Setif 1 University-Ferhat ABBAS
Faculty of Sciences
Department of mathematics

Advanced Probabilities

Conforming to the first-year Master of Computer Science – Quantum Computing Curriculum (2024/2025)

Dr. Raouf ZIADI

Contents

Introduction

Probability theory forms a fundamental pillar of modern science and engineering, offering powerful tools to analyze uncertainty, model random phenomena, and make informed decisions under incomplete information. From elementary counting techniques to advanced probabilistic models, the study of probability provides a rigorous framework for reasoning about chance, variability, and complex systems.

In the field of quantum computing, probability plays a particularly crucial role. Unlike classical computing, where outcomes are deterministic, quantum systems evolve in terms of probability amplitudes, and measurement outcomes are inherently probabilistic. A solid understanding of probability theory is therefore essential for interpreting quantum algorithms, analyzing quantum states, and modelling phenomena such as superposition, entanglement, and decoherence.

This booklet has been designed for first-year Master's students in Computer Science, option: Quantum Computing. By blending theoretical foundations with illustrative examples, it aims to strengthen both intuition and analytical rigor. The material not only introduces students to key probabilistic tools but also prepares them to apply these concepts directly within the framework of quantum information theory, quantum algorithms, and related computational paradigms.

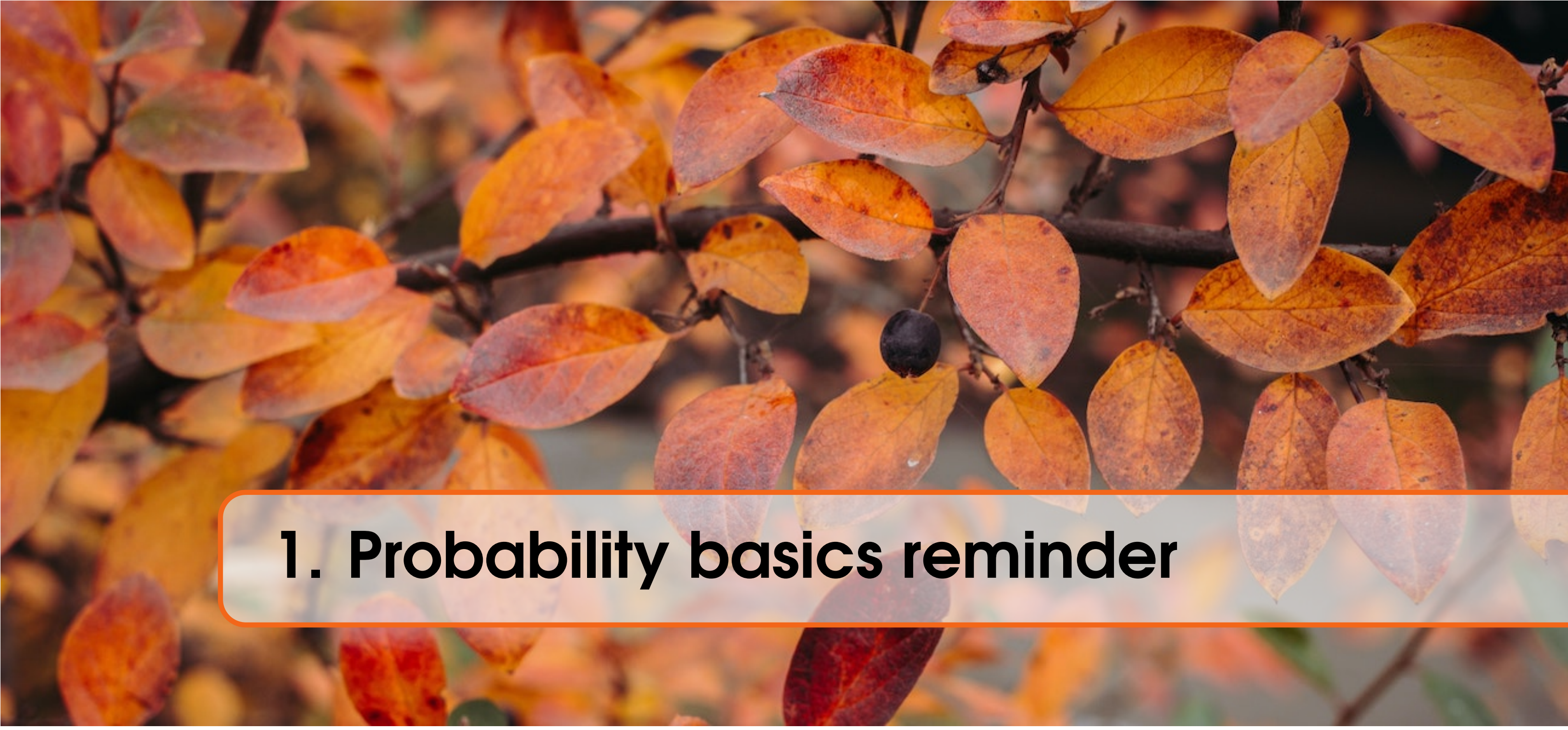
This booklet is intended as a concise yet comprehensive introduction to probability and random variables. Each section is accompanied by illustrative examples, and at the end of each chapter, students are encouraged to work on additional exercises to reinforce theoretical concepts and enhance problem-solving skills.

The booklet is organized into four main chapters. The first chapter introduces the fundamental principles of probability, including counting techniques, sample spaces, and the concept of independence. Conditional probability and Bayes' theorem are presented early on to provide a solid foundation for probabilistic reasoning.

Building on these basics, the second chapter explores discrete and continuous random variables, covering key concepts such as probability distributions, expected value, variance, and standard deviation. Special attention is given to widely used distributions, including the Bernoulli, Binomial, Poisson, Normal, Exponential, and Gamma distributions, which play a central role in statistics, data science, and various applied fields.

The third chapter addresses probability distributions of combined random variables. Topics such as joint distributions, conditional expectations, and their applications are detailed to enable learners to analyze more complex systems involving multiple sources of randomness.

Finally, the last chapter focuses on conditional probabilities and independence, emphasizing the role of conditional distributions, their properties, and their importance in both theoretical and applied contexts.



1. Probability basics reminder

1.1 Techniques of counting

In this section we present some techniques for determining without direct enumeration the number of possible outcomes of a particular experiment or the number of elements in a particular set. Such techniques are sometimes referred to as combinatorial analysis. We begin with the following basic principle.

Factorial notation

The product of the positive integers from 1 to n inclusive occurs very often in mathematics and hence is denoted by the special symbol $n!$ (read "n factorial"):

$$n! = 1.2.3.\ldots,.(n-2)(n-3)n.$$

It is also convenient to define $0! = 1! = 1$.

1.1.1 Permutations

Permutations without repetitions:

An arrangement of a set of n objects in a given order is called a permutation of the objects (taken all at a time). An arrangement of any $r \leq n$ of these objects in a given order is called an r-permutation or a permutation of the n objects taken r at a time. The number of permutations of n objects taken r at a time will be denoted by $P(n,r)$

Theorem 1.1 The number of permutations of r objects selected from a set of n distinct objects is

$$P(n,r) = n(n-1)(n-2)\dots(n-r+1),$$

or, in factorial notation,

$$P(n,r) = \frac{n!}{(n-r)!}..$$

In the case where all the objects are taken at a time, then

$$P(n,n) = n(n-1)(n-2)\dots,1 = n!.$$

■ **Example 1.1** An electronic controlling mechanism requires 5 distinct, but interchangeable, memory chips. In how many ways can this mechanism be assembled:
(a) by placing the 5 chips in the 5 positions within the controller?
(b) by placing 3 chips in the odd numbered positions within the controller?
Solution:
(a) When all 5 chips must be placed, the answer is 5!. Alternatively, in the permutation notation with $n = 5$ and $r = 5$, the first formula yields

$$P(5,5) = 5! = 5.4.3.2.1 = 120$$

(b)For $n = 5$ chips placed in $r = 3$ positions, the permutation is

$$P(5,3) = \frac{5!}{2!} = 5.4.3 = 60.$$

■ **Example 1.2** You are going on a road trip with 4 friends in a car that fits 5 people. How many different ways can everyone sit if you have to drive the whole way?
Solution:
You have to sit in the driver's seat. Hence, there are 4 options for the 1st passenger seat. Once that person is seated, there are 3 options for the next passenger seat. This goes on until there is one person left with 1 seat.

$$1.P(4,4) = 1.4! = 24.$$

Permutations with repetitions:

Frequently we want to know the number of permutations of objects some of which are alike, as illustrated below. The general formula follows.

Theorem 1.2 The number of permutations of n objects of which n_l are alike, n_2 are alike,...,n_k are alike is

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

■ **Example 1.3** How many different signals, each consisting of 8 flags hung in a vertical line, can be formed from a set of 4 indistinguishable red flags, 3 indistinguishable white flags, and a blue flag?

Solution:

We seek the number of permutations of 8 objects of which 4 are alike (the red flags) and 3 are alike (the white flags). By the above theorem, there are $\frac{8!}{4!3!} = 280$ different signals.

1.1.2 Combinations

There are many problems in which we must find the number of ways in which r objects can be selected from a set of n objects, **but we do not care about the order** in which the selection is made and where these objects are taken at a time. Therefore, to find the number of ways in which r objects can be selected from a set of n distinct objects, also called the number of combinations of n objects taken r at a time and denoted by $C(n,r)$.

Theorem 1.3 The number of ways in which r objects can be selected from a set of n distinct objects is

$$C(n,r) = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!},$$

or, in factorial notation,

$$C(n,r) = \frac{P(n,r)}{r!} = \frac{n!}{r!(n-r)!}.$$

Note that $C(n,n) = 1$ and $C(n,0) = 1$ since there is only one way to choose a set of (all) n elements or no elements, and $C(n,1) = n$ since there are n subsets of size 1.

■ **Example 1.4** There are 20 hockey players on a pro NHL team, 2 of whom are goalies. How many sets of 5 skaters and 1 goalie can be on the ice at the same time?
Solution:
The question asks for how many on the ice, implying that order does not matter. This is combination problem with 2 combinations. You need to choose 1 goalie out of a possible of 2, and choose 5 skaters out of a possible 18.

$$C(2,1)C(18,5) = 2.\frac{18!}{5!13!}$$

■ **Example 1.5** How many different ways could you score a 70% on a 10-question test, where each question is weighted equally and is either right or wrong?
Solution:
The order of the questions you got right does not matter, so this is a combination problem,

$$C(10,7) = \frac{10!}{7!3!} = 120.$$

Ordered partitions

Theorem 1.4 Let A contain n elements and let $n_1, n_2, \ldots, n_r$ be positive integers with $n_1 + n_2 + \cdots + n_r = n$. Then there exist $\frac{n!}{n_1!n_2!...n_r!}$ different ordered partitions of A of the form $(A_1, A_2, ..., A_r)$ where A_1 contains n_1 elements, A_2 contains n_2 elements,..., and A_r contains n_r elements.

■ **Example 1.6** In how many ways can 9 toys be divided between 4 children if the youngest child is to receive 3 toys and each of the other children 2 toys?
Solution:
We wish to find the number of ordered partitions of the 9 toys into 4 cells containing 3,2, 2 and 2 toys respectively. By the above theorem, there are

$$C(9,3).C(6,2).C(4,2).C(2,2) = \frac{9!}{3!2!2!2!} = 7560$$

such ordered partitions.

1.2 Sample spaces, events, and their probabilities

Sample spaces

A **random experiment** is a mechanism that produces a definite outcome that cannot be predicted with certainty. The **sample space** S, is the set of all possible outcomes of a random experiment. Each outcome in a sample space is called an element or a member of the sample space.

For example, there are only two outcomes for tossing a coin, and the sample space is

$$S = \{Heads, Tails\} = \{H, T\}.$$

If we toss a coin three times, then the sample space is

$$S = \{HHH, HHT, HTH, THH, HTT, TTH, THT, TTT\}.$$

■ **Example 1.7** Toss a coin until a head appears and then count the number of times the coin was tossed. The sample space of this experiment is $S = \{1, 2, 3, \ldots, \infty\}$. Here ∞ refers to the case when a head never appears and so the coin is tossed an infinite number of times. This is an example of a sample space which is countably infinite.

■ **Example 1.8** Consider rolling a fair die twice and observing the dots facing up on each roll. There are 36 possible outcomes in the sample space S, where

$$S = \{(1,1), (1,2), \ldots, (1,6), (2,1)(2,2), \ldots, (2,6), \ldots, (6,1) \ldots, (6,6)\}$$

Events

An event is a subset of a sample space S. An event A is said to occur on a particular trial of the experiment if the outcome observed is an element of the set A.

■ **Example 1.9** 1-Construct a sample space for the experiment that consists of rolling a single die.
2-Find the events that correspond to the phrases A: "an even number is rolled" and B:"a number greater than two is rolled."
Solution:
1- The outcomes could be labelled according to the number of dots on the top face of the die. Then the sample space is the set $S = \{1, 2, 3, 4, 5, 6\}$.

2- The outcomes that corresponds to the event A is the set $\{2,4,6\}$, so we set $A = \{2,4,6\}$. Similarly the outcomes that corresponds to event B is the set $\{3,4,5,6\}$, and we set $B = \{3,4,5,6\}$.

1.2.1 Probability of an event

Probability is a real-valued function P that assigns to each event A in a sample space S a number called the probability of the event A, denoted by $P(A)$, such that the following three properties are satisfied:

A1 - For every event A, $0 \leq P \leq 1$,
A2 - $P(S) = 1$,
A3 - If A and B are disjoint events(i.e, $A \cap B = \phi$), then

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

A4 - If $A_1, A_2, A_3, \ldots$ is a sequence of disjoint events, then

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

Based on the above axioms, we can derive many other results which play important roles in applications:

Theorem 1.5 I- If ϕ is the empty set, then $P(\phi) = 0$.
II- If A is any event in S, then $P(A^c) = 1 - P(A)$, A^c is the complement of an event A.
III- If A and B are any two events, then $P(A - B) = P(A) - P(A \cap B)$.
IV- If $A \subset B$, then $P(A) \leq P(B)$.
V- If A and B are any events in S, then: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof.

I- Since we have S and ϕ are disjoint events, then from A2 and A3 $P(S) = P(S \cup \phi) = P(S) + P(\phi) = 1$. Then $P(\phi) = 0$.

II- Since $S = A \cup A^c$, and A and A^c are disjoints, then from A3 it results that : $P(S) = P(A) + P(A^c) = 1$. Hence, $P(A^c) = 1 - P(A)$.

III- The event A can be decomposed into the mutually disjoint events $A - B$ and $A \cap B$; that is, $A = (A - B) \cup (A \cap B)$. Thus by A3, $P(A) = P(A - B) + P(A \cap B)$, from which our result follows.

IV- Since $B = A \cup (B \cap A^c)$, then from A3, $P(B) = P(A) + P(B \cap A^c) \geq P(A)$.

V- Note that the event $A \cup B$ can be decomposed into disjoint events $A - B$ and B;that is, $A \cup B = (A - B) \cup B$. Thus by A3, $P(A \cup B) = P(A - B) + P(B) = P(A) - P(A \cap B) + P(B)$, which is the desired result. ■

1.2.2 Independence of events

Definition 1.1 Two events A and B are independent if and only if their joint probability equals the product of their probabilities

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

$A \cap B \neq \phi$ indicates that two independent events A and B have common elements in their sample space so that they are not mutually exclusive (disjoint).

