



Polycopié expertisé et validé par
les instances scientifiques

Introduction to Probability



Course for Students in Third year undergraduate

Mathematics & Informatics

Author: ABDELGHANI HAMAZ

Establishment: Faculty of Sciences. Mouloud Mammeri University of Tizi-Ouzou

Date: December 2023

FOREWORDS

★ ★ ★ ★★

The lecture notes and exercises contained in this manuscript are part of the five years of teaching given to students in the third year of the undergraduate specialisation in Mathematics and Informatics.

No prior knowledge of probability or statistics is necessary for the reading of this document. The mathematics required is a full year of classical calculus, including series, integration and linear algebra.

This collection, in line with the official syllabus of the above-mentioned courses, covers all the fundamental aspects of the introduction to probabilities. Particular attention has been paid to form, emphasising a simple presentation of the various concepts, supplemented where necessary by illustrations, comments and application examples. There are also a number of exercises, mainly on the basis of the documents in the bibliography.

There are four chapters in this document. All contain definitions, theorems, demonstrations and proofs based on the main points of the course. At the end of each chapter, there is a series of exercises with corrections, in order to consolidate the knowledge which has been acquired.

The choice of the exercises is based on the objectives of the course: to make the student familiar with the basic and fundamental concepts of probability calculus.

★ ★ ★ ★★

Contents

1	Probability	1
1.1	Random experiments	1
1.2	Events	2
1.3	Algebra of Sets	3
1.4	Probability Space.	7
1.5	Probability space	8
1.6	Equally Likely Events	11
1.7	Conditional Probability	12
1.8	Bayes' Rule	12
1.9	Total Probability	12
1.10	Independent Events	12
1.11	Chapter Summary	13
1.12	Exercises	14
2	Random variables	20
2.1	Definitions	20
2.2	Distribution Functions	21
2.3	Density Functions	23
2.4	Continuous Random Variables	24
2.5	Mean And Variance	26
2.6	Moments and generating function	30
2.7	Some Special Discrete Distributions	33
2.8	Continous Distributions	37
2.9	Comments	41
2.10	Exercises	42
3	Vector Random variables	49
3.1	Introduction	49
3.2	Bivariate distribution	49
3.3	Random pairs with density	62
3.4	Exercises	65

4	Convergence of random variables	78
4.1	Introduction	78
4.2	Convergence in Probability	78
4.3	Convergence in r^{th} mean, $r > 0$	81
4.4	Convergence in Distribution	83
4.5	almost sure convergence	84
4.6	Relations between convergence modes	84
4.7	Exercises	85
	Statistical Tables	90
	Bibliography	105

Chapter Probability

Introduction

- Event
- Probability
- Conditional probability
- Bayes' Theorem f

One of the fundamental tools of statistics is probability, which had its formal beginning with games of chance in the seventeenth century. The theory of probability deals with averages of mass phenomena occurring sequentially or simultaneously: electron mass, telephone calls, radar detection, quality control, games of chance, statistical mechanics, noise, birth and death rates, and queueing theory, among many others. The purpose of this chapter is to define *probability* and discuss some of its properties. Also, the basic concepts of probability theory are presented. It is a brief essay on some of all different meanings that have been attached to probability and may be omitted by those who are interested only in mathematical (axiomatic) probability.

1.1 Random experiments

Definition 1.1. Random Experiments

In the study of probability, any process of observation is referred to as an experiment. The results of an observation are called the outcomes of the experiment. An experiment is called a random experiment if

- *its outcome cannot be predicted and if it has more than one possible outcomes,*
- *if this experiment is repeated under the same conditions the outcome may be different*
- *the set of outcomes is known in advance.*



Typical examples of a random experiment are the roll of a die, the toss of a coin, drawing a card from a deck, or selecting a message signal for transmission from several messages.

Definition 1.2. Sample Space

The set of all possible outcomes of a random experiment is called the sample space (or universal set connected with that experiment), and it is denoted by Ω . An element in Ω is called a sample point. Each outcome of a random experiment corresponds to a sample point.

When we repeat a random experiment several times, we call each one of them a trail.



