.6 Specific types of continuous random variables

3.6.1 Uniform distribution

Definition 3.6.1 (Uniform distribution)**.** Let $M = (a, b) \subseteq \mathbb{R}$. The random variable $X : \Omega \to (a, b)$ is said to be **uniformly** distributed over the interval (a, b) if its p.d.f is given by

$$f_X(x) := \begin{cases} \frac{1}{b-a}, & \text{if } x \in (a, b)\\ 0, & \text{otherwise.}\end{cases} \tag{3.8}$$

Note that Equation (3.8) is a density function, since $f_X(x) \geq 0$ and

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{dx} = \int_a^b \frac{1}{b-a}\mathrm{dx} = \frac{1}{b-a}(b-a) = 1.$$

Theorem 3.6.1. Let X be a uniformly distributed random variable over the interval (a, b). Then

(1) Its c.d.f is given by

$$F_X(x) := \begin{cases} 0, & \text{if } x \leq a, \\ \dfrac{x-a}{b-a}, & \text{if } x \in (a,b), \\ 1, & \text{otherwise.} \end{cases}$$

(2) For expectation and variance we have

$$\mathbb{E}[X] = \frac{a+b}{2},$$

$$\mathrm{Var}(X) = \frac{(a-b)^2}{12}.$$

Proof. (1) If $x \leq a$, then $F_X(x) = 0$.

If $x \in (a, b)$, we have

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\mathrm{dy} = \underbrace{\int_{-\infty}^{a} f_X(y)\mathrm{dy}}_{=0} + \int_{a}^{x} f_X(y)\mathrm{dy} = \int_{a}^{x} \frac{1}{b-a}\mathrm{dy} = \frac{x-a}{b-a}.$$

If $x \geq b$,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\mathrm{dy} = \underbrace{\int_{-\infty}^{a} f_X(y)\mathrm{dy}}_{=0} + \int_{a}^{b} f_X(y)\mathrm{dy} + \underbrace{\int_{b}^{x} f_X(y)\mathrm{dy}}_{=0} = \int_{a}^{b} \frac{1}{b-a}\mathrm{dy} = 1.$$

(2) By the definition of expectation we have

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x)\mathrm{dx} = \frac{1}{b-a}\int_{a}^{b} x\mathrm{dx} \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}. \end{aligned}$$

Moreover,

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{dx} = \frac{1}{b-a}\int_{a}^{b} x^2\mathrm{dx} \\ &= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\frac{b^3-a^3}{3} = \frac{a^2+ab+b^2}{3}. \end{aligned}$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{(a-b)^2}{12}.$$

$\square$

Example 3.6.1. If X is uniformly distributed over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 6$, and (c) $3 < X < 8$.

Solution:

(a) $\mathbb{P}(X < 3) = \int_0^3 \frac{1}{10}\mathrm{dx} = \frac{3}{10}$.

(b) $\mathbb{P}(X > 6) = \int_6^{10} \frac{1}{10}\mathrm{dx} = \frac{4}{10}$.

(c) $\mathbb{P}(3 < X < 8) = \int_3^8 \frac{1}{10}\mathrm{dx} = \frac{1}{2}$.

3.6.2 Exponential random variables

Definition 3.6.2 (Exponential random variable)**.** Let $M = (0, +\infty)$ and $\theta > 0$. The random variable $X : \Omega \to M$ is said to be **exponential** distributed with parameter θ if its p.d.f is given by

$$f_X(x) := \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & \text{if } x \in [0, +\infty), \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

Note that Equation (3.9) is a density function, since $f_X(x) \geq 0$ and

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{dx} = \frac{1}{\theta}\int_0^{+\infty} e^{-x/\theta}\mathrm{dx} \overset{y=x/\theta}{=} \int_0^{+\infty} e^{-y}\mathrm{dy} = 1.$$

This distribution very often occurs in practice as description of the time elapsing between unpredictable events (such as telephone calls, earthquakes, emissions of radioactive particles, and arrival of buses, and so on).

Theorem 3.6.2. Let X be an exponential distributed random variable with parameter θ. We have

(1) Its c.d.f is given by

$$F_X(x) = \begin{cases} 1 - e^{-x/\theta}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(2) For expectation and variance, we have

$$\mathbb{E}[X] = \theta.$$

$$\mathrm{Var}(X) = \theta^2.$$

Proof. (1) We first determine the c.d.f.