■ **Example 1.10** The probability that Tom forgets his PE kit is $0,3$. The probability that Noah forgets his PE kit is $0,1$. The events are independent. Calculate the probability that both Tom and Noah forget their PE kits on the same day.

Solution

The first event is that Tom forgets his PE kit. The probability that Tom forgets his PE kit is $0,3$. The second event is that Noah forgets his PE kit. The probability that Noah forgets his PE kit is $0,1$. The probability that Tom and Noah both forget their PE kits is $0,3 \times 0,1 = 0,03$.

■ **Example 1.11** Rachel tosses three fair coins. Find the probability that all three coins land on tails.

Solution

The outcome for each coin is not affected by the other coins, therefore the events are independent. Since the coins are fair, for each coin the probability that it lands on tails is $0,5$. Hence, the probability of getting three tails is $0,5^3 = 0,125$.

1.3 Conditional probability and Bayes' theorem

Conditional probability

The conditional probability of an event B is the probability that the event will occur given the knowledge that an event A has already occurred i.e., $P(A) > 0$. This probability is written

$P(B|A)$: the probability of B given A, is defined as follows:

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

- In the case where events A and B are independent, the conditional probability of event B given event A is simply the probability of event B, that is $P(B)$.
- If events A and B are dependent, then the probability of the intersection of A and B (the probability that both events occur) is defined by $P(A \cap B) = P(A)P(B|A)$.

■ **Example 1.12** You toss a fair coin three times. Given that you have observed at least one head, what is the probability that you observe at least two heads?

Solution

Let $A1$ be the event that you observe at least one head, and $A2$ be the event that you observe at least two heads. Then

$$A_1 = S - \{TTT\}, \text{ and } P(A_1) = \frac{7}{8};$$

$$A_2 = \{HHT, HTH, THH, HHH\}, \text{ and } P(A_2) = \frac{4}{8}.$$

Thus, we can write

$$\begin{aligned} P(A_2|A_1) &= \frac{P(A_2 \cap A_1)}{P(A_1)} \\ &= \frac{P(A_2)}{P(A_1)} \\ &= \frac{4}{8} . \frac{8}{7} = \frac{4}{7} \end{aligned}$$

1.3.1 Bayes' Rule

Now we are ready to state one of the most useful results in conditional probability: Bayes' rule. Suppose that we know $P(A|B)$, but we are interested in the probability $P(B|A)$. Using the definition of conditional probability, we have

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

Dividing by $P(A)$, we obtain

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

which is the famous Bayes' rule. Often, in order to find $P(A)$ in Bayes' formula we need to use the law of total probability, so sometimes Bayes' rule is stated as

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)},$$

where $B_1, B_2, \cdots, B_n$ form a partition of the sample space.

Definition 1.2 (**Bayes' Rule**).

- For any two events A and B, where $P(A) \neq 0$, we have

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

- If $B_1, B_2, B_3, \cdots$ form a partition of the sample space S, and A is any event with $P(A) \neq 0$, we have

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

▪ **Example 1.13** Three machines A, B and C produce respectively 50%, 30% and 20% of the total number of items of a factory. The percentages of defective output of these machines are 3%, 4% and 5%. If an item is selected at random, find the probability that the item is defective.
Solution

Let D be the event that an item is defective. Then using Bayes' formula and as illustrated in Figure 2.1 we get,

$$\begin{aligned} P(D) &= P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) \\ &= (0.5)(0.03) + (0.3)(0.04) + (0.2)(0.05) = 0.037 \end{aligned}$$

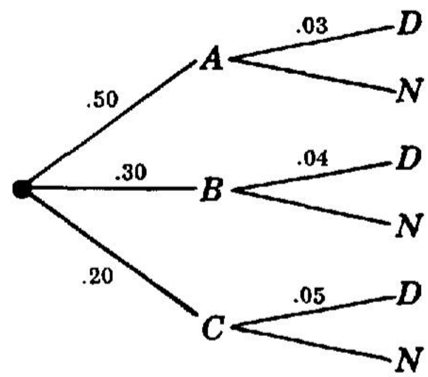


Figure 1.1: Diagram of Example 3.1.

■ **Example 1.14** Consider the factory in the preceding example. Suppose an item is selected at random and is found to be defective. Find the probability that the item was produced by machine A; that is, find $P(A|D)$.

Solution

By Bayes' theorem,

$$P(A|D) = \frac{P(A)P(D|A)}{P(A)P(D|A)+P(B)P(D|B)+P(C)P(D|C)}$$
$$= \frac{(0.5)(0.03)}{(0.5)(0.03)+(0.3)(0.04)+(0.2)(0.05)} = \frac{15}{37}$$

1.4 Solved exercises

Exercise 1.1 Three light bulbs are chosen at random from 15 bulbs where 5 of them are defective. Find the probability that :

(i) none is defective

(ii) exactly one is defective

(iii) at least one is defective.

Solution:

(i) $p = \frac{C(10,3)}{C(15,5)} = \frac{120}{455} = \frac{24}{91}$.

(ii)$p = \frac{C(5,1)C(10,2)}{C(15,5)} = \frac{225}{455} = \frac{45}{91}$

(iii)$p = 1 - \frac{24}{91}$.

Exercise 1.2 A class contains 10 men and 20 women of which half the men and half the women have brown eyes. Find the probability that a person chosen randomly has brown eyes or is a man.

Solution

Let A be the event that the chosen person is a man, and B the event that the chosen person has brown eyes. Then

$$P(A) = \frac{10}{30} = 1/3$$
$$P(B) = \frac{15}{30} = 1/2$$
$$P(A \cap B) = \frac{5}{30} = \frac{1}{6}$$

Thus $p = P(A) + P(B) - P(A \cap B) = 2/3$.

Exercise 1.3 A class has 12 boys and 4 girls. If three students are selected at random from the class, what is the probability that they are all boys?

Solution

There exist $C(16,3)$ ways of selecting three students, and $C(12,3)$ ways of selecting three boys, then

$$p = \frac{C(16,3)}{C(12,3)} = \frac{11}{28}$$

Exercise 1.4 Ten numbered cards are there from 1 to 15, and two cards a chosen at random such that the sum of the numbers on both the cards is even. Find the probability that the chosen cards are odd-numbered.

Solution:

Let A be the event of selecting two odd-numbered cards and B the event of selecting cards whose sum is even. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$= \frac{C(8,2)}{C(8,2) + C(7,2)} = 4/7.$$

Exercise 1.5 The probability of a student passing in science is $4/5$ and the of the student passing in both science and maths is $1/2$. What is the probability of that student passing in maths knowing that he passed in science?

Solution:

Let A be the event of passing in science and B the event of passing in maths. Then, probability of passing maths after passing in science is

$$\begin{aligned} P(B|A) &= \frac{P(A\cap B)}{P(A)} \\ &= \frac{1/2}{4/5} = \frac{5}{8} \end{aligned}$$

Exercise 1.6 Three persons A, B and C have applied for a job in a private company. The chance of their selections is in the ratio $1,2,4$. The probabilities that A, B and C can introduce changes to improve the profits of the company are $0.8, 0.5$ and 0.3, respectively. If the change does not take place, find the probability that it is due to the appointment of C.

Solution

Let E1, E2 and E3 respectively the events :"person A get selected", "person B get selected" and "person C get selected". Let also A the event "Changes introduced but profit not happened". Then, $P(E1) = 1/(1+2+4) = 1/7, P(E2) = 2/7$ and $P(E3) = 4/7$. We have also,

$$\begin{aligned} P(A|E1) &= 1-0.8 = 0.2 \\ P(A|E2) &= 1-0.5 = 0.5 \\ P(A|E3) &= 1-0.3 = 0.7 \end{aligned}$$

Applying Bays' formula it follows that

$$\begin{aligned} P(E3|A) &= \frac{P(A|E3)P(E3)}{P(A|E1)P(E1)+P(A|E2)P(E2)+P(A|E3)P(E3)} \\ &= \frac{0.7\frac{4}{7}}{0.2\frac{1}{7}+0.5\frac{2}{7}+0.7\frac{4}{7}} = \frac{7}{10} \end{aligned}$$

Exercise 1.7 A doctor is called to see a sick child. The doctor has prior information that 90% of sick children in that neighborhood have the flu, while the other 10% are sick with measles. Let F stand for an event of a child being sick with flu and M stand for an event of a child being sick with measles. Assume for simplicity that $F \cup M = S$, i.e., that there no other maladies in that neighbourhood.

A well-known symptom of measles is a rash (the event of having which we denote R). Assume that the probability of having a rash if one has measles is $P(R|M) = 0.95$. However, occasionally children with flu also develop rash, and the probability of having a rash if one has flu is $P(R|F) = 0.08$. Upon examining the child, the doctor finds a rash. What is the probability that the child has measles?

Solution

Using Bayes' formula we get

$$P(M|R) = \frac{P(R|M)P(M)}{P(R|M)P(M) + P(R|F)P(F)} \qquad = \frac{0.95 \times 0.10}{0.95 \times 0.10 + 0.08 \times 0.90} = 0.568$$

Exercise 1.8 Suppose we have 3 cards identical in form except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side is colored black. The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Solution

Let RR, BB, and RB denote, respectively, the events that the chosen cars is the red-red, the black-black, or the red-black card. Letting R be the event that the upturned side of the chosen card is red, we have that the desired probability is obtained by

$$\begin{aligned} P(RB|R) &= \frac{P(RB \cap R)}{P(R)} \\ &= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)} \\ &= \frac{1/21/3}{1 \times 1/3 + 1/2 \times 1/3 + 0 \times 1/3} = 1/3 \end{aligned}$$

Exercise 1.9 It is estimated that50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%.

Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

Solution

Let A the event that an email is detected as spam, and B the event that an email is spam. Hence by the Bayes's formula we have

$$\begin{aligned} P(B^c|A) &= \frac{P(B^c \cap A)}{P(A)} \\ &= \frac{P(A|B^c)P(B^c)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \\ &= \frac{0.05 \times 0.5}{0.05 \times 0.5 + 0.99 \times 0.5} = 5/104. \end{aligned}$$

Exercise 1.10 In a study, physicians were asked what the odds of breast cancer would be in a woman who was initially thought to have a 1% risk of cancer but who ended up with a positive mammogram result (a mammogram accurately classifies about 80% of cancerous tumours and 90% of benign tumours.)
95% of physicians estimated the probability of cancer to be about 75%. Do you agree?
Solution
Let M be the event "tumour is malignant", B for the "tumour is benign" and M^+ for the event "mammogram result is positive". Hence using Bayes' formula we get:

$$P(M|M^+) = \frac{P(M^+|M)P(M)}{P(M^+|M)P(M) + P(M^+|B)P(B))} \tag{1.1}$$

$$= \frac{0.80 \times 0.01}{0.80 \times 0.01 + 0.10 \times 0.99} = 0.075. \tag{1.2}$$

So the chance would be 7.5%. A far cry from a common estimate of 75%.

2. Random Variables

A random variable is a numerical description of the outcome of a statistical experiment. A random variable that may assume only a finite number or an infinite sequence of values is said to be discrete; one that may assume any value in some interval on the real number line is said to be continuous. For instance, a random variable representing the number of auto-mobiles sold at a particular dealership on one day would be discrete, while a random variable representing the weight of a person in kilograms would be continuous.

Definition 2.1 A random variable is a function which maps from sample space of an experiment S to the real numbers. Mathematically, Random Variable is expressed as,

$$X : S \to \mathbb{R}$$

■ **Example 2.1** The sample space of this experiments is $S = \{TT, HT, TH, HH\}$. If Y is the random variable that refers to the number of heads from tossing two coins, then Y takes the values $\{0, 1, 2\}$. This means that we could have no heads, one head, or both heads on a two-coin toss.

■ **Example 2.2** A non-biased coin is tossed three times. The sample space is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let X be the random variable which assigns to each point in S the largest number of successive heads which occurs. Then R.V. X takes the values $0,1,2$ and 3. Indeed,

$$X(TTT) = 0, X(HTH) = 1, X(HTT) = 1, X(THT) = 1,$$

$$X(TTH) = 1, X(HHT) = 2, X(THH) = 2, X(HHH) = 3.$$

2.1 Discrete random variables.

Definition 2.2 A random variable X is discrete if its support R_X is countable and there exists a function $p_X : \mathbb{R} \to [0,1]$, called Probability Mass Function (PMF) of X, such that

$$p_X(x) = P(X = x) = P(X^{-1}(x))$$

where $P(X = x)$ is the probability that X that takes the value x.

■ **Example 2.3** 1- The PMF that corresponds to the random variable Y of Example 2.1 is

$$p_Y(x) = 1/4 \text{ if } x \in \{0,1,2\}, \text{ and } p_Y(x) = 0 \text{ elswere.}$$

$$p_Y(x) = \begin{cases} 1/4, & \text{if } x \in \{0,2\} \\ 1/2, & \text{if } x = 1 \\ 0, & \text{otherwise.} \end{cases}$$

2- The PMF that corresponds to the random variable X of Example 2.2 is

$$p_X(x) = \begin{cases} 1/8, & \text{if } x \in \{0,3\} \\ 3/8, & \text{if } x \in \{1,2\} \\ 0, & \text{otherwise.} \end{cases}$$

2.1.1 Cumulative distribution function

In probability, the Cumulative Distribution Function (CDF) of a random variable X, is the probability that X takes a value less than or equal to x.

Definition 2.3 Let X be a random variable. The cumulative distribution function (CDF) or the distribution function of X is defined as

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

for any $x \in \mathbb{R}$. If X is discrete , then $F_X(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} p_X(x)$.

■ **Example 2.4** We toss a coin three times successively, and let X the number of obtained heads. The PMF that corresponds to the random variable X is

$$p_X(x) = \begin{cases} 1/8, \text{ if } x \in \{0,3\} \\ 3/8, \text{ if } x \in \{1,2\} \\ 0, \text{ otherwise.} \end{cases}$$

Therefore, by summing up the probabilities up to the value of x we get the following CDF,

$$F_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{8} & 0 \leq x < 1 \\ \frac{4}{8} & 1 \leq x < 2 \\ \frac{7}{8} & 2 \leq x < 3 \\ 1 & 3 \leq x < +\infty \end{cases}$$

Proposition 2.1 The CDF of a random variable X has the following properties:

1- F_X is right continuous and increasing, that is, if $x < y$, then $F(x) \leq F(y)$,

2- $\lim_{x \to +\infty} F_X(x) = 1$,

3- $\lim_{x \to -\infty} F_X(x) = 0$.

Independent random variables

The concept of independent random variables is very similar to independent events. Remember, two events A and B are independent if we have $P(A,B) = P(A)P(B)$(comma means "and", i.e., $P(A,B) = P(A \cap B)$). Similarly, we have the following definition for independent discrete random variables.