Example 1.1 Find the sample space for the experiment of tossing a coin (a) and (b) twice.

a. There are two possible outcomes, heads or tails. Thus:

$$\Omega = \{H, T\}$$

where H and T represent head and tail, respectively.

b. There are four possible outcomes. There are pairs of heads and tails. Thus:

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

1.2 Events

Our definition of sample space is precise and satisfactory, whereas our definitions of event and event space are not entirely given. Rather than developing the necessary mathematics to precisely define which subsets of Ω constitute our event space \mathcal{A} . An event is a subset of the sample space. Let us state some properties of \mathcal{A} that it seems reasonable to require:

- (i) $\Omega \in \mathcal{A}$.
- (ii) If $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
- (iii) If A_1 and $A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$.

We said earlier that we were interested in events mainly because we would be interested in probability that an event happens. Surely, then, we would want \mathcal{A} to include Ω , the sure event.

Any collection of events with properties (i) and (iii) is called a **Boolean algebra** or just **algebra**, of events. We might note that the collection of all subsets of Ω necessarily satisfies the above properties. Several results follow from the above assumed properties of \mathcal{A} .

Since we have identified a sample space \mathcal{A} as the set of all possible outcomes of a random experiment, we will review some set notations in the following.

If ω is an element of Ω (or belongs to Ω), then we write

$$\omega \in \Omega$$

If ω is not an element of Ω (or does not belong to Ω), then we write

$$\omega \notin \Omega$$

A set A is called a subset of B , denoted by

$$A \subset B$$

if every element of A is also an element of B . Any subset of the sample space S is called an event. A sample point of ω is often referred to as an elementary event. Note that the sample space Ω is the subset of itself: that is, $\Omega \subset \Omega$. Since Ω is the set of all possible outcomes, it is often called the certain event.

1.3 Algebra of Sets

1. Equality. Two sets A and B are equal, denoted $A = B$, if and only if $A \subset B$ and $B \subset A$.

2. Complement. Suppose $A \subset \Omega$. The complement of set A , denoted \bar{A} , is the set containing all elements in Ω but not in A .

$$\bar{A} = \{\omega : \omega \in \Omega \text{ and } \omega \notin A\}$$

3. Union. The union of sets A and B , denoted $A \cup B$, is the set containing all elements in either A or B or both.

$$A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$$

4. Intersection. The intersection of sets A and B , denoted $A \cap B$, is the set containing all elements in both A and B .

$$A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$$

5. Difference. The difference of sets A and B , denoted $A \setminus B$, is the set containing all elements in A but not in B .

$$A \setminus B = \{\omega : \omega \in A \text{ and } \omega \notin B\}$$

Note that $A \setminus B = A \cap \bar{B}$.

6. Symmetrical Difference. The symmetrical difference of sets A and B , denoted $A \Delta B$, is the set of all elements that are in A or B but not in both.

$$A \Delta B = \{\omega : \omega \in A \text{ or } \omega \in B \text{ and } \omega \notin A \cap B\}$$

Note that $A \Delta B = (A \cap \bar{B}) \cup (\bar{A} \cap B) = (A \setminus B) \cup (B \setminus A)$.

7. Null Set. The set containing no element is called the null set, denoted \emptyset . Note that

$$\emptyset = \bar{\Omega}$$

8. Disjoint Sets. Two sets A and B are called disjoint or mutually exclusive if they contain no common element, that is, if $A \cap B = \emptyset$.

The definitions of the union and intersection of two sets can be extended to any finite number of sets as follows:



Extension

$$\begin{aligned} \bigcup_{i=1}^n A_i &= A_1 \cup A_2 \cup \cdots \cup A_n \\ &= \{\omega : \omega \in A_1 \text{ or } \omega \in A_2 \text{ or } \cdots \omega \in A_n\} \\ \bigcap_{i=1}^n A_i &= A_1 \cap A_2 \cdots \cap A_n \\ &= \{\omega : \omega \in A_1 \text{ and } \omega \in A_2 \text{ and } \cdots \omega \in A_n\} \end{aligned}$$

Note that these definitions can be extended to an infinite number of sets:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \cup \dots$$