If $x \leq 0$, then $F_X(x) = 0$. If $x > 0$, then

$$F_X(x) := \int_{-\infty}^{x} f_X(y)\mathrm{dy} = \int_0^x e^{-y/\theta}\mathrm{d(y/\theta)} \overset{\mathrm{u=y/\theta}}{=} \int_0^{\mathrm{x}/\theta} e^{-\mathrm{u}}\mathrm{du} = \left[-e^{-\mathrm{u}}\right]_0^{\mathrm{x}/\theta} = 1 - e^{-\mathrm{x}/\theta}.$$

(2) For the expectation, using integration by parts, we have

$$\begin{aligned}
\mathbb{E}[X] = &\int_{-\infty}^{+\infty} x f_X(x)\mathrm{dx} = \theta \int_0^{+\infty} \frac{x}{\theta} e^{-x/\theta}\mathrm{d(x/\theta)} \\
\overset{y=x/\theta}{=} &\theta \int_0^{+\infty} y e^{-y}\mathrm{dy} \\
= &\theta \left(\underbrace{\left[-ye^{-y}\right]_0^{+\infty}}_{=0} + \underbrace{\int_0^{+\infty} e^{-y}\mathrm{dy}}_{=1} \right) = \theta.
\end{aligned}$$

By a similar argument,

$$\begin{aligned}
\mathbb{E}[X^2] = &\int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{dx} = \theta^2 \int_0^{+\infty} \frac{x^2}{\theta^2} e^{-x/\theta}\mathrm{d(x/\theta)} \\
\overset{y=x/\theta}{=} &\theta^2 \int_0^{+\infty} y^2 e^{-y}\mathrm{dy} \\
= &\theta^2 \left(\underbrace{\left[-y^2e^{-y}\right]_0^{+\infty}}_{=0} + 2\underbrace{\int_0^{+\infty} y e^{-y}\mathrm{dy}}_{=1} \right) = 2\theta^2.
\end{aligned}$$

Thus,

$$\mathrm{Var}(X) = 2\theta^2 - \theta^2 = \theta^2.$$

And so the standard deviation is θ.

□

Example 3.6.2. In a storm the time elapsed between two consecutive thunderbolts is an exponential random variable. Its standard deviation is 1 minute.

(1) What is the probability that the time gap between two thunderbolts is at most 2 mins?

(2) What is the probability that the time gap between two thunderbolts is at least 1 min?

Solution: Denote by X the time gap in minutes between two consecutive thunderbolts. Using Theorem 3.6.2, which says that both the mean and the standard deviation of an exponential random variable is θ, we conclude that its value is 1.

Hence, the c.d.f is

$$F_X(x) = 1 - e^{-x}, \quad x > 0.$$

Thus,

(1) $\mathbb{P}(X \leq 2) = F_X(2) = 1 - e^{-2} \approx 0.8647$.

(2) $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 1 - F_X(1) = 1 - (1 - e^{-1}) = e^{-1} \approx 0.3679$.

3.6.3 Normal random variables

Definition 3.6.3 (Normal random variables)**.** We say that X is a **normal random variable or X is normally distributed**, with parameters μ and σ^2 if the p.d.f of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad \text{for } -\infty < x < \infty.$$

This p.d.f is a bell-shaped curve that is symmetric about $x = \mu$ (See the figure below).