Definition 2.4 Consider n discrete random variables $X_1, X_2, X_3, ..., X_n$. We say that $X_1, X_2, X_3, ..., X_n$

are independent if

$$P\left(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right) = P(X_1 = x_1)P(X_2 = x_2)\ldots P(X_n = x_n).$$

In general, if n random variables are independent, then

$$P\left(X_1 \in A_1, X_2 \in A_2 \ldots X_n \in A_n\right) = P(X_1 \in A_1)P(X_2 \in A_2)\ldots P(X_n \in A_n).$$

■ **Example 2.5** I toss a coin twice and define X to be the number of heads I observe. Then, I toss the coin two more times and let Y be the number of heads that I observe in last two tosses. Find

$$P\left(X \leq 1, Y = 2\right).$$

Since X and Y are the result of different independent coin tosses, the two random variables X and Y are independent. We can write

$$\begin{aligned} P\left(X \leq 1, Y = 2\right) &= P(X \leq 1)P(Y = 2) \qquad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= \big(p_X(0) + p_X(1)\big)p_Y(2) \\ &= \left(\frac{1}{4} + \frac{1}{2}\right)\frac{1}{4} = 3/16 \end{aligned}$$

2.1.2 Mean and standard deviation of discrete random variables

The most important characteristics of any probability distribution are the mean, expected value, (or even average value) and the standard deviation.

2.1.2.1 The expected value of discrete random variables

The mean or expected value of a random variable X is written as $\mathbb{E}(x)$ or μ_x. If we observe n random values of X, then the mean of the n values will be approximately equal to $\mathbb{E}(x)$ for large n. The expected value is defined differently for continuous and discrete random variables.

Definition 2.5 Let X be a discrete random variable with PMF $p_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x)$$

The following example illustrates how to calculate the mean of a discrete random variable.

Example 2.6 Find the mean number of heads obtained in 3 flips of a balanced coin.

Solution

The PMF that corresponds to the r.v. X is

$$p_X(x) = \begin{cases} 1/8, \text{ if } x \in \{0,3\} \\ 3/8, \text{ if } x \in \{1,2\} \\ 0, \text{ otherwise.} \end{cases}$$

Therefore,

$$\mathbb{E}(X) = \sum_{x \in \{0,1,2,3\}} x p_X(x) = \frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2}$$

Remark 2.1 If g is a function of the random variable X then the expectation of $g(X)$ is given by

$$\mathbb{E}(g(X)) = \sum_{x \in R_X} g(X) p_X(x)$$

2.1.2.2 Properties of the expected value

Theorem 2.1 Let a be a constant, then $\mathbb{E}(a) = a$

Proof. From the mathematical definition of expectation we know that

$$\begin{aligned} \mathbb{E}(c) =& \sum_x x p_c(x) \\ =& c \times 1 = c \end{aligned}$$

■

Theorem 2.2 Let X and Y be random variables on the same sample space S. Then $E(X+Y) = E(X)+E(Y)$. More generally, if $X_1, X_2, \ldots, X_n$ be n random variables on the same sample space S. Then, $\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i)$.

Theorem 2.3 .
1- Let X be a random variable on the same sample space S and c is a constant. Then , $\mathbb{E}(cX) = c\mathbb{E}(X)$.
2- Let X be a random variables on the same sample space S, and a and c are two constants, then $\mathbb{E}(aX+c) = a\mathbb{E}(X)+c$.

Proof. 1-

$$\begin{aligned}\mathbb{E}(cX) = &\sum_{x\in R_X} cx p_X(x)\\ =&c\sum_{x\in R_X} x p_X(x) = c\mathbb{E}(X)\end{aligned}$$

2-

$$\begin{aligned}\mathbb{E}(aX+c) &= \mathbb{E}(aX)+\mathbb{E}(c)\\ &= a\mathbb{E}(X)+c\end{aligned}$$

■

As a consequence, the mean value of the difference of two random variables are equal to the difference of their expectations i.e. $\mathbb{E}(X-Y) = \mathbb{E}(X)-\mathbb{E}(Y)$.

Theorem 2.4 Let X and Y be random variables on the same sample space S. If X and Y are independent random variables then, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

■ **Remark 2.2** If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. However, the converse is not generally true: it is possible for $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ even though X and Y are dependent.

2.1.2.3 Variance and standard deviation of discrete random variables

The variance is an important characteristic of a random variable. It is a measure of dispersion of the random variable, it measures how much the probability mass is spread out around the mean value.

Definition 2.6 If X is a random variable with mean $\mathbb{E}(X) = \mu_X$, then the variance of X is defined by

$$Var(x) = \mathbb{E}\left((X - \mu_X)^2\right) = \mathbb{E}(X^2) - \mu_X^2.$$

The standard deviation σ of X is defined by

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

2.1.2.4 Properties of variance

The following properties are most useful for computing variance

1- If X and Y are independent then $Var(X+Y) = Var(X) + Var(Y)$.
2- For a constant c, $Var(cX) = c^2 Var(X)$
3- For a constant c, $Var(X+c) = Var(X)$.

2.1.3 Some discrete distributions

This section presents some common probability distributions for discrete random variables. In each case, we will define a probability mass function by specifying an explicit formula and we define the characteristics that will help to determine when to use a certain distribution in a given context.

2.1.3.1 Bernoulli distribution

Bernoulli distribution the simplest model that can be used in many real-life applications. A random experiment that can have only two outcomes 1 or 0, is known as a Bernoulli trial. In other words, the random variable can be 1 with a probability p or it can be 0 with a probability $q = (1-p)$. Typically, a value 1 is assigned to "success" and the value 0 indicates a "failure". The parameter p in the Bernoulli distribution refers to the probability of "success".

Definition 2.7 A random variable X is governed by the Bernoulli random distribution with parameter p, noted as $X \sim Bernoulli(p)$, if its PMF is as follows

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 2.2 Suppose X is a Bernoulli random variable with parameter p, then

$$\mathbb{E}(X) = p$$
$$Var(X) = pq$$

2.1.3.2 Binomial distribution

In general, we can connect binomial random variables to Bernoulli random variables. If X is a binomial random variable, with parameters n and p, then it can be written as the sum of n independent Bernoulli random variables, $X_1, \ldots, X_n$. If we define the random variable X_i, for $i = 1, \ldots, n$, to be 1 when the $i-th$ trial is a "success", and 0 when it is a "failure", then the sum

$$X = X_1 + \cdots + X_n$$

gives the total number of success in n trials.

Definition 2.8 A random variable X is governed by the binomial distribution with parameters n and p, noted as $X \sim Binomial(n, p)$, if its PMF is as follows

$$p_X(k) = \begin{cases} C_n^k p^k q^{n-k} & \text{for } k = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where $q = 1 - p$.

Proposition 2.3 Suppose X is a binomial random variable with parameters n and p, then

$$\mathbb{E}(X) = np$$
$$Var(X) = npq$$

■ **Example 2.7** A basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot, and let X be the r.v. that counts the number of successes in all trials. This is an example of a Bernoulli experiment with 4 trials, and we write $X \sim Binomial(4, 0.7)$.

Example 2.8 Flip a coin 12 times, and let X be the r.v. that counts the number of heads. This is an example of a Bernoulli Experiment with 12 trials, and we write $X \sim Binomial(12, 0.5)$.

2.1.3.3 Geometric and Pascal distribution

The geometric and Pascal distributions are related to the binomial distribution in that the underlying probability experiment is the same, i.e., independent trials with two possible outcomes. However, the random variable defined in the geometric and negative binomial case highlights a different aspect of the experiment, namely the number of trials needed to obtain a specific number of "successes". We start with the geometric distribution.

Definition 2.9 Suppose that a sequence of independent Bernoulli trials is performed, with $p = P(\text{"}success\text{"})$ for each trial. Let X be the random variable that refers to the number of trial at which the first success occurs. Then X is governed by the geometric distribution with parameter p, noted as $X \sim Geometric(p)$ and its PMF is as follows

$$\begin{aligned} p_X(k) &= P(1^{st} \text{ success on } k^{th} \text{ trial}) \\ &= P(1^{st}\ (k-1) \text{ trials are failures } \& \ k^{th} \text{ trial is success}) \\ &= q^{k-1}p, \quad \text{for } k = 1, 2, 3, \dots \end{aligned}$$

where $q = 1 - p$.

Proposition 2.4 If X is a Geometric random variable with parameter p, then

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{p} \\ Var(X) &= \frac{q}{p^2} \end{aligned}$$

Example 2.9 Both of the following examples are random variables governed by the geometric distribution.

1- Toss a fair coin until the first heads occurs. In this case, a "success" is getting a heads and so the parameter $p = 0.5$.

2- Buy lottery tickets until getting the first win. In this case, a "success" is getting a lottery ticket that wins money, and a "failure" is not winning. The parameter p will depend on the odds of wining for a specific lottery.

2.1.3.4 Pascal distribution

Pascal distribution (negative binomial) generalizes the geometric distribution. It is a repetition of independent trials of random experiments until getting r successes.

Definition 2.10 Suppose that a sequence of independent Bernoulli trials is performed, with $p = P(\text{"}success\text{"})$ for each trial. Let $r \geq 2$ integer and X be the random variable that refers to the r-th success occurs. Then X is governed by the Pascal distribution with parameters r and p, noted as $X \sim Pascal(r, p)$, if its PMF is as follows

$$\begin{aligned} P(X = k) &= P(r^{th} \text{ success is on } k^{th} \text{ trial}) \\ &= \underbrace{P(1^{st}\ (r-1) \text{ successes in } 1^{st}\ (k-1) \text{ trials})}_{\text{bniomial with } n=k-1} \times P(r^{th} \text{ success on } k^{th} \text{ trial}) \\ &= C_{r-1}^{k-1} p^{r-1} q^{(k-1)-(r-1)} \times p \\ &= C_{r-1}^{k-1} p^r q^{k-r}, \quad \text{for } k = r, r+1, r+2, \ldots \end{aligned}$$

Proposition 2.5 If X is a Pascal random variable with parameters r and p, then

$$\mathbb{E}(X) = \frac{rp}{q}$$
$$Var(X) = \frac{rp}{q^2}$$

■ **Example 2.10** For examples of the Pascal distribution, we can alter the geometric examples given in 2.9

- Toss a fair coin until get 8 heads. In this case, the parameter p is still given by $p = 0.5$, but now we also have the parameter $r = 8$, the number of desired "successes", i.e., heads.
- Buy lottery tickets until win 5 times. In this case, the parameter p is still given by the odds of winning the lottery, but now we also have the parameter $r = 5$, the number of desired wins.

2.1.3.5 Poison distribution

The Poisson distribution is a probability distribution that describes the number of events that occur in a fixed interval of time or space, given a known average rate of occurrence and assuming that events happen independently of each other. It's often used in situations where events happen randomly and at a constant average rate, such as the number of phone calls received by a call center in an hour, the number of accidents at a particular intersection in a day, or the number of emails received in an hour. Poisson distribution is used under certain conditions. They are:

- The number of trials $n$tends to infinity
- Probability of success p tends to zero
- the parameter $\lambda = np$ is finite

Definition 2.11 A random variable X is governed by the Poisson distribution with parameters λ (the average rate of events per interval), noted as $X \sim Poisson(\lambda)$, if its PMF is as follows

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

Proposition 2.6 Suppose X is a Poisson random variable with parameter λ, then

$$\mathbb{E}(X) = \lambda$$
$$Var(X) = \lambda$$

Example 2.11 Suppose on average there are 5 homicides per month in a given city. What is the probability there will be at most 1 in a certain month?
Solution If X is the number of homicides, we are given that $\lambda = 5$. Therefore

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-5} + 5e^{-5}.$$

Example 2.12 Assume on average there is one large earthquake per year in a country. What is the probability that next year there will be exactly 2 large earthquakes?
Solution:
We have $\lambda = 1$, then $P(X = 2) = e^{-1}\frac{1}{2}$.

2.2 Continuous random variables.

A continuous random variable can be defined as a random variable that can take on an infinite number of possible values. It is used to denote measurements such as height, weight, time, etc. Due to this, the probability that a continuous random variable will take on an exact value is 0. The cumulative distribution function (CDF) and the probability density function (PDF) are used to describe the characteristics of a continuous random variable. The main difference between continuous and discrete random variables is that continuous probability is measured over intervals, while discrete probability is calculated on exact points.

■ **Example 2.13** - The time it takes to complete an exam for a 60 minute test. Possible values = all real numbers on the interval $[0,60]$.
- Age of a fossil. Possible values = all real numbers on the interval [minimum age, maximum age].
- Miles per gallon for a such car. Possible Values = all real numbers on the interval [minimum MPG, maximum MPG].

Definition 2.12 A random variable X is said to be continuous if and only if the probability that it will belong to an interval $[a,b]$ can be expressed as an integral:

$$P(X \in [a,b]) = \int_a^b f_X(x)dx$$

where the integrand function $f_X(x) : \mathbb{R} \to [0,+\infty)$ is called the probability density function (PDF) of X.

■ **Example 2.14** Let X be a continuous random variable that can take any value in the interval $[0,1]$.

$$f_X(x) = \begin{cases} 4x^3 & \text{if } x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

The probability that X takes a value between 0.5 and 1 can be computed as follows:

$$\begin{aligned} P(X \in [0.5,1]) &= \int_{0.5}^1 f_X(x)dx \\ &= \int_{0.5}^1 4x^3 dx = \frac{15}{16} \end{aligned}$$

2.2.1 Probability density function (PDF)

Recall that continuous random variables have uncountably many possible values (think of intervals of real numbers). Just as for discrete random variables, we can talk about probabilities for continuous random variables using density functions.

Definition 2.13 The probability density function (PDF), denoted f_X, of a continuous random variable X satisfies the following properties:

1- $f_X(x) \geq 0$, for all $x \in \mathbb{R}$.
2- f_X is piecewise continuous.
3- $\int_{\mathbb{R}} f(x)\,dx = 1$
4- $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

Note that, unlike discrete random variables, continuous random variables have zero point probabilities, i.e., the probability that a continuous random variable equals a single value is always given by 0. Formally, this follows from properties of integrals:

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

2.2.2 Cumulative distribution function (CDF)

Recall Definition 2.3, the definition of the CDF, which applies to both discrete and continuous random variables. For continuous random variables we can further specify how to calculate the CDF with a formula as follows.

Definition 2.14 Let X be a continuous random variable that has PDF f_X, then the CDF F_X is given by

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

In other words, the CDF for a continuous random variable is found by integrating its PDF.

2.2.3 Mean and standard deviation of continuous random variables

Definition 2.15 If X is a continuous random variable with PDF f_X, then the mean of X is given by

$$\mu = \mu_X = \mathrm{E}[X] = \int_{\mathbb{R}} x \cdot f_X(x)\,dx.$$

- For the variance of a continuous random variable, the definition is the same as given by Definition 2.6, only we now integrate to calculate the value:

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mu^2 = \left(\int_{\mathbb{R}} x^2 \cdot fX(x)\, dx \right) - \mu^2$$

Example 2.15 Let X be a continuous random variable with PDF

$$f_X(x) = \begin{cases} \frac{3}{x^4} & \text{if } x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and variance of X.
Solution

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x f_X(x) dx \\ &= \int_1^{+\infty} \frac{3}{x^3} dx \\ &= \left[-\frac{3}{2} x^{-2} \right]_1^{+\infty} = \frac{3}{2} \end{aligned}$$

Then, $\mu_X = 3/2$. Next, we have

$$\begin{aligned} E[X^2] &= \int_{\mathbb{R}} x^2 f_X(x) dx \\ &= \int_1^{+\infty} \frac{3}{x^2} dx \\ &= \left[-3x^{-1} \right]_1^{+\infty} = 3 \end{aligned}$$

Thus, we have

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

2.2.4 Some continuous distributions

2.2.4.1 Uniform distribution

The uniform distribution is a continuous probability distribution and is concerned with events that are equally likely to occur.