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap A_3 \cap \dots$$

In our definition of event, we state that every subset of S is an event, including S and the null set \emptyset . Then

$$S = \text{the certain event}$$

$$\emptyset = \text{the impossible event}$$

If A and B are events in S , then

$$\bar{A} = \text{the event that } A \text{ did not occur}$$

$$A \cup B = \text{the event that either } A \text{ or } B \text{ or both occurred}$$

$$A \cap B = \text{the event that both } A \text{ and } B \text{ occurred}$$



Generalization

Similarly, if A_1, A_2, \dots, A_n are a sequence of events in S , then

$$\bigcup_{i=1}^n A_i = \text{the event that at least one of the } A_i \text{ occurred}$$

$$\bigcap_{i=1}^n A_i = \text{the event that all of the } A_i \text{ occurred}$$

9. Partition of Ω . If $A_i \cap A_j = \emptyset$ for $i \neq j$ and , then the collection $\{A_i; 1 \leq i \leq n\}$ is said to form a partition of Ω .

10. Size of Set. When sets are countable, the size (or cardinality) of set A , denoted $|A|$, is the number of elements contained in A . When sets have a finite number of elements, it is easy to see that size has the following properties:

Proposition 1.1

- (i) if $A \cap B = \emptyset$, then $|A \cup B| = |A| + |B|$.
- (ii) $|\emptyset| = 0$.
- (iii) If $A \subset B$, then $|A| \leq |B|$.
- (iv) $|A \cup B| + |A \cap B| = |A| + |B|$.

Note that the property (iv) can be easily seen if A and B are subsets of a line with length $|A|$ and $|B|$, respectively.

11. Product of Sets. The product (or Cartesian product) of sets A and B , denoted by $A \times B$, is the set of ordered pairs of elements from A and B .

$$C = A \times B = \{(a, b) : a \in A, b \in B\}$$

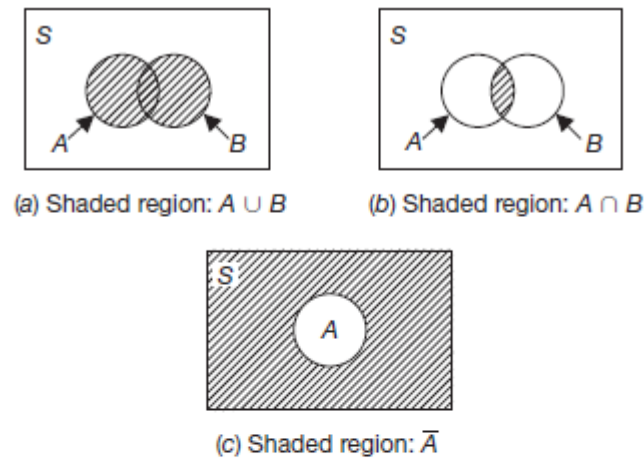


Fig. 1-1

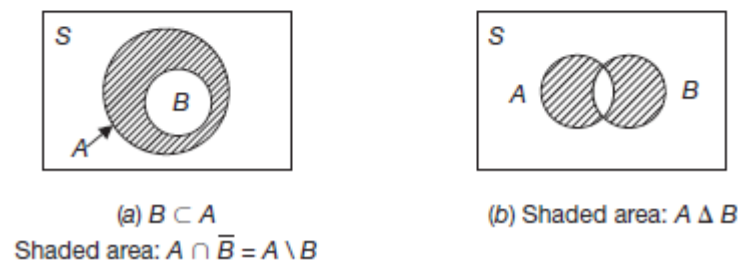


Figure 1.1

Note that

$$A \times B \neq B \times A, \text{ and } |C| = |A \times B| = |A| \times |B|.$$

Example 1.2 Let $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2\}$. Then

$$C = A \times B = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2), (a_3, b_1), (a_3, b_2)\}$$

$$D = B \times B = \{(b_1, a_1), (b_1, a_2), (b_1, a_3), (b_2, a_1), (b_2, a_2), (b_2, a_3)\}$$

1.3.1 Venn Diagram.

A graphical representation that is very useful for illustrating set operation is the Venn diagram. For instance, in the three Venn diagrams shown in Fig. 1-1, the shaded areas represent, respectively, the events $A \cup B$, $A \cap B$, and \bar{A} .

The Venn diagram in Fig. 1-2(a) indicates that $B \subset A$, and the event $A \cap \bar{B}$ is shown as the shaded area. In Fig. 1-2(b), the shaded area represents the event $A \Delta B$.