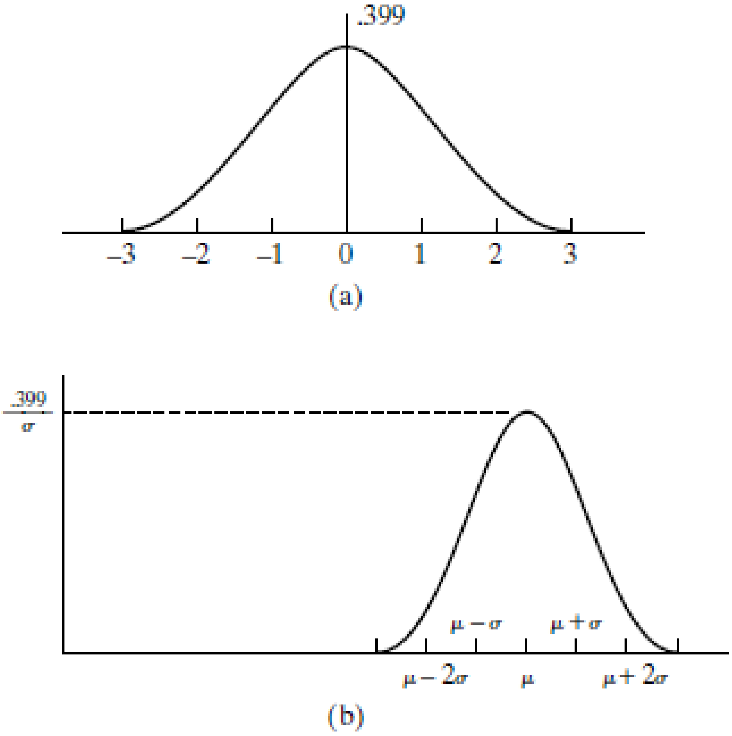


Figure 3.2: Normal density function: (a) $\mu = 0$, $\sigma = 1$; (b) arbitrary μ, σ.

Remark 3.6.3. (1) To prove that $f_X(x)$ is indeed a p.d.f, we need to show that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} \mathrm{dx} = 1.$$

Making the substitution $y = (x - \mu)/\sigma$, we see that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} \mathrm{dx} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} \mathrm{dy}.$$

Hence, we must show that

$$\int_{-\infty}^{+\infty} e^{-y^2/2} \mathrm{dy} = \sqrt{2\pi}. \tag{3.10}$$

Toward this end, let $I = \int_{-\infty}^{+\infty} e^{-y^2/2}\mathrm{dy}$. Then

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-y^2/2}\mathrm{dy} \int_{-\infty}^{+\infty} e^{-x^2/2}\mathrm{dx} \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(y^2+x^2)/2}\mathrm{dy}\,\mathrm{dx}. \end{aligned}$$

We now evaluate the double integral by means of a change of variables to polar coordinates. (That is, let $x = r\cos\theta$, $y = r\sin\theta$, and $\mathrm{dydx} = r\,\mathrm{d}\theta\,\mathrm{dr}$.) Thus,

$$\begin{aligned} I^2 &= \int_0^{+\infty} \int_0^{2\pi} e^{-r^2/2} r\,\mathrm{d}\theta\,\mathrm{dr} \\ &= 2\pi \int_0^{+\infty} r e^{-r^2/2}\,\mathrm{dr} \\ &= -2\pi \left[e^{-r^2/2} \right]_0^{+\infty} \\ &= 2\pi. \end{aligned}$$

Hence, $I = \sqrt{2\pi}$, and the result is proved.

(2) If X is normally distributed with parameters μ and σ^2, then we use the shorthand notation:

$$X \sim \mathrm{Normal}(\mu, \sigma^2).$$

In the particular case, when $\mu = 0$ and $\sigma^2 = 1$, i.e.

$$X \sim \mathrm{Normal}(0, 1),$$

X is called a **standard normal** random variable.

Theorem 3.6.4. If $X \sim \mathrm{Normal}(\mu, \sigma^2)$, then

$$\begin{aligned} \mathbb{E}[X] &= \mu, \\ \mathrm{Var}(X) &= \sigma^2. \end{aligned}$$

Proof. By the definition, the p.d.f. of X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

Let $y := \dfrac{x-\mu}{\sigma}$, then

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{+\infty} x f_X(x)\mathrm{dx} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{dx} \\ &= \frac{\sigma}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (\sigma y + \mu) e^{-\frac{y^2}{2}}\,\mathrm{dy} \\ &= \frac{1}{\sqrt{2\pi}} \left[\sigma \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}}\,\mathrm{dy} + \mu \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,\mathrm{dy} \right] = \mu. \end{aligned}$$

Since $y \mapsto ye^{-\frac{y^2}{2}}$ is an odd function over $\mathbb{R}$, so the first integration in the brackets above equals 0, the second integration equals $\sqrt{2\pi}$, due to (3.10).
Thus,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mu^2.$$