Definition 2.16 A random variable X is governed by the Uniform distribution with parameters a and b, noted as $X \sim U(a,b)$, if its density function is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

Proposition 2.7 Let $X \sim U(a,b)$, then

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

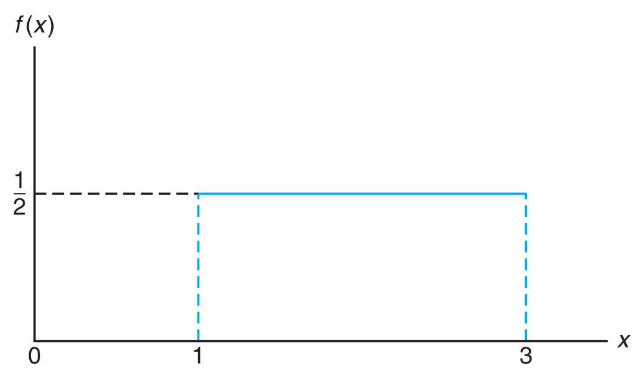


Figure 2.1: The density function for a uniform random variable on the interval $[1,3]$.

■ **Example 2.16** Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval $[0,4]$.
a- What is the probability density function?

b- What is the probability that any given conference lasts at least 3 hours?

Solution

a- The appropriate density function for the uniformly distributed random variable X in this situation is

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

b-

$$P(X \geq 3) = \int_3^4 \frac{1}{4} dx = \frac{1}{4}.$$

2.2.4.2 Normal (Gaussian) distribution

The normal distribution, which is continuous, is the most important of all the probability distributions. Its graph is bell-shaped, see Figure 3.2. This bell-shaped curve is used in almost all disciplines. Since it is a continuous distribution, the total area under the curve is one. The parameters of the normal are the mean μ and the standard deviation σ. A special normal distribution, called the standard normal distribution is the distribution of z-scores. Its mean is **zero**, and its standard deviation is **one**.

Definition 2.17 A random variable X is governed by the Normal(Gaussian) distribution with parameters μ (mean) and σ (standard deviation), noted as $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density function is

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

Figure 3.3 shows two normal curves having different means and different standard deviations. Clearly, they are centred at different positions on the horizontal axis and their shapes reflect the two different values of σ.

Proposition 2.8 Let $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}(X) = \mu$$
$$Var(X) = \sigma^2$$

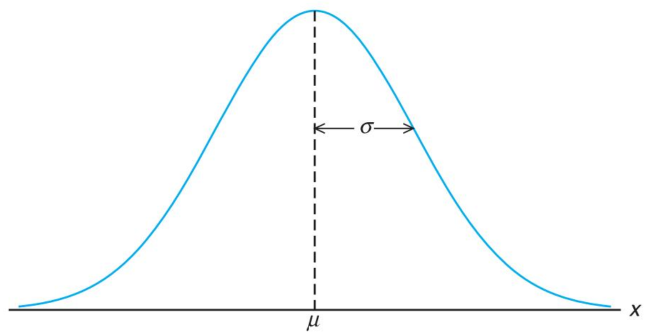


Figure 2.2: The density function for a Normal random variable.

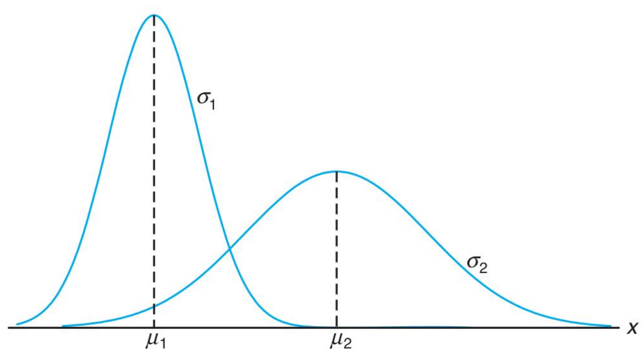


Figure 2.3: Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.

Definition 2.18 The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.

Theorem 2.5 If Z is a standard normal random variable i.e., $Z \sim \mathcal{N}(0,1)$ and $X = \sigma Z + \mu$, then X is a normal random variable with mean μ and variance σ^2, i.e,

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Conversely, if $X \sim \mathcal{N}(\mu, \sigma^2)$, the random variable defined by $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable, i.e., $Z \sim \mathcal{N}(0,1)$.

Proof. Let $\Phi(.)$ the CDF of Z, then to compute the CDF of X, we can write

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(\sigma Z + \mu \leq x) \qquad \big(\text{where } Z \sim \mathcal{N}(0,1)\big) \\
&= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{x-\mu}{\sigma}\right).
\end{aligned}$$

It results that

$$F_X(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

and

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

■

■ **Example 2.17** Let $X \sim \mathcal{N}(-5,4)$. Find the probabilities

a- $P(X < 0)$

b- $P(-7 < X < -3)$

c- $P(X > -3 | X > -5)$

Solution

Since $\mu = -5$ and $\sigma = 2$, then

a-

$$\begin{aligned}
P(X < 0) &= F_X(0) \\
&= \Phi\left(\frac{0-(-5)}{2}\right) \\
&= \Phi(2.5) \approx 0.99
\end{aligned}$$

b-

$$\begin{aligned}
P(-7 < X < -3) &= F_X(-3) - F_X(-7) \\
&= \Phi\left(\frac{(-3)-(-5)}{2}\right) - \Phi\left(\frac{(-7)-(-5)}{2}\right) \\
&= \Phi(1) - \Phi(-1) \\
&= 2\Phi(1) - 1 \qquad \big(\text{since } \Phi(-x) = 1 - \Phi(x)\big) \approx 0.68
\end{aligned}$$

c-

$$
\begin{aligned}
P(X > -3 | X > -5) &= \frac{P(X > -3, X > -5)}{P(X > -5)} \\
&= \frac{P(X > -3)}{P(X > -5)} \\
&= \frac{1 - \Phi\left(\frac{(-3)-(-5)}{2}\right)}{1 - \Phi\left(\frac{(-5)-(-5)}{2}\right)} \\
&= \frac{1 - \Phi(1)}{1 - \Phi(0)} \\
&\approx \frac{0.1587}{0.5} \approx 0.32
\end{aligned}
$$

Theorem 2.6 If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, and $Y = aX + b$, where $a, b \in \mathbb{R}$, then $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ where

$$\mu_Y = a\mu_X + b, \quad \sigma_Y^2 = a^2 \sigma_X^2.$$

Proof. Since we have

$$X = \sigma_X Z + \mu_X \qquad \text{where } Z \sim \mathcal{N}(0,1),$$

it follows that

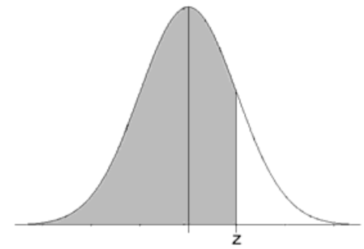
$$
\begin{aligned}
Y &= a(\sigma_X Z + \mu_X) + b \\
&= (a\sigma_X)Z + (a\mu_X + b).
\end{aligned}
$$

Then

$$Y \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2).$$

■

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

2.2.4.3 Exponential distribution

Exponential distribution is often used to predict the waiting time until the next event occurs, such as a success, failure, or arrival. For example the amount of time you need to wait for the bus to arrive.

Definition 2.19 A random variable X is governed by the exponential distribution with parameter $\lambda > 0$, noted as $X \sim \exp(\lambda)$, if its density function is

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

Figure 2.4 shows the PDF of the exponential distribution for some values of λ.

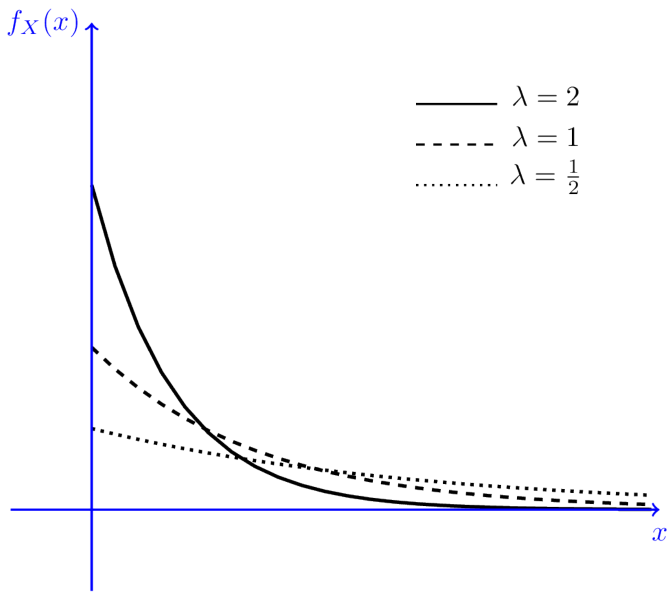


Figure 2.4: Probability density functions of the exponential distribution for some values of λ.

Proposition 2.9 Let $X \sim \exp(\lambda), \lambda > 0$, then

$$\mathbb{E}(X) = \lambda$$

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

■ **Example 2.18** Assume that, you usually get 2 phone calls per hour. calculate the probability, that a phone call will come within the first hour.

Solution
It is given that, 2 phone calls per hour. Then, it would expect that one phone call at every half-an-hour. So, we can take $\lambda = 0.5$. Hence, the computation is as follows:

$$p(0 \leq X \leq 1) = \int_0^1 0.5e^{-0.5x}$$

Therefore, the probability of arriving the phone calls within the first hour is 0.393

2.2.4.4 Gamma distribution

Gamma distributions are usually used to model wait times until events. Unlike the exponential distribution which models the time until the first event, gamma models the time until the α-th event. Before introducing the gamma random variable, we need to introduce the gamma function.

Definition 2.20 Let us take a parameter $\alpha > 0$. Gamma function $\Gamma(\alpha)$ is an extension of the factorial function to real (and complex) numbers i.e., for $\alpha \in \{1, 2, 3, \dots\}$, then

$$\Gamma(\alpha) = (\alpha - 1)!$$

More generally, for any positive real number α the gamma function is defined as follows

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Proposition 2.10 Let α a strictly positive real number, then gamma function has the following properties:

- $\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}, \qquad \text{for } \lambda > 0;$
- $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha);$
- $\Gamma(\alpha) = (\alpha - 1)!, \text{ for } \alpha = 1, 2, 3, \cdots;$

Now we define the gamma distribution by its PDF

Definition 2.21 A random variable X is governed by the gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$, noted as $X \sim Gamma(\alpha, \lambda)$, if its density function is

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Figure 2.5 shows the PDF of Gamma distribution for some values of λ and α.

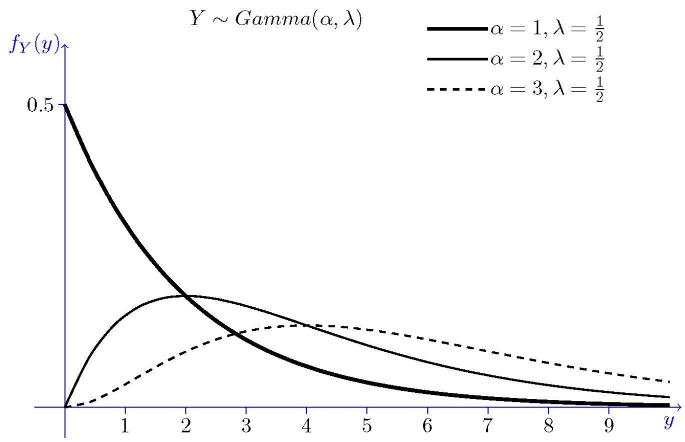


Figure 2.5: Probability density functions of the gamma distribution for some values of λ and α.

Proposition 2.11 Let $X \sim Gamma(\alpha, \lambda)$, then

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}$$
$$Var(X) = \frac{\alpha}{\lambda^2}.$$

■ **Example 2.19** On average, someone sends a money order once per our. What is the probability someone sends 3 money orders in less than 3 hours?.

2.3 Solved exercises

Exercise 2.1 Classify each random variable as either discrete or continuous.
1- The number of arrivals at an emergency room between midnight and 6:00 a.m.
2- The weight of a box of cereal labelled “18 ounces.”
3- The duration of the next outgoing telephone call from a business office.
4- The number of kernels of popcorn in a 1-pound container.
5- The number of applicants for a job.
Solution

1- Discrete.
2- Continuous.
3- Continuous.
4- Discrete.
5- Discrete.

Exercise 2.2 I roll two dice and observe two numbers X and Y.
a- Find S_X and S_Y and the PMFs of X and Y.
b- Compute $P(X=2, Y=6)$
c- Compute $P(X>3|Y=2)$
d- Let $Z=X+Y$. Find the range and PMF of Z
e- $P(X=4|Z=8)$.
Solution
a- $S_X = S_Y = \{1,2,3,4,5,6\}$

$$P_X(k) = \begin{cases} \frac{1}{6} & \text{for } k=1,2,3,4,5,6 \\ 0 & \text{otherwise} \end{cases}$$

Similarly,

$$P_Y(k) = \begin{cases} \frac{1}{6} & \text{for } k=1,2,3,4,5,6 \\ 0 & \text{otherwise} \end{cases}$$

b- Since X and Y are independent random variables, then

$$\begin{aligned} P(X=2,Y=6) &= P(X=2)P(Y=6) \\ &= \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36} \end{aligned}$$

c- Since X and Y are independent, then

$$\begin{aligned} P(X>3|Y=2) &= \frac{P(X>3,Y=2)}{P(Y=2)} \\ &= \frac{P(X>3)P(Y=2)}{P(Y=2)} \\ &= P_X(4)+P_X(5)+P_X(6) \\ &== \frac{1}{6}+\frac{1}{6}+\frac{1}{6} = \frac{1}{2}. \end{aligned}$$

d- Since we have $S_Z = \{2,3,4,...,12\}$. Thus, PMF of Z is as follows

$$\begin{aligned}
P_Z(2) &= P(Z=2) = P(X=1,Y=1) \\
&= P(X=1)P(Y=1) \text{ (since } X \text{ and } Y \text{ are independent)} \\
P_Z(3) &= P(Z=3) = P(X=1,Y=2) + P(X=2,Y=1) \\
&= P(X=1)P(Y=2) + P(X=2)P(Y=1) \\
&= \frac{1}{6}\cdot\frac{1}{6} + \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{18}
\end{aligned}$$

We continuous this process we get $P_Z(4) = 3\cdot\frac{1}{36} = \frac{1}{12}, P_Z(5) = 1/9, P_Z(6) = 5/36, P_Z(7) = 1/6, P_Z(8) = 5/36, P_Z(9)1/9, P_Z(10) = 1/12, P_Z(11) = 1/18, P_Z(12) = 1/36$

d- Since $Z = X+Y$, it is clear that Z depends on the values of X. So,

$$\begin{aligned}
P(X=4|Z=8) &= \frac{P(X=4,Z=8)}{P(Z=8)} \\
&= \frac{P(X=4,Y=4)}{P(Z=8)} \\
&= \frac{P(X=4)P(Y=4)}{P(Z=8)} \text{ (since } X \text{ and } Y \text{ are independent)} \\
&= \frac{\frac{1}{6}\cdot\frac{1}{6}}{\frac{5}{36}} = \frac{1}{5}
\end{aligned}$$

Exercise 2.3 Let X be a continuous random variable with PDF

$$f_X(x) = \begin{cases} 4x^3 & 0 < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Compute $P(X \leq \frac{2}{3}|X > \frac{1}{3})$.