1.3.2 Identities.

By the above set definitions or reference to Fig. 1-1, we obtain the following identities:

$$\begin{aligned}\bar{S} &= \emptyset \\ \bar{\emptyset} &= S \\ \bar{A} &= A \\ S \cup A &= S \\ A \cup \bar{A} &= S \\ A \cap \bar{A} &= \emptyset \\ A \cup \emptyset &= A \\ A \cap \emptyset &= \emptyset \\ A \setminus B &= A \cap \bar{B} \\ S \setminus A &= \bar{A} \\ A \setminus \emptyset &= A \\ A \Delta B &= (A \cap \bar{B}) \cup (\bar{A} \cap B)\end{aligned}$$

The union and intersection operations also satisfy the following laws ;

Commutative Laws.

$$\begin{aligned}A \cup B &= B \cup A \\ A \cap B &= B \cap A\end{aligned}$$

Associative Laws.

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

Distributive Laws.

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



De Morgan's Laws

$$\begin{aligned}(A \bar{\cup} B) &= \bar{A} \cap \bar{B} \\ (A \bar{\cap} B) &= \bar{A} \cup \bar{B}\end{aligned}$$

These relations are verified by showing that any element that is contained in the set on the left side of the equality sign is also contained in the set on the right side, and vice versa. One way of showing this is by means of a Venn diagram. The distributive laws can be extended as follows:

Distributive laws Extension

$$A \cap \left(\bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i)$$

$$A \cup \left(\bigcap_{i=1}^n B_i \right) = \bigcap_{i=1}^n (A \cup B_i)$$

Similarly, De Morgan's laws also can be extended as follows (Prob.):

De Morgan's Laws Extension

$$\overline{\left(\bigcup_{i=1}^n A_i \right)} = \bigcap_{i=1}^n \overline{A_i}$$

$$\overline{\left(\bigcap_{i=1}^n A_i \right)} = \bigcup_{i=1}^n \overline{A_i}$$

1.4 Probability Space.

Definition 1.3. Event and Event space

We have defined that events are subsets of the sample space S . In order to be precise, we say that a subset A of S can be an event if it belongs to a collection \mathcal{F} of subsets of S , satisfying the following conditions:

- i) $\Omega \in \mathcal{A}$.
- ii) if $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$.
- iii) if $A_i \in \mathcal{A} \forall i \geq 1$, the $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The collection \mathcal{A} is called an event space. In mathematical literature, event space is known as σ -field sigma field or σ -algebra.

Using the above conditions, we can show that if A and B are in \mathcal{A} , then so are

$$A \cap B, A \setminus B, A \Delta B.$$



We said earlier that we were interested in events mainly because we would be interested in the probability that an event happens. Then, surely, we would want \mathcal{A} to include Ω , the sure event. Also, if A is an event, meaning we can talk about the probability that A occurs, then \bar{A} should also be an event so that we can talk about the probability that A does not occur. Similarly, if A_1 and A_2 are events, so should $A_1 \cup A_2$ be an event.

Example1.3 The experiment is the tossing of a single die (a regular six-sided polyhedron or cube marked on each face with one to six spots) and noting which face up. Now the die can land with any one of the six faces up; so there are six possible outcomes of the experiment;

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

Example 1.4 The experiment is to record the number of traffic deaths in the state of Paris next year. Any nonnegative integer is a conceivable outcome of this experiment; so

$\Omega = \{0, 1, 2, \dots\}$. $A = \{\text{fewer than 500 deaths}\} = \{0, 1, \dots, 499\}$ is an event. $A_i = \{\text{exactly } i \text{ deaths}\}$ is an element event.

Example 1.5 Select a light bulb, and record the time in hours that it burns before burning out. Any nonnegative number is a conceivable outcome of this experiment; so $\Omega = \{x : x \geq 0\}$. For this sample space not all subsets of Ω are events; however, any subset that can be exhibited will be an event. For example, let

$$\begin{aligned} A &= \{\text{bulb burns for at least } k \text{ hours but burns out before } m \text{ hours}\} \\ &= \{x : k \leq x < m\}; \end{aligned}$$

then A is an event for any $0 \leq k < m$.

Theorem 1.1

$$\emptyset \in \mathcal{A}$$

Proof By property (i), $\Omega \in \mathcal{A}$, by (ii) $\bar{\Omega} \in \mathcal{A}$; but $\bar{\Omega} = \emptyset$; so $