And

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{dx} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{dx} \\
&= \frac{\sigma}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (\sigma y + \mu)^2 e^{-\frac{y^2}{2}} \mathrm{dy} \\
&= \frac{1}{\sqrt{2\pi}} \left[\sigma^2 \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} \mathrm{dy} + 2\sigma\mu \underbrace{\int_{-\infty}^{+\infty} ye^{-\frac{y^2}{2}} \mathrm{dy}}_{=0} + \mu^2 \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} \mathrm{dy}}_{=\sqrt{2\pi}} \right] \\
&= \left(\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} \mathrm{dy} \right) + \mu^2 \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(- \underbrace{\left[ye^{-\frac{y^2}{2}} \right]_{-\infty}^{+\infty}}_{=0} + \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} \mathrm{dy}}_{=\sqrt{2\pi}} \right) + \mu^2, \text{ (using IBP, with } u = y \text{ and } \mathrm{dv} = ye^{-y^2/2}\mathrm{dy}) \\
&= \sigma^2 + \mu^2.
\end{aligned}$$

Hence, we have $\mathrm{Var}(X) = \sigma^2$. □

Remark 3.6.5. (1) It is customary to denote the c.d.f of a standard normal random variable by Φ. That is, $\forall x \in \mathbb{R}$,

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} \mathrm{dy}.$$

The values of $\Phi(x)$ with nonnegative x are given in the table in Page 57. For negative values of x, $\Phi(x)$ can be obtained from the relationship

$$\Phi(-x) = 1 - \Phi(x), \quad \text{for } -\infty < x < \infty. \tag{3.11}$$

Equation (3.11) follows from the symmetry of the standard normal density (The proof is left as an exercise). This equation states that if $Z \sim \mathrm{Normal}(0, 1)$, then

$$\mathbb{P}(Z \leq -x) = \mathbb{P}(Z > x) \quad \text{for } -\infty < x < +\infty.$$

See the figure in Page 52 for illustration.

(2) Let $Y = aX + b$, with two non-null constant $a, b \in \mathbb{R}$. If $X \sim \mathrm{Normal}(\mu, \sigma^2)$, then

$$Y \sim \mathrm{Normal}(a\mu + b, a^2\sigma^2).$$

To prove this statement, suppose that $a > 0$. (The proof when $a < 0$ is similar.) Let F_Y denote the c.d.f of Y. Then

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \le x) \\ &= \mathbb{P}(aX + b \le x) \\ &= \mathbb{P}(X \le \frac{x-b}{a}) \\ &= F_X(\frac{x-b}{a}), \end{aligned}$$

where F_X is the c.d.f. of X. By differentiation, the p.d.f of Y is then

$$\begin{aligned} f_Y(x) = \frac{1}{a} f_X(\frac{x-b}{a}) &= \frac{1}{\sqrt{2\pi} a\sigma} \exp\left\{ -\frac{(\frac{x-b}{a} - \mu)^2}{2\sigma^2} \right\} \\ &= \frac{1}{\sqrt{2\pi} a\sigma} \exp\left\{ -\frac{[x - (a\mu + b)]^2}{2(a\sigma)^2} \right\}, \end{aligned}$$

which shows that Y is normal with parameters $a\mu + b$ and $a^2\sigma^2$ from the definition.

(3) An important implication of (2) is that if $X \sim \text{Normal}(\mu, \sigma^2)$, then $Z := (X - \mu)/\sigma \sim \text{Normal}(0, 1)$.
Moreover, the c.d.f of X at any point a can be expressed as

$$F_X(a) := \mathbb{P}(X \le a) = \mathbb{P}(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}) = \mathbb{P}(Z \le \frac{a-\mu}{\sigma}) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Example 3.6.3. If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find (a) $\mathbb{P}(2 < X < 5)$; (b) $\mathbb{P}(X > 0)$; (c) $\mathbb{P}(|X - 3| > 6)$.

Solution: Let $Z := \dfrac{X-3}{3}$.

(a)

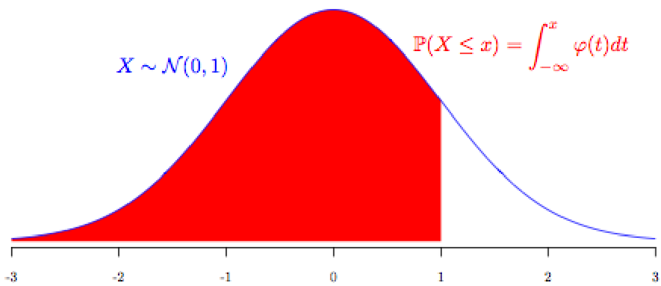
$$\begin{aligned} \mathbb{P}(2 < X < 5) &= \mathbb{P}\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) \\ &= \mathbb{P}\left(-\frac{1}{3} < Z < \frac{2}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right), \quad \text{since } Z \sim \text{Normal}(0,1) \\ &= \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right] \approx 0.3779. \end{aligned}$$