Solution

$$\begin{aligned}
P(X \leq \frac{2}{3}|X > \frac{1}{3}) &= \frac{P(\frac{1}{3} < X \leq \frac{2}{3})}{P(X > \frac{1}{3})} \\
&= \frac{\int_{\frac{1}{3}}^{\frac{2}{3}} 4x^3 dx}{\int_{\frac{1}{3}}^{1} 4x^3 dx} = \frac{3}{16}.
\end{aligned}$$

Exercise 2.4 Suppose that the length of a phone call in minutes is an exponential random variable with average length 10 minutes.
a- What is the probability of your phone call being more than 10 minutes?
b- What is the probability of your phone call being Between 10 and 20 minutes?
Solution
a- We have Here $\lambda = \frac{1}{10}$, thus

$$P(X > 10) = e^{-0.1 \times 10} = e^{-1} = 0.368.$$

b-

$$P(10 < X < 20) = F_X(20) - F_X(10) = e^{-1} - e^{-2} = 0.233.$$

Exercise 2.5 Suppose the life of a smart-phone has exponential distribution with mean life of 4 years. Let X denote the life of a phone (or time until it dies). Given that the phone has lasted 3 years, what is the probability that it will 5 more years.
Solution
We have $\lambda = 4$, then

$$\begin{aligned} P(X > 5 + 3 | X > 3) &= \frac{P(X > 8)}{P(X > 3)} \\ &= \frac{e^{-4 \times 8}}{e^{-4 \times 3}} = e^{-4 \times 5} = P(X > 5)) \end{aligned}$$

Exercise 2.6 Suppose that the time (in minutes) required to check out a book at the library can be represented by an exponentially distributed random variable with parameter $\lambda = \frac{2}{11}$.
a- What is the probability that it will take at least 5 minutes to check out a book?
b- What is the probability that it will take at least 11 minutes to check out a book given that you have already waited for 6 minutes?
Solution
a-

$$P(X > 5) = e^{-\frac{10}{11}}$$

b- Since the exponential distribution is a memoryless one then

$$\begin{aligned} P(X > 11 | X > 6) &= P(X > 6 + 5 | X > 6) \\ &= P(X > 5) = e^{-\frac{10}{11}} \end{aligned}$$

Exercise 2.7 Let's say that the time (measured in minutes) needed to complete the checkout process for a book at the library follows an exponential distribution with parameter $\lambda = 2/11$.
1- What is the probability that it will take at least 5 minutes to check out a book?
2- What is the probability that it will take at least 11 minutes to check out a book given that you have already waited for 6 minutes?
Solution
1-

$$\mathbb{P}(X > 5) = e^{-\frac{10}{11}}$$

2- Using the memoryless propriety

$$\begin{aligned}\mathbb{P}(X > 11 \mid X > 6) &= \mathbb{P}(X > 6+5 \mid X > 6) \\ &= \mathbb{P}(X > 5) = e^{-\frac{10}{11}}\end{aligned}$$

Exercise 2.8 An insurance company insures a large number of homes. The insured value, denoted as X, of any randomly chosen home, is presumed to follow to a distribution characterized by its density function.

$$f_X(x) = \begin{cases} \frac{8}{x^3} & x > 2 \\ 0 & \text{otherwise} \end{cases}$$

1- Knowing that a randomly selected home is insured for at most 4, find the probability that it is insured for less than 3.
2- Knowing that a randomly selected home is insured for at least 3, find the probability that it is insured for less than 4.
Solution
1- Since we have

$$\mathbb{P}(X < 4) = \int_2^4 \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_2^4 = -\frac{1}{4} + 1 = \frac{3}{4}$$

and

$$\mathbb{P}(X < 3) = \mathbb{P}(2 < X < 3) = \int_2^3 \frac{8}{x^3} dx = -\frac{4}{x^2}\Big|_2^3 = -\frac{4}{9} + 1 = \frac{5}{9}.$$

Using the Conditional probability definition, we get

$$\mathbb{P}(X < 3 \mid X < 4) = \frac{\mathbb{P}(X < 3)}{\mathbb{P}(X < 4)} = \frac{\frac{5}{9}}{\frac{3}{4}} = \frac{20}{27} \approx 0.74074074.$$

2- Since we have

$$\mathbb{P}(X>3)=\int_3^\infty \frac{8}{x^3}dx=-\left.\frac{4}{x^2}\right|_3^\infty=\frac{4}{9}$$
$$\mathbb{P}(3<X<4)=\int_3^4 \frac{8}{x^3}dx=-\left.\frac{4}{x^2}\right|_3^4=\frac{7}{36}.$$

Then, by the conditional probability definition, it follows that

$$\mathbb{P}(X<4\,|\,X>3)=\frac{\mathbb{P}(3<X<4)}{\mathbb{P}(X>3)}=\frac{\frac{7}{36}}{\frac{4}{9}}=\frac{7}{16}\approx 0.4375$$

Exercise 2.9 A company sets prices for its hurricane insurance based on the following assumptions:
1- At most one hurricane can occur in any given calendar year.
2- The probability of a hurricane in any given calendar year is 0.05.
3- The occurrences of hurricanes in different calendar years are mutually independent.
Using these assumptions, calculate the probability that there are at most 2 hurricanes in a 20-year period.
Solution
From the assumptions it seems that $X \sim Binomial(20, 0.05)$. Therefore,

$$\mathbb{P}(X\leqslant 2)=C(20,0)(0.05)^0(0.95)^{20}+C(20,1)(0.05)^1(0.95)^{19}+C(20,2)(0.05)^2(0.95)^{18}$$
$$=0.9245.$$

Exercise 2.10 The expected number of of typographical errors on a page of the new Harry Potter book is 0.2. Explain what assumptions you used to find the probability that the next page you read contains::
1- 0 typographical errors?
2- At least 2 typographical errors?
Solution

Considering that each word has a small probability of being a typo, the number of typos should be approximately distributed according to a Poisson distribution. Hence
1- $e^{-0.2}$
2- $1-e^{-0.2}-0.2e^{-0.2}=1-1.2e^{-0.2}$.

Exercise 2.11 A certain type of storage battery lasts, on average, 3 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, calculate the probability that a given battery will last less than 2.3 years.

Solution

We have $Z = \frac{2.3-3}{0.5} = -1.4$, therefore

$$P(X < 2.3) = P(Z < -1.4) = 0.0808$$

Exercise 2.12 In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped?

Solution

We have $Z_1 = \frac{2.99-3}{0.005} = -2$ and $Z_2 = \frac{3.01-3}{0.005} = 2$. Hence

$$\begin{aligned} P(2.99 < X < 3.01) &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z < -2) = 2P(Z < 2) - 1 = 2(0.9772) - 1 = 0.9544 \end{aligned}$$

Consequently, it is expected that approximately 4.56% of manufactured ball bearings will be scrapped on average.

Exercise 2.13 Gauges are utilized to reject all components for which a certain dimension does not fall within the specification of $1.50 \pm d$. It is established that this measurement follows a normal distribution with a mean of 1.5 and a standard deviation of 0.2. Find the value d such that the specifications "cover" 95% of the measurements.

Solution

From standard normal distribution table we have

$$P(-1.96 < Z < 1.96) = 0.95.$$

It follows that

$$1.96 = \frac{(1.50 + d) - 1.50}{0.2}$$

from which we obtain,

$$1.96 = \frac{(1.50 + d) - 1.50}{0.2}$$

Exercise 2.14 A specific machine manufactures electrical resistors with a mean resistance of 40 ohms and a standard deviation of 2 ohms. It is assumed that the resistance follows a normal distribution and can be measured with any degree of precision, what percentage of resistors will have a resistance exceeding 43 ohms?
Solution

$$z = \frac{43-40}{2} = 1.5$$

Then,

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668.$$

Therefore, 6.68% of the resistors will have a resistance that exceed 43 ohms.

Exercise 2.15 This week, a financial regulator from the Federal Reserve (FED) will assess two banks. For each evaluation, the regulator will choose with equal probability between two different stress tests. Failing the first stress test incurs a penalty of 10000$ for a bank, while failing the second test results in a 5000$ penalty. The probability of the first bank failing either test is 0.4, while independently, the second bank has a 0.5 probability of failing either test. Let X represent the total fees collected by the regulator after evaluating both banks. Determine the cumulative distribution function of X.
Solution
The random variable X can take the values $0, 5000, 10000, 15000$ and 20000 depending on which test was applied to each bank, and if the bank fails the evaluation or not. Denote by B_i the event that the i-th bank fails and by T_i the event that test i applied. Then

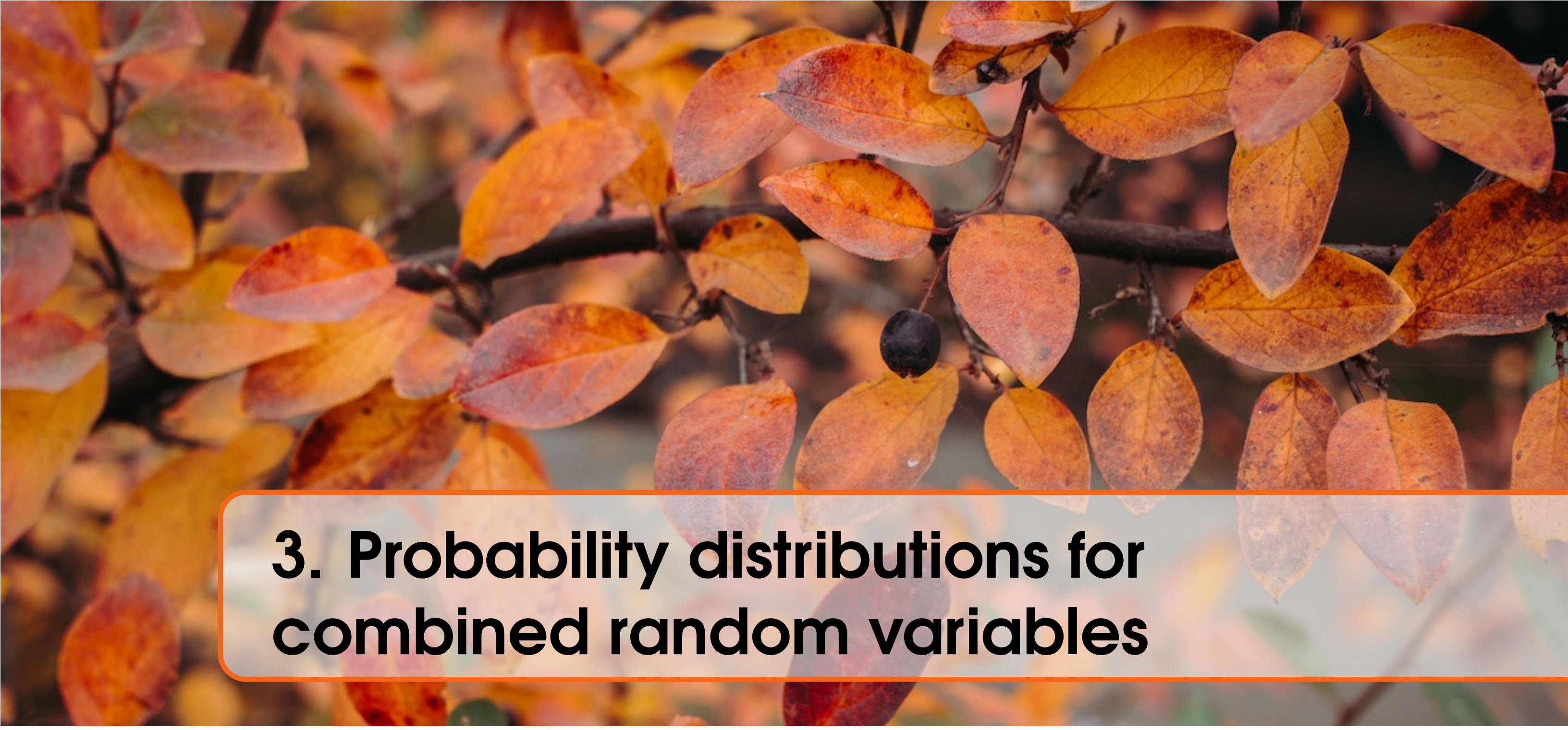
$$\mathbb{P}(T_1) = \mathbb{P}(T_2) = 0.5, \mathbb{P}(B_1) = \mathbb{P}(B_1 \mid T_1) = \mathbb{P}(B_1 \mid T_2) = 0.4$$
$$\mathbb{P}(B_2) = \mathbb{P}(B_2 \mid T_1) = \mathbb{P}(B_2 \mid T_2) = 0.5$$

Since banks and tests are independent we have

$$\mathbb{P}(X=0) = \mathbb{P}(B_1^c \cap B_2^c) = \mathbb{P}(B_1^c) \cdot \mathbb{P}(B_2^c) = 0.6 \cdot 0.5 = 0.3,$$
$$\mathbb{P}(X=5000) = \mathbb{P}(B_1)\mathbb{P}(T_2)\mathbb{P}(B_2^c) + \mathbb{P}(B_1^c)\mathbb{P}(B_2)\mathbb{P}(T_2) = 0.25,$$
$$\mathbb{P}(X=10000) = \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2^c) + \mathbb{P}(B_1)\mathbb{P}(T_2)\mathbb{P}(B_2)\mathbb{P}(T_2) + \mathbb{P}(B_1^c)\mathbb{P}(B_2)\mathbb{P}(T_1) = \frac{3}{10}$$
$$\mathbb{P}(X=15000) = \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2)\mathbb{P}(T_2) + \mathbb{P}(B_1)\mathbb{P}(T_2)\mathbb{P}(B_2)\mathbb{P}(T_1) = 0.1$$
$$\mathbb{P}(X=20000) = \mathbb{P}(B_1)\mathbb{P}(T_1)\mathbb{P}(B_2)\mathbb{P}(T_1) = 0.05.$$

Therefore, the CDF is given as follows

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.3 & 0 \leqslant x < 5000 \\ 0.55 & 5000 \leqslant x < 10000 \\ 0.85 & 10000 \leqslant x < 15000 \\ 0.95 & 15000 \leqslant x < 20000 \\ 1 & x \geqslant 20000 \end{cases}$$



3. Probability distributions for combined random variables

In this chapter, we focus on how we model the probability distribution of two (or more) random variables jointly, that are defined within the same sample space. We begin with the discrete case by seeking for the joint probability mass function of two discrete random variables. Next, we will consider the continuous case.

3.1 Joint distributions of discrete random variables

Definition 3.1 (**joint probability mass function**) Let X and Y, discrete random variables that are defined on the same sample space S, then their joint probability mass function is defined as

$$P_{XY}(x,y) = P(X = x, Y = y) = p(x,y)$$

where (x,y) is a pair of possible values for the pair of random variables (X,Y), and $p(x,y)$ satisfies the following conditions:

- $0 \leq p(x,y) \leq 1$
- $\sum_{x \in A} \sum_{y \in A} p(x,y) = 1$
- $P((X,Y) \in A)) = \sum_{x \in A} \sum_{y \in A} p(x,y)$

Definition 3.2 (**joint cumulative probability function**) Let X and Y are two discrete random variables that are defined on the same sample space S, then their joint cumulative distribution function is obtained by summing the joint PMF's, i.e:

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p\left(x_i, y_j\right)$$

where x_i and y_j denote respectively the possible values of X and Y.