(b)

$$\begin{aligned} \mathbb{P}(X > 0) = \mathbb{P}\left(\frac{X-3}{3} > \frac{0-3}{3}\right) &= \mathbb{P}(Z > -1) \\ &= 1 - \mathbb{P}(Z \le -1) \\ &= 1 - \Phi(-1), \quad \text{since } Z \sim \text{Normal}(0,1) \\ &= 1 - [1 - \Phi(1)] \\ &= \Phi(1) \approx 0.8413. \end{aligned}$$

(c)

$$\begin{aligned}
\mathbb{P}(|X-3|>6) &= \mathbb{P}(X-3<-6 \text{ or } X-3>6)\\
&= \mathbb{P}(X>9)+\mathbb{P}(X<-3)\\
&= \mathbb{P}\left(\frac{X-3}{3}>\frac{9-3}{3}\right)+\mathbb{P}\left(\frac{X-3}{3}<\frac{-3-3}{3}\right)\\
&= \mathbb{P}(Z>2)+\mathbb{P}(Z<-2)\\
&= 1-\Phi(2)+\Phi(-2), \quad \text{since } Z\sim \text{Normal}(0,1)\\
&= 1-\Phi(2)+1-\Phi(2), \quad \text{since } \Phi(-2)=1-\Phi(2)\\
&= 2[1-\Phi(2)]\\
&\approx 0.0456.
\end{aligned}$$



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Figure 3.3: Area $\Phi(x)$ under the standard normal curve to the left of x.

EXERCISES

3.1. Consider a bag containing five balls, with the balls numbered 1 through to 5. Three balls are drawn from the bag without replacement.

(a) What is the probability the balls numbered 1, 2 and 3 are drawn from the bag, in the order 1, 2, 3?

(b) What is the probability the balls numbered 1, 2 and 3 are drawn in any order?

(c) Complete the table below from the probabilities of the sum of the 3 balls drawn from the bag.

sum of balls	6	7	8	9	10	11	12
probability							$\frac{1}{10}$

3.2. Consider two fair dice A and B. Die A is 6-sided and is numbered 1 through to 6 whilst die B is 4-sided and is numbered 1 though to 4. Both dice are rolled. Let $X = A + B$. Find the probability mass function and probability distribution function of X.

3.3. The random variable Z has the probability mass function below.

Z	0	1	2	3
$f_Z(z)$	0.2	0.16	0.41	a

(a) What is the value of a?

(b) What is $\mathbb{P}(1 \leq Z < 3)$?

(c) What is $F_Z(1.7)$?

3.4. The probability mass function of random variable Y is

$$f_Y(y) := \begin{cases} \frac{ky}{y^2+1}, & \text{if } y \in \{2,3\}; \\ \frac{ky}{2(y^2-1)}, & \text{if } y \in \{4,5\}; \\ 0, & \text{otherwise.} \end{cases}$$

Find the value of the constant k.

3.5. Let X be a discrete random variable and

$$f(x) := \begin{cases} C(4 - |x-4|), & \text{if } x \in \{1,2,3,4,5,6,7\}, \\ 0, & \text{otherwise,} \end{cases}$$

where C is a real-valued constant.

(a) Find C such that f is a probability mass function.

(b) Assuming that f is the probability mass function of X, calculate the mean, variance, and standard deviation of X. Round your answers to the 3rd decimal place if necessary.

(c) Determine the probability

$$P(1 < X \leq 5).$$

3.6. A component of a particular assessment is a multiple choice test consisting of two problems. The first has three possible answers, the second has five. The student did not revise much before the test, and chooses at random one answer from each of the problems independently. What is the expected number of correct answers? Is random answering a good strategy?

3.7. The probability that Ms. Brown will sell a piece of property at a profit of \$3,000 is 3/20, the probability that she will sell it at a profit of \$1,500 is 7/20, the probability that she will break even is 7/20, and the probability that she will lose \$1,500 is 3/20. What is her expected profit?

3.8. Assume that a game of chance is considered fair, or equitable, if each player's expectation is equal to zero. If someone pays us \$10 each time that we roll a 3 or a 4 with a balanced die, how much should we pay that person when we roll a 1, 2, 5 or 6 to make the game equitable?

3.9. The probability that a fluorescent bulb burns for at least 500 hours is 0.90. Of 8 such bulbs, find the probability that

(a) all 8 burn for at least 500 hours,

(b) 7 burn for at least 500 hours.

3.10. Samuel Pepys being about to place a bet asked Newton which of the following events is more likely

A. at least one six when 6 dice are thrown

B. at least a double six when 12 dice are thrown

C. at least a triple six when 18 dice are thrown.

Newton gave the correct answer. What is that?

3.11. The following gambling game, known as the wheel of fortune (or chuk-a-luck), is quite popular at many carnivals and gambling casinos: A player bets on one of the numbers 1 through 6. Three dice are then rolled, and if the number bet by the player appears i times, $i = 1, 2, 3$, then the player wins i unites; on the other hand, if the number bet by the player does not appear on any of the dice, then the player loses 1 unit. Is this game fair to the player? (Actually the game is played by spinning a wheel that comes to rest on a slot labeled by three of the numbers 1 through 6, but it is mathematically equivalent to the dice version.)

3.12. Suppose that a particular trait (such as eye color or left handedness) of a person is classified on the basis of one pair of genes and suppose that d represents a dominant gene and r a recessive gene. Thus a person with dd genes is pure dominance, one with rr is pure recessive, and one with rd is hybrid. The pure dominance and the hybrid are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total 4 children, what is the probability that 3 out of 4 children have the outward appearance of the dominant gene?

3.13. (See also Theorem 3.2.3, (2).) Let X and Y be two discrete random variables, prove that if X and Y are independent, then for any constants a, $b \in \mathbb{R}$,

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y).$$

Hint: Two random variables X and Y are said to be independent if

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

3.14. (See also Exercise 3.3.1) Prove that if X follows Geometric distribution with parameter p, then

$$\mathbb{E}[X] = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

3.15. The number of trucks X arriving on any one day at a truck depot in a certain city is a Poisson distributed r.v. Suppose that the average number of X is 12. What is the probability that on a given day fewer than nine trucks will arrive at this depot?

3.16. Let $N(t)$ denote the number of earthquakes during the time interval $(0, t]$, for $t > 0$, in the western portion of the United States. It is known that the $N(t)$ is a Poisson r.v. with parameter λt, for any $t > 0$. Let t be of 1 week as the unit of time and $\lambda = 2$.

(a) Find the probability that at least 3 earthquakes occur during the next 2 weeks.

(b) Find the probability distribution of the time, starting from now, until the next earthquake.

3.17. The lifetime of a component $\mathcal{C}$ is an random variable X with the probability density function (p.d.f) given by

$$f(t) = \begin{cases} k(2t - t^2) & \text{if } 0 \le t < 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a real valued constant.

(a) Determine the value k.

(b) Determine the cumulative distribution function (c.d.f) $F(x)$ of X.

(c) Determine the probability $\mathbb{P}(X \le 1)$.

(d) Determine the probability $\mathbb{P}(X \leq \frac{1}{2})$.

(e) Determine the probability $\mathbb{P}(X \geq \frac{3}{2})$.

(f) Determine the conditional probability $\mathbb{P}(X \geq \dfrac{3}{2}|X \geq 1)$.

3.18. The lifetime of an insect is an random variable X with c.d.f $F(x)$ given by

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - (x+1)e^{-x} & \text{if } x > 0. \end{cases}$$

(a) Determine, rounding to the second decimal place, the probability $\mathbb{P}(X \leq 1)$.

(b) Prove that the random variable X admits an p.d.f $f(t)$ and determine it.

3.19. The lifetime of an organism is an random variable X with p.d.f

$$f(t) = \begin{cases} k(10t^2 - t^3) & \text{if } 0 \leq t \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

where k is a real valued constant.

(a) Determine k.

(b) Determine $F(x)$, the c.d.f of X.

(c) Calculate the probability $\mathbb{P}(X \geq 1)$.

(d) Calculate the probability $\mathbb{P}(1 \leq X \leq 2)$.

(e) Calculate the conditional probability $\mathbb{P}(1 \leq X \leq 2|X \geq 1)$, rounding your answer to the fourth decimal place.

(f) Determine $\mathbb{E}(X)$.