From the above definitions, we can obtain the PMF of X from its joint PMF with Y. Indeed,

$$\begin{aligned} P_X(x) &= P(X = x) \\ &= \sum_{y_j \in S_Y} P(X = x, Y = y_j) \\ &= \sum_{y_j \in S_Y} P_{XY}(x, y_j). \end{aligned}$$

Here, $P_X(x)$ is called the marginal probability mass function of X. Similarly, we can derive the marginal PMF of Y from its joint PMF with X.

Definition 3.3 Let X and Y are discrete random variable and let $p(x,y)$ their joint PMF. The marginal probability mass functions of X and Y are respectively defined as follows:

$$P_X(x) = \sum_{y_j \in S_Y} p(x, y_j), \qquad \text{for any } x \in S_X$$

$$P_Y(y) = \sum_{x_i \in S_X} p(x_i, y), \qquad \text{for any } y \in S_Y$$

■ **Example 3.1** We toss a fair coin three times and record the sequence of heads (H) and tails (T). Let X the random variable of the number of obtained heads and Y denotes the winnings earned in a single play of a game with the following rules, based on the outcomes of the probability experiment:
- player wins 1\$ if first heads occurs on the first toss
- player wins 2\$ if first heads occurs on the second toss
- player wins 3\$ if first heads occurs on the third toss
- player loses 1\$ if no heads occur

The possible values of X are $S_x = \{0,1,2,3\}$, and the possible values of Y are $S_y = \{-1,1,2,3\}$. We represent the joint PMF in Table 3.1 below. Given the joint PMF, the marginal PMF's of both X and Y is represented in Table 3.2, below. Also, the joint CDF for X and Y is represented

Table 3.1: Joint PMF of X and Y

$p(x,y)$	X			
Y	0	1	2	3
-1	1/8	0	0	0
1	0	1/8	2/8	1/8
2	0	1/8	1/8	0
3	0	1/8	0	0

Table 3.2: Marginal PMF's for X and Y

x	$p_X(x)$	y	$p_Y(y)$
0	1/8	-1	1/8
1	3/8	1	1/2
2	3/8	2	1/4
3	1/8	3	1/8

Table 3.3: Joint CDF of X and Y

$F(x,y)$	X			
Y	0	1	2	3
-1	1/8	1/8	1/8	1/8
1	1/8	1/4	1/2	5/8
2	1/8	3/8	3/4	7/8
3	1/8	1/2	7/8	1

in Table 1.3 below, by summing over values of the joint frequency function. For example, consider $F(1,1)$:

$$F(1,1) = P(X \leq 1 \text{ and } Y \leq 1) = \sum_{x \leq 1} \sum_{y \leq 1} p(x,y) = p(0,-1) + p(0,1) + p(-1,1) + p(1,1) = \frac{1}{4}$$

Definition 3.4 (Independent Random Variables)
Discrete random variables $X_1, X_2, \ldots, X_n$ are considered independent if the joint probability mass function factors into a product of the marginal PMF's i.e,

$$p(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

This condition holds also for the cumulative distribution functions.

For example in Table 3.2 of the above example, in the case where $(x, y) = (0, -1)$, we have

$$p(0, -1) = \frac{1}{8}, p_X(0) = \frac{1}{8}, p_Y(-1) = \frac{1}{8},$$

this mean that $p(0, -1) \neq p_X(0) \cdot p_Y(-1)$, so the variables X and Y are not independent (in other word they are dependent).

3.1.1 Expectations of joint discrete distributions

Theorem 3.1 Let X and Y are discrete random variables and let $p(x, y)$ their joint PMF. If $g(X, Y)$ is a function of these two random variables, then its mean value is given as follows:

$$\mathrm{E}[g(X, Y)] = \sum_{x \in \S_X} \sum_{y \in \S_Y} g(x, y) p(x, y)$$

■ **Example 3.2** Lets consider Example 3.1. Find $E(XY), E(X)$ and $E(Y)$.
Solution
1- For the first case, we define $g(x, y) = xy$. Therefore, the expected value of XY is :

$$\begin{aligned}
\mathrm{E}[XY] = \sum_{x \in S_X} \sum_{y \in S_Y} xy \cdot p(x, y) = (0)(-1)\left(\frac{1}{8}\right) \\
&+ (1)(1)\left(\frac{1}{8}\right) + (2)(1)\left(\frac{2}{8}\right) + (3)(1)\left(\frac{1}{8}\right) \\
&+ (1)(2)\left(\frac{1}{8}\right) + (2)(2)\left(\frac{1}{8}\right) \\
&+ (1)(3)\left(\frac{1}{8}\right) \\
&= \frac{17}{8} = 2.125
\end{aligned}$$

2- For the second case, we define $g(x) = x$, so the expected value of X :

$$\begin{aligned}
\mathrm{E}[X] = \sum_{x \in S_X} x \cdot p(x,y) &= (0)\left(\frac{1}{8}\right) \\
&+ (1)\left(\frac{1}{8}\right) + (2)\left(\frac{2}{8}\right) + (3)\left(\frac{1}{8}\right) \\
&+ (1)\left(\frac{1}{8}\right) + (2)\left(\frac{1}{8}\right) \\
&+ (1)\left(\frac{1}{8}\right) \\
&= \frac{12}{8} = 1.5
\end{aligned}$$

3. For the last case, we define $g(x,y) = y$, and the expected value of Y is :

$$\begin{aligned}
\mathrm{E}[Y] = \sum_{y \in S_Y} y \cdot p(x,y) &= (-1)\left(\frac{1}{8}\right) \\
&+ (1)\left(\frac{1}{8}\right) + (1)\left(\frac{2}{8}\right) + (1)\left(\frac{1}{8}\right) \\
&+ (2)\left(\frac{1}{8}\right) + (2)\left(\frac{1}{8}\right) \\
&+ (3)\left(\frac{1}{8}\right) \\
&= \frac{10}{8} = 1.25
\end{aligned}$$

Theorem 3.2 If X and Y are independent random variables, then $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

3.2 Joint distributions of continuous random variables

As considered in the discrete case, we interest now by the joint distributions for continuous case.

Definition 3.5 (joint probability density function) Let X and Y are two continuous random variables that defined on the same sample space S, then their joint probability density function is a piecewise continuous function, denoted $f(x,y)$, that satisfies the following properties.
1- $f(x,y) \geq 0$, for all $(x,y) \in \mathbb{R}^2$
2- $\iint_{\mathbb{R}^2} f(x,y)dxdy = 1$
3- $P((X,Y) \in A) = \iint_A f(x,y)dxdy$, for any $A \subseteq \mathbb{R}^2$

Definition 3.6 (joint cumulative distribution function) Let X and Y are continuous random variables that are defined on the same sample space S, then their joint cumulative distribution function is obtained by integrate the joint PDF i.e, if A is given as

$$A = \left\{(x,y) \in \mathbb{R}^2 \mid X \leq a \text{ and } Y \leq b\right\}$$

where a and b are constants, then the joint CDF of X and Y, at the point (a,b), is given by

$$F(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)dxdy$$

Similarly to the discrete case, we can derive the individual marginal probability density functions (PDF's) of X and Y from the joint PDF.

Definition 3.7 Let X and Y are continuous random variables and let $f(x,y)$ their joint PDF. The marginal probability density functions of X and Y are respectively defined as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad (\text{ fix a value of } X, \text{ and integrate over all possible values of } Y)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \quad (\text{ fix a value of } Y, \text{ and integrate over all possible values of } X)$$

Similarly, in the continuous case, we can also define independent random variables in the same manner as we did for discrete random variables.

Definition 3.8 (Independent random variables)
Continuous random variables $X_1, X_2, \ldots, X_n$ are considered independent if the joint probability density function factors into a product of the marginal PDF's i.e,

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

This condition holds also for the cumulative distribution functions.

3.2.1 Expectations of joint continuous distributions

Let X and Y are continuous random variable and let $f(x,y)$ their joint PDF. If $g(X,Y)$ is a function of these two random variables, then its mean value is given as follows:

$$\mathrm{E}[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)f(x,y)dxdy$$

Example 3.3 Let X and Y be two joint continuous random variable having the following joint PDF

$$f_{XY}(x,y) = \begin{cases} x+y & 0 \le x,y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- Determine $E[XY^2]$

Solution

$$\begin{aligned} E[XY^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy^2 f_{XY}(x,y)\ dxdy \\ &= \int_0^1\int_0^1 xy^2(x+y)\ dxdy \\ &= \int_0^1\int_0^1 x^2y^2+xy^3\ dxdy \\ &= \int_0^1 \left(\frac{1}{3}y^2+\frac{1}{2}y^3\right) dy \\ &= \frac{17}{72}. \end{aligned}$$

3.3 Solved exercises

Exercise 3.1 Assume that a radioactive particle is confined in a unit square and X and Y are the random variables, which represent the particle's location in the unit square, with the bottom left corner placed at the origin. Radioactive particles follow entirely random behavior that is uniformly distributed over the unit square, having the following joint PDF

$$f(x,y) = \begin{cases} c, & \text{if } 0 \le x, y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

where c is a constant.
1- Determine the value of the constant c.
2- Let the set $A = \{(x,y) \mid x - y > 0.5\}$. Find the probability $P((X,Y) \in A)$
3- Find the marginal PDF's of X and Y
4- Determine $E[XY]$.

Solution

1- From joint PDF definition, it results that $\iint_{\mathbb{R}^2} f(x,y)dxdy = 1$, therefore,

$$\int_0^1 \int_0^1 c dxdy = 1 \quad \Rightarrow \quad c \int_0^1 \int_0^1 1 dxdy = 1 \quad \Rightarrow \quad c = 1$$

2- Figure 1 represents the graph of region A, then

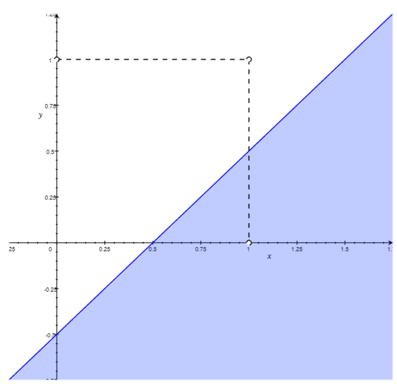


Figure 1: Graph of region A, shaded in blue.

$$P(X - Y > 0.5) = \iint_A f(x,y)dxdy = \int_0^{0.5} \int_{y+0.5}^1 1 dxdy = 0.125$$

3-

$$f_X(x) = \int_0^1 1 dy = 1, \quad \text{for } 0 \le x \le 1$$
$$f_Y(y) = \int_0^1 1 dx = 1, \quad \text{for } 0 \le y \le 1$$

4-The expected value of XY

$$\mathrm{E}[XY] = \iint_{\mathbb{R}^2} xy \cdot f(x,y)dxdy = \int_0^1 \int_0^1 xy \cdot 1dxdy = \int_0^1 \left(\frac{x^2}{2} y \Big|_0^1 \right) dy = \frac{1}{4}$$

Exercise 3.2 Let $X,Y \sim U(0,1)$ and they are independents. Find the cumulative joint distribution.

Solution

We have

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$

and

$$F_Y(y) = \begin{cases} 0 & \text{for } y < 0 \\ y & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

Since X and Y are independent, then

$$F_{XY}(x,y) = F_X(x)F_Y(y) = \begin{cases} 0 & \text{for } y < 0 \text{ or } x < 0 \\ xy & \text{for } 0 \leq x,y \leq 1 \\ y & \text{for } x > 1, 0 \leq y \leq 1 \\ x & \text{for } y > 1, 0 \leq x \leq 1 \\ 1 & \text{for } x > 1, y > 1 \end{cases}$$

Exercise 3.3 Two points are chosen uniformly and independently along a stick of length 1. determine the mean distance between those two points?

Solution

We aim to find $E[|X-Y|]$. Since X and Y are uniformly chosen, then their joint PDF is

$$f(x,y) = \begin{cases} 1 & 0 < x,y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Hence,

$$
\begin{aligned}
E[|X-Y|] &= \int_0^1 \int_0^1 |x-y| \cdot dx\,dy \\
&= \iint_{x \geq y} (x-y)\,dx\,dy + \iint_{y>x} (y-x)\,dx\,dy \\
&= \int_0^1 \int_y^1 (x-y)\,dx\,dy + \int_0^1 \int_0^y (y-x)\,dx\,dy \\
&= \frac{1}{3}.
\end{aligned}
$$

Exercise 3.4 Gasoline is stocked in a bulk tank each week at a particular gas station. Let X be the random variable that represents the proportion of the tank's capacity stocked in a given week, and Y be the random variable that represents the proportion of the tank's capacity sold in the same week. Note that the gas station cannot sell more than what was stocked in a given week, which implies that the value of Y cannot exceed the value of X. A possible joint pdf of X and Y is determined as follows

$$
f(x,y) = \begin{cases} 3x, & \text{if } 0 \leq y \leq x \leq 1 \\ 0, & \text{otherwise .} \end{cases}
$$

1- Find the joint CDF of X and Y at the point $(x,y) = (1/2, 1/3)$.
2- What is the probability that the amount of gas sold is less than half the amount that is stocked in a given week.
Solution

1-

$$
\begin{aligned}
F\left(\frac{1}{2}, \frac{1}{3}\right) = P\left(X \leq \frac{1}{2}, Y \leq \frac{1}{3}\right) &= \int_0^{1/3} \int_y^{0.5} 3x dx dy \\
&= \int_0^{1/3} \left(\frac{3}{2} x^2 \Big|_y^{0.5} \right) dy = \int_0^{1/3} \left(\frac{3}{8} - \frac{3}{2} y^2 \right) dy \\
&= \frac{3}{8} y - \frac{1}{2} y^3 \Big|_0^{1/3} \approx 0.1065
\end{aligned}
$$

2- Finding the probability that the amount of gas sold is less than half the amount that is stocked in a given week means that to compute $P(Y < 0.5X)$. Figure 3.3 illustrates the

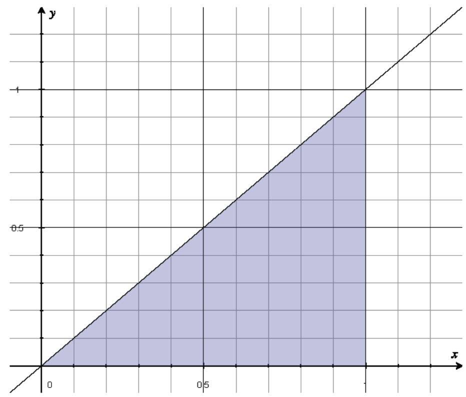


Figure 3.1: Region over which joint pdf $f(x,y)$ is nonzero.

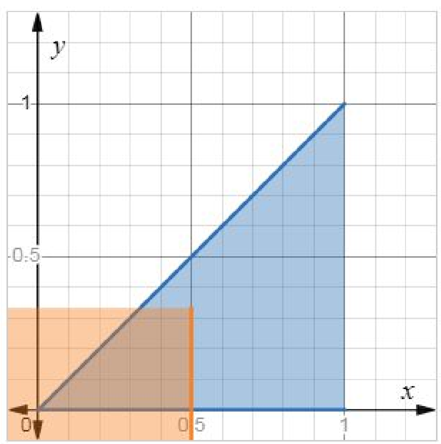


Figure 3.2: Intersection of $\{(x,y) \mid x \leq 1/2, y \leq 1/3\}$ with the region over which joint pdf $f(x,y)$ is nonzero.

intersection of $\{(x,y) \mid y < 0.5x\}$ with the feasible PDF region over which joint pdf $f(x,y)$ is nonzero.

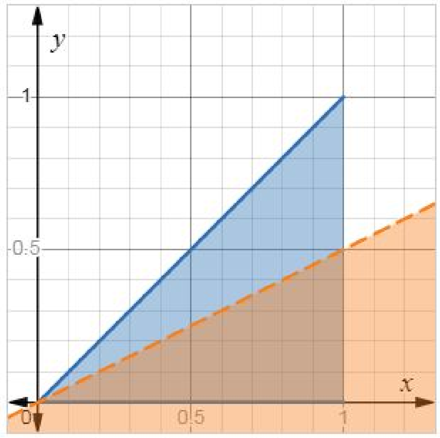


Figure 3.3: Intersection of $\{(x,y) \mid y < 0.5x\}$ with the feasible PDF region over which joint pdf $f(x,y)$ is nonzero.

$$\begin{aligned} P(Y < 0.5X) &= \int_0^1 \int_0^{0.5x} 3xdydx \\ &= \int_0^1 \left(3xy|_0^{0.5x} \right) dx \\ &= \int_0^1 \left(\frac{3}{2}x^2 - 0 \right) dx = \frac{1}{2}x^3 \Big|_0^1 \\ &= \frac{1}{2} \end{aligned}$$

Exercise 3.5 Let X and Y two random variables with joint PMF given in Table3.4 below.
1- Find $P(X \leq 2, Y \leq 4)$.
2- Find the marginal PMF's of X and Y.
3- Find $P(Y = 2|X = 1)$.
4- Are X and Y are independent.

Solution

1-

$$\begin{aligned} P(X \leq 2, Y \leq 4) &= P_{XY}(1,2) + P_{XY}(1,4) + P_{XY}(2,2) + P_{XY}(2,4) \\ &= \frac{1}{12} + \frac{1}{24} + \frac{1}{6} + \frac{1}{12} = \frac{3}{8}. \end{aligned}$$

$f(x,y)$	Y		
X	2	4	5
1	1/12	1/24	1/24
2	1/6	1/12	1/8
3	1/4	1/8	1/12

Table 3.4: Joint PMF of X and Y

2- We have $S_X = \{1,2,3\}$ and $S_Y = \{2,4,5\}$, then

$$P_X(x) = \begin{cases} \frac{1}{6} & x = 1 \\ \frac{3}{8} & x = 2 \\ \frac{11}{24} & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

and

$$P_Y(y) = \begin{cases} \frac{1}{2} & y = 2 \\ \frac{1}{4} & y = 4 \\ \frac{1}{4} & y = 5 \\ 0 & \text{otherwise} \end{cases}$$

3- By the conditional probability formula, we have

$$\begin{aligned} P(Y = 2|X = 1) &= \frac{P(X = 1, Y = 2)}{P(X = 1)} \\ &= \frac{P_{XY}(1,2)}{P_X(1)} \\ &= \frac{\frac{1}{12}}{\frac{1}{6}} = \frac{1}{2}. \end{aligned}$$

4- Since we have

$$P(X = 2, Y = 2) = \frac{1}{6} \neq P(X = 2)P(Y = 2) = \frac{3}{16}.$$

it results that X and Y are not independent.

Exercise 3.6 Let X and Y two random variables with joint PMF given in Table3.5 below.
1- Find $P(X=0, Y \leq 1)$.
2- Find the marginal PMF's of X and Y.
3- Find $P(Y=1|X=0)$.
4- Are X and Y are independent.

$f(x,y)$	Y		
X	0	1	2
0	1/6	1/4	1/8
1	1/8	1/6	1/6

Table 3.5: Joint PMF of X and Y

Solution
1-

$$P(X=0, Y \leq 1) = P_{XY}(0,0) + P_{XY}(0,1) = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}.$$

2- We have $S_X = \{0,1\}$ and $S_Y = \{0,1,2\}$. To find for example $P_X(0)$, we can write

$$\begin{aligned} P_X(0) &= P_{XY}(0,0) + P_{XY}(0,1) + P_{XY}(0,2) \\ &= \frac{1}{6} + \frac{1}{4} + \frac{1}{8} \\ &= \frac{13}{24}. \end{aligned}$$

We use the same technique to compute the other probabilities, we obtain

$$P_X(x) = \begin{cases} \frac{13}{24} & x = 0 \\ \frac{11}{24} & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$P_Y(y) = \begin{cases} \frac{7}{24} & y = 0 \\ \frac{5}{12} & y = 1 \\ \frac{7}{24} & y = 2 \\ 0 & \text{otherwise} \end{cases}$$

3-

$$
\begin{aligned}
P(Y=1|X=0) &= \frac{P(X=0,Y=1)}{P(X=0)} \\
&= \frac{P_{XY}(0,1)}{P_X(0)} \\
&= \frac{\frac{1}{4}}{\frac{13}{24}} = \frac{6}{13}.
\end{aligned}
$$

4- Since we have

$$P(Y=1|X=0) = \frac{6}{13} \neq P(Y=1) = \frac{5}{12}.$$

it results that X and Y are not independent.

Exercise 3.7 Let X and Y be two continuous random variables having the following joint PDF

$$f_{XY}(x,y) = \begin{cases} x + cy^2 & 0 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

1- Find the constant c
2- Find $P\left(0 \leq X, Y \leq \frac{1}{2}\right)$.
3- Find the marginal PMF's of X and Y
4- Find the joint CDF for X and Y

Solution

1- Since we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$$

Thus,

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\
&= \int_0^1\int_0^1 x+cy^2 dxdy \\
&= \int_0^1 \left[\frac{1}{2}x^2+cy^2x\right]_{x=0}^{x=1} dy \\
&= \int_0^1 \frac{1}{2}+cy^2 dy \\
&= \left[\frac{1}{2}y+\frac{1}{3}cy^3\right]_{y=0}^{y=1} \\
&= \frac{1}{2}+\frac{1}{3}c
\end{aligned}$$

Therefore, we obtain $c=\frac{3}{2}$.
2-

$$\begin{aligned}
P\left(0\le X,Y\le\frac{1}{2}\right) &= \int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}}\left(x+\frac{3}{2}y^2\right)dxdy \\
&= \int_0^{\frac{1}{2}}\left[\frac{1}{2}x^2+\frac{3}{2}y^2x\right]_0^{\frac{1}{2}} dy \\
&= \int_0^{\frac{1}{2}}\left(\frac{1}{8}+\frac{3}{4}y^2\right)dy \\
&= \frac{3}{32}
\end{aligned}$$

3- The marginal PMF's of X and Y.
For $0\le x\le 1$, we have

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\
&= \int_0^1\left(x+\frac{3}{2}y^2\right)dy \\
&= \left[xy+\frac{1}{2}y^3\right]_0^1 \\
&= x+\frac{1}{2}
\end{aligned}$$

Thus,

$$f_X(x) = \begin{cases} x + \frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, for $0 \le y \le 1$, we have

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx \\ &= \int_0^1 \left(x + \frac{3}{2}y^2\right) dx \\ &= \left[\frac{1}{2}x^2 + \frac{3}{2}y^2 x\right]_0^1 \\ &= \frac{3}{2}y^2 + \frac{1}{2} \end{aligned}$$

Thus,

$$f_Y(y) = \begin{cases} \frac{3}{2}y^2 + \frac{1}{2} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

4- To find the joint CDF for $x > 0$ and $y > 0$, we need to integrate the joint PDF:

$$\begin{aligned} F_{XY}(x,y) &= \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(u,v)dudv \\ &= \int_0^y \int_0^x f_{XY}(u,v)dudv \\ &= \int_0^{\min(y,1)} \int_0^{\min(x,1)} \left(u + \frac{3}{2}v^2\right) dudv. \end{aligned}$$

For $0 \le x, y \le 1$, we obtain

$$\begin{aligned} F_{XY}(x,y) &= \int_0^y \int_0^x \left(u + \frac{3}{2}v^2\right) dudv \\ &= \int_0^y \left[\frac{1}{2}u^2 + \frac{3}{2}v^2 u\right]_0^x dv \\ &= \int_0^y \left(\frac{1}{2}x^2 + \frac{3}{2}xv^2\right) dv \\ &= \frac{1}{2}x^2 y + \frac{1}{2}xy^3. \end{aligned}$$

For $0 \leq x \leq 1$ and $y \geq 1$, we use the fact that F_{XY} is continuous to obtain

$$\begin{aligned} F_{XY}(x,y) &= F_{XY}(x,1) \\ &= \frac{1}{2}x^2 + \frac{1}{2}x. \end{aligned}$$

Similarly, for $0 \leq y \leq 1$ and $x \geq 1$, we obtain

$$\begin{aligned} F_{XY}(x,y) &= F_{XY}(1,y) \\ &= \frac{1}{2}y + \frac{1}{2}y^3. \end{aligned}$$

Exercise 3.8 Let X and Y be two continuous random variables having the following joint PDF

$$f_{XY}(x,y) = \begin{cases} cx^2y & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

1- Determinate $S_{X,Y}$ and show it in the plane.
2- Determinate the constant c.
3- Determinate the marginal PDFs, $f_X(x)$ and $f_Y(y)$.
4- Determinate $P\left(Y \leq \frac{X}{2}\right)$.
5- Determinate $P\left(Y \leq \frac{X}{4} \middle| Y \leq \frac{X}{2}\right)$.
Solution
We have

$$S_{X,Y} = \left\{(x,y) \in \mathbb{R}^2 \mid 0 \leq y \leq x \leq 1\right\}$$

Figure 3.4 shows S_{XY}.

2- To find the constant c, we can write

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\ &= \int_0^1 \int_0^x cx^2 y dy dx \\ &= \int_0^1 \frac{c}{2}x^4 dx \\ &= \frac{c}{10} \end{aligned}$$

Thus, $c = 10$

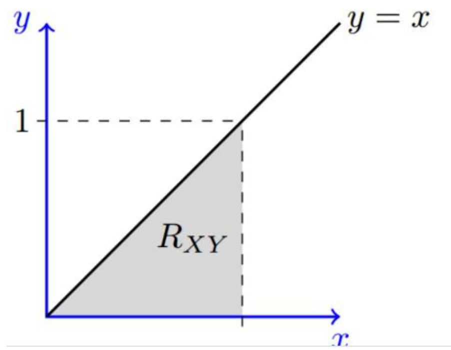


Figure 3.4: Region over which joint pdf $f(x,y)$ is nonzero.

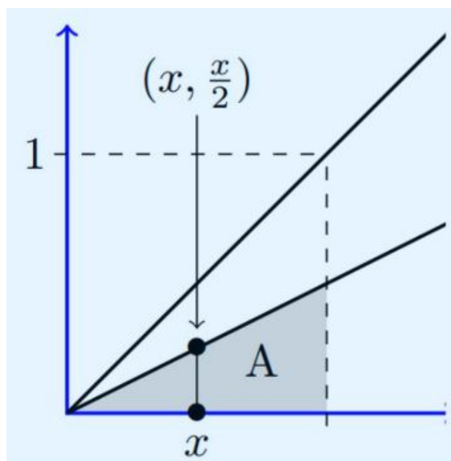


Figure 3.5: Intersection of $\{(x,y) \mid y < 0.5x\}$ with the feasible PDF region over which joint pdf $f(x,y)$ is nonzero.

3- The marginal PMF's of X and Y. For $0 \le x \le 1$, we have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\ &= \int_0^x 10x^2 y dy \\ &= 5x^4 \end{aligned}$$

Thus,

$$f_X(x) = \begin{cases} 5x^4 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, for $0 \le y \le 1$, we set

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx \\ &= \int_{y}^{1} 10x^2 y dx \\ &= \frac{10}{3} y \left(1 - y^3\right) \end{aligned}$$

Thus,

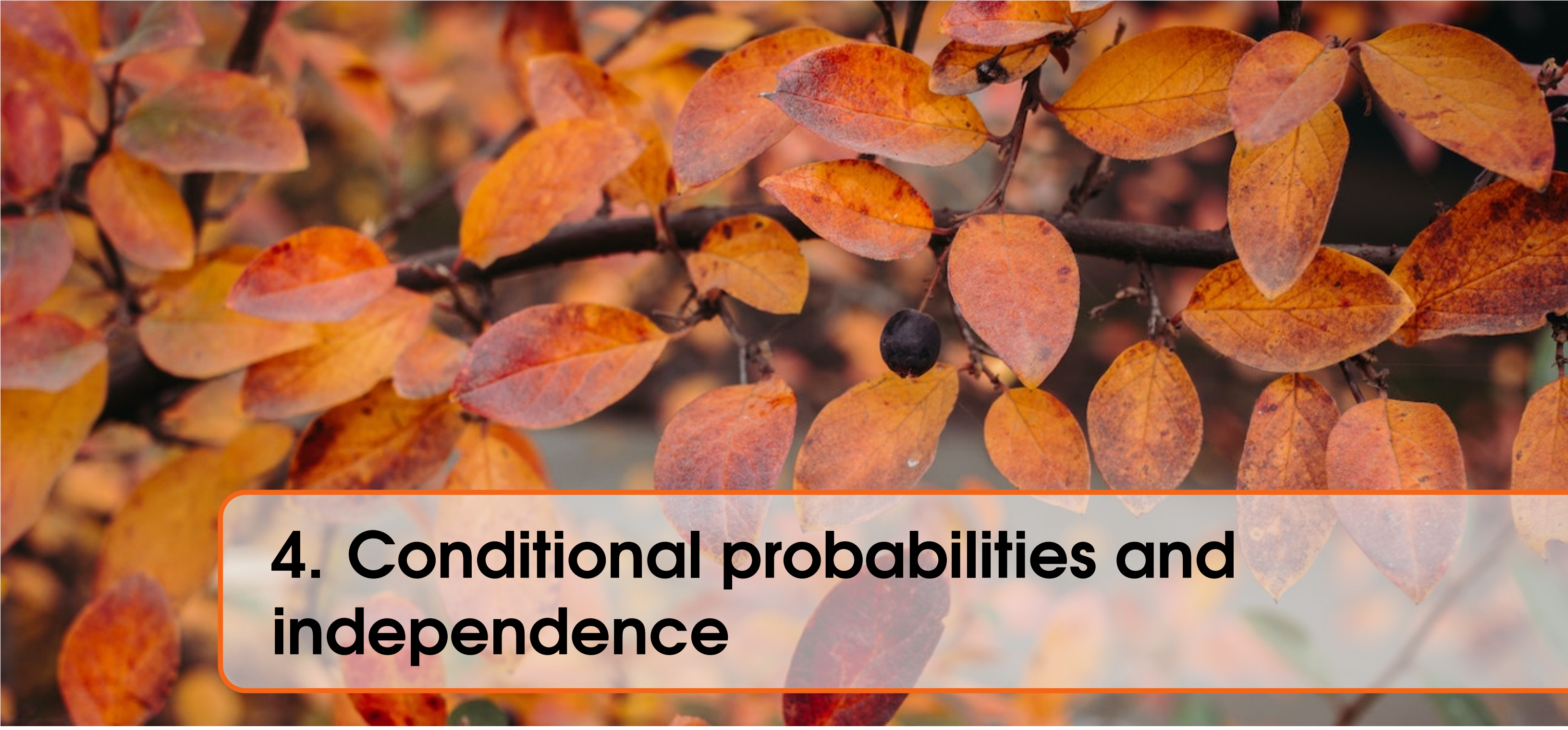
$$f_Y(y) = \begin{cases} \frac{10}{3} y \left(1 - y^3\right) & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

4-

$$\begin{aligned} P\left(Y \le \frac{X}{2}\right) &= \int_{-\infty}^{\infty} \int_{0}^{\frac{x}{2}} f_{XY}(x,y)dydx \\ &= \int_{0}^{1} \int_{0}^{\frac{x}{2}} 10x^2 ydydx \\ &= \int_{0}^{1} \frac{5}{4} x^4 dx \\ &= \frac{1}{4} \end{aligned}$$

5-

$$\begin{aligned} P\left(Y \le \frac{X}{4} \middle| Y \le \frac{X}{2}\right) &= \frac{P\left(Y \le \frac{X}{4}, Y \le \frac{X}{2}\right)}{P\left(Y \le \frac{X}{2}\right)} \\ &= 4P\left(Y \le \frac{X}{4}\right) \\ &= 4\int_{0}^{1} \int_{0}^{\frac{x}{4}} 10x^2 ydydx \\ &= 4\int_{0}^{1} \frac{5}{16} x^4 dx \\ &= \frac{1}{4} \end{aligned}$$



4. Conditional probabilities and independence

In this chapter, we define the probability distribution of one random variable given information about another random variable. This type of conditional distribution relies on the joint distribution of the two random variables introduced earlier. We begin with discrete random variables and then move on to the continuous case.

4.1 Conditional distribution of a discrete random variable

Recall that for any two events A and B such that $P(B) > 0$, the conditional probability is defined as follows

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

We use this same concept for events to define conditional probabilities for random variables.

Definition 4.1 Let X and Y be discrete random variables and $p(x,y)$ is the joint probability mass function. The conditional probability mass function of X, given that $Y = y$ is defined as

$$p_{X|Y}(x|y) = P(X = x|Y = y)$$
$$= \begin{cases} \frac{p_{X,Y}(x,y)}{p_Y(y)} & \text{if } p_Y(y) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

where $p_Y(y)$ is the marginal PMF of Y for any x and y such that $p_Y(y) > 0$

Definition 4.2 (**Conditional CDF of discrete random variable**) Let X and Y be discrete random variables, the conditional cumulative distribution function of X given $Y = y$ is

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{x' \leq x} p_{X|Y}(x'|y).$$

4.1.1 Properties of the conditional probability mass function

-The conditional probability mass function for X, given $Y = y$, for a fixed y, is a PMF satisfying the following:

$$0 \leq p_{X|Y}(x|y) \leq 1 \qquad \text{and} \qquad \sum_x p_{X|Y}(x|y) = 1,$$

indeed,

$$\sum_x p_{X|Y}(x|y) = \sum_x P(X = x|Y = y)$$
$$= \sum_x \frac{p_{X,Y}(x,y)}{p_Y(y)}$$
$$= \frac{1}{p_Y(y)} \sum_x p_{X,Y}(x,y)$$
$$= \frac{1}{p_Y(y)} p_Y(y)$$
$$= 1.$$

- Typically, the conditional distribution of X given Y is different from the conditional distribution of Y given X, i.e.

$$p_{X|Y}(x|y) \neq p_{Y|X}(y|x).$$

- If X and Y are independent, discrete random variables, then:

$$p_{X|Y}(x|y) = p_X(x)$$
$$p_{Y|X}(y|x) = p_Y(y)$$

■ **Example 4.1** We have investigated the correlation between hair and eye color among randomly selected Saint Mary's students. Here is the joint probability mass function that we have obtained, along with the marginal PMF's provided in the margins:

$p(x,y)$	Hair Color (X)				
Eye Color (Y)	blonde (1)	red (2)	brown (3)	black (4)	$p_Y(y)$
blue (1)	0.12	0.05	0.12	0.01	0.30
green (2)	0.12	0.07	0.09	0	0.28
brown (3)	0.16	0.07	0.16	0.03	0.42
$p_X(x)$	0.40	0.19	0.37	0.04	1.00

1- Find the portion of SMC students with blue eyes having red hair.
2- Find portion of SMC students with blonde hair having green eyes.
solution
1 - First, let's determine $p_{X|Y}(2 \mid 1)$:

$$p_{Y|X}(1 \mid 2) = \frac{p_{Y,X}(1,2)}{p_Y(2)} = \frac{0.05}{0.19} = \frac{5}{19} \approx 0.167$$

There are approximately 16.7% of students with blue eyes having red hair.
2 - Now, let's reverse the order of X and Y, and find $p_{Y|X}(2 \mid 1)$:

$$p_{X|Y}(1 \mid 2) = \frac{p_{X,Y}(1,2)}{p_Y(2)} = \frac{0.12}{0.28} = \frac{3}{7}$$

There are approximately 30% of SMC students with blonde hair having green eyes.

■ **Example 4.2** Suppose the joint PMF of X and Y is given by the following probability table.

X/Y	y = 0	y = 1	y = 2	y = 3
x = 0	0	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{3}{42}$
x = 1	$\frac{2}{42}$	$\frac{3}{42}$	$\frac{4}{42}$	$\frac{5}{42}$
x = 2	$\frac{4}{42}$	$\frac{5}{42}$	$\frac{6}{42}$	$\frac{7}{42}$

- Determine the conditional PMF of Y given $X=1$.

Solution

We have

$$P_{Y|X}(y \mid x=1) = \frac{p_{X,Y}(x=1,y)}{p_X(x=1)} = \frac{p_{X,Y}(x=1,y)}{14/42}$$

Hence, the conditional PMF of Y given $X=1$ is as follows

$$P_{Y|X}(y \mid x=1) = \begin{cases} \frac{2/42}{14/42} = \frac{2}{14}, & \text{if } y=0 \\ \frac{3/42}{14/42} = \frac{3}{14}, & \text{if } y=1 \\ \frac{4/42}{14/42} = \frac{4}{14}, & \text{if } y=2 \\ \frac{5/42}{14/42} = \frac{5}{14}, & \text{if } y=3 \end{cases}$$

4.2 Conditional distributions of continuous random variables

The following definition gives the formulas for conditional distributions of continuous random variables by simply replacing sums with integrals and PMF's with PDF's

Definition 4.3 **(Conditional PDF)** If X and Y are continuous random variables with joint probability density function (PDF) given by $f(x,y)$, is the joint probability mass function. The conditional probability density function of X, given that $Y=y$, is defined by

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}, & \text{if } f_Y(y) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 4.4 **(Conditional CDF of continuous random variable)** If X and Y are continuous random variables, the conditional cumulative distribution function of X given $Y=y$ is

$$F_{X|Y}(x|y) = P(X \le x|Y=y) = \int_{-\infty}^{x} f_{X|Y}(t|y)dt.$$

4.2.1 Properties of conditional probability density function

1. The conditional PDF for X, given $Y = y$, for a fixed y, is a PDF satisfying the following:

$$0 \leq f_{X|Y}(x \mid y) \quad \text{and} \quad \int_{\mathbb{R}} f_{X|Y}(x \mid y) dx = 1$$

2. Typically, the conditional distribution of X given Y does not equal the conditional distribution of Y given X, i.e.,

$$f_{X|Y}(x \mid y) \neq f_{Y|X}(y \mid x)$$

3. If X and Y are independent, continuous random variables, then the following are true:

$$f_{X|Y}(x \mid y) = f_X(x)$$
$$f_{Y|X}(y \mid x) = f_Y(y)$$

4.3 Conditional expectation and variance

Definition 4.5 Let X and Y discrete random variables, then the conditional expected value of X, given $Y = y$, is given by

$$\mu_{X|Y=y} = \mathrm{E}[X \mid Y = y] = \sum_{x} x p_{X|Y}(x \mid y)$$

and the conditional variance of X, given $Y = y$, is given by

$$\sigma^2_{X|Y=y} = \mathrm{Var}(X \mid Y = y) = \mathrm{E}\left[\left(X - \mu_{X|Y=y}\right)^2 \mid Y = y\right] = \sum_{x \in S_X} \left(x - \mu_{X|Y=y}\right)^2 p_{X|Y}(x \mid y)$$
$$= \mathrm{E}\left[X^2 \mid Y = y\right] - \mu^2_{X|Y=y} = \left(\sum_{x \in S_X} x^2 p_{X|Y}(x \mid y)\right) - \mu^2_{X|Y=y}.$$

Similarly, if X and Y are continuous random variables with joint PDF given by $f(x,y)$, then the conditional expected value of X, given $Y = y$, is

$$\mathrm{E}[X \mid Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x \mid y) dx$$

and the conditional variance of X, given $Y = y$, is

$$\begin{aligned}\operatorname{Var}(X \mid Y = y) &= \mathrm{E}\left[X^2 \mid Y = y\right] - (\mathrm{E}[X \mid Y = y])^2 \\ &= \int_{\mathbb{R}} x^2 f_{X|Y}(x \mid y)dx - \left(\int_{\mathbb{R}} x f_{X|Y}(x \mid y)dx\right)^2.\end{aligned}$$

4.4 Solved exercises

Exercise 4.1 Suppose that the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} e^{-\left(\frac{x}{y}+y\right)} y^{-1} & 0 < x, y \\ 0 & \text{otherwise} \end{cases}$$

For $y > 0$, determine:
1- $P(X > 1 \mid Y = y)$;
2- $E[X \mid Y = y]$.
Solution
1- For $y > 0$, we have

$$\begin{aligned}f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_0^{\infty} e^{-\left(\frac{x}{y}+y\right)} y^{-1} dx \\ &= e^{-y}\end{aligned}$$

Hence, for $y > 0$,

$$\begin{aligned}f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \begin{cases} e^{-x/y} y^{-1} & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}\end{aligned}$$

Thus,

$$\begin{aligned}P(X > 1 \mid Y = y) &= \int_1^{\infty} f_{X|Y}(x \mid y)dx \\ &= \int_1^{\infty} e^{-x/y} y^{-1} dx \\ &= e^{-1/y}\end{aligned}$$

2-

$$\begin{aligned} E[X \mid Y = y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx \\ &= \int_{0}^{\infty} \frac{x}{y} e^{-x/y} dx \\ &= y \end{aligned}$$

Exercise 4.2 Let X and Y be continuous random variables having the following joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 24x(1-x-y), & \text{if } x,y \geq 0, x+y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

1- Determine the conditional PDF of X given $Y = y$.
2- Determine the conditional PDF of X given $Y = \frac{1}{2}$.
Solution
1- We have the marginal PDf of Y is

$$f_Y(y) = \begin{cases} 4(1-y)^3, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} f_{X|Y}(x \mid y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \begin{cases} \frac{24x(1-x-y)}{4(1-y)^3}, & \text{if } x,y \geq 0, x+y \leq 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

2-

$$\begin{aligned} f_{X|Y}(x|0.5) &= \frac{f_{X,Y}(x,0.5)}{f_Y\left(\frac{1}{2}\right)} \\ &= \begin{cases} \frac{24x(0.5-x)}{4(0.5)^3} = 48x(0.5-x), & \text{if } 0 \leq x \leq 0.5 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Exercise 4.3 Let consider the context of Exercise 3,of the previous chapter, where X gave the amount of gas stocked and Y gave the amount of gas sold at a given.
1- Find the conditional PDF for the amount of gas sold in a given week, when only half of the tank was stocked; in other word, find the conditional pdf of Y given that $X = 0.5$.
2- Compute the conditional mean value of Y, given that $X = 0.5$

Solution
1- The marginal PDF for X is:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_0^x 3xdy = 3xy|_0^x = 3x^2, \quad \text{for } 0 \leq x \leq 1$$

Hence, if $X = 0.5$, then $f_X(0.5) = 3(0.5)^2 = 0.75$, and the conditional PDF of Y in this case is

$$f_{Y|X}(y \mid 0.5) = \frac{f(0.5,y)}{f_X(0.5)} = \frac{3(0.5)}{0.75} = 2, \quad \text{for } 0 \leq y \leq 0.5$$

2-

$$\mathrm{E}[Y \mid X = 0.5] = \int_0^{0.5} y f_{Y|X}(y \mid 0.5)dy = 0.25.$$

This means that, when 50% of the tank is stocked, we expect that 25% will be sold.

Exercise 4.4 Suppose that the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} 3y(x+\frac{1}{4}y) & 0 \leq x,y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

1- Find the conditional PDF of X given $Y = y$, for $y \in (0,1]$.
2- Determine $E[X|Y = y]$, for $y \in (0,1]$.
3- Find $P(X > \frac{1}{2}|Y = 1)$
Solution
1-

$$\begin{aligned} f_Y(y) &= \int_0^1 3y\left(x+\frac{1}{4}y\right)dx = \left[3y\left(\frac{1}{2}x^2+\frac{1}{4}xy\right)\right]_0^1 \\ &= 3y\left(\frac{1}{2}+\frac{1}{4}y\right) = \frac{3}{4}y(2+y). \end{aligned}$$

Therefore for $y \in (0,1]$,

$$\begin{aligned} f_{X|Y}(x|y) &= \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)}, & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} \frac{3y(x+\frac{1}{4}y)}{\frac{3}{4}y(2+y)}, & 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases} \\ &= \begin{cases} \frac{4x+y}{2+y}, & 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

2-

$$
\begin{aligned}
E[X|Y=y] &= \int_0^1 x f_{X|Y}(x|y)dx \\
&= \int_0^1 x\frac{4x+y}{2+y}dx = \frac{1}{2+y}\int_0^1 x(4x+y)dx \\
&= \frac{1}{2+y}\left[\frac{4}{3}x^3+\frac{1}{2}x^2y\right]_0^1 = \frac{1}{2+y}\left(\frac{4}{3}+\frac{1}{2}y\right) \\
&= \frac{8+3y}{6(2+y)}.
\end{aligned}
$$

3-

$$
\begin{aligned}
P\left(X>\frac{1}{2}\middle|Y=1\right) &= \int_{1/2}^1 f_{X|Y}(x|y=1)dx \\
&= \frac{1}{3}\int_{1/2}^1 (4x+1)dx \\
&= \frac{1}{3}\left[2x^2+x\right]_{1/2}^1 \\
&= \frac{1}{3}\left(2+1-\frac{2}{4}-\frac{1}{2}\right) = \frac{2}{3}.
\end{aligned}
$$

Bibliography

[1] Borkar, V. S. (2012). *Probability theory: An advanced course*. Springer.

[2] Devore, J. L. (2000). *Probability and statistics*. Brooks/Cole.

[3] DeCoursey, W. (2003). *Statistics and probability for engineering applications*. Elsevier.

[4] Johnson, R. A., Miller, I., & Freund, J. E. (2011). *Probability and statistics for engineers*. Prentice Hall.

[5] Kuter, K. (2023). *Math 345 - Probability*. LibreTexts.

[6] Pfeiffer, P. E. (2012). *Probability for applications*. Springer.

[7] Sokol, A., & Rønn-Nielsen, A. (2013). *Advanced probability*. Department of Mathematical Sciences, University of Copenhagen.

[8] Sousi, P. (2013). *Advanced probability*. University of Cambridge.