(g) Determine $\mathbb{E}\left(X^2\right)$.

(h) Determine the variance $\mathrm{Var}(X)$ of X.

3.20. The concentration of an urinary metabolite is an random variable X with p.d.f defined by

$$f(t) = \begin{cases} k(1-t) & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Draw the curve of the function f and calculate k.

(b) Determine the expression of the c.d.f $F(x)$ of the random variable X by the position of x in comparison to 0 and 1. Draw the curve of the function F.

(c) Evaluate $\mathbb{P}(0 < X < 1)$.

(d) Calculate $\mathbb{E}(X)$.

(e) Calculate $\mathrm{Var}(X)$.

Let Y be another random variable given by

$$Y = 1 - X.$$

(f) Calculate $\mathbb{E}(Y)$.

(g) Calculate $\mathrm{Var}(Y)$.

(h) Calculate $\mathbb{P}\left(\frac{1}{3} \leq Y \leq \frac{2}{3}\right)$.

3.21. Suppose that X is a uniformly distributed random variable over the interval $(-10, 10)$. The functions $f(t)$ and $F(x)$ are respectively the p.d.f and c.d.f of X. Determine $f(t)$, $F(x)$, $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

3.22. The lifetime of a radioactive atom is an exponential random variable with parameter $\theta = 1/2$.

(a) Determine the expectation $\mathbb{E}(X)$ of X.

(b) Determine the variance $\mathrm{Var}(X)$ of X.

(c) For $x \geq 0$, what is the expression of the c.d.f, $F(x)$, of X?

(d) Determine the probability $\mathbb{P}(X \geq 1)$, rounding your answer to the fourth decimal place.

(e) Determine the probability $\mathbb{P}(1 \leq X \leq 2)$, rounding your answer to the fourth decimal place.

(f) Determine the conditional probability $\mathbb{P}(X \leq 2 | X \geq 1)$, rounding your answer to the fourth decimal place.

3.23. At a particular point of the electric grid the time of a voltage drop is an exponential random variable. The mean time elapsed between consecutive voltage drops is 20 days.

(a) What is the probability that there is at least a 30 days gap between consecutive voltage drops?

(b) What is the probability that within 30 days at least two voltage drops occur?

3.24. A loom stops automatically as soon as a thread breaks. Let X be the time elapsed from the start of the weaving process to the first stop caused by this failure. Assuming that X is of exponential distribution with a mean of 2.5 hours, what is the probability that on an average 8 hours workday the machine does not fail?

3.25. An component has lifetime X, which is an exponential random variable with parameter $\theta = 250$.

(a) Determine $\mathbb{E}(X)$.

(b) Determine $\mathrm{Var}(X)$.

(c) Evaluate the probability that the lifetime of the component is > 100, rounding your answer to the second decimal place.

(d) Determine the conditional probability $\mathbb{P}(X > 200|X > 100)$, rounding your answer to the second decimal place.

(e) Determine the smallest integer n such that $\mathbb{P}(X > n) \leq 0.05$.

3.26. The weight of an egg in a farm follows normal distribution with mean 65 and standard deviation 10, denoted by $\mathrm{Normal}(65, 100)$.

(a) Determine the probability $\mathbb{P}(X < 50)$.

(b) Determine the probability $\mathbb{P}(X > 80)$.

(c) Determine the probability

$$\mathbb{P}(X > 80|X \geq 65).$$

Hint: Use the table in Figure 4.3.

3.27. In certain population, the rate of glucose in the body of an individual is a random variable X, which follows a normal distribution with mean m (g/l) and standard deviation σ (g/l), denoted by $\mathrm{Normal}(m, \sigma^2)$.

It follows from a study performed on a large number of independent people that

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X \geq 0.8) = 0.84.$$

(a) Determine m and σ, rounding your answers to the second decimal place.

In the following question, use the values m and σ that you obtained in Part (a).

(b) Determine the probability $\mathbb{P}(X \geq 1.1)$.

Hint: Use the table in Figure 3.3.

Bibliography

[1] Sheldon M. Ross. *A first course in probability*, Pearson, 2010, 8th edition

[2] Geoffery Grimmett. David Sttirzaker *Probability and random processes*, Oxford University Press, 2001, 3rd edition

[3] Sheldon M. Ross. *Introduction to probability and statistics for engineers and scientists*, Elsevier Academic Press, 2009, 4th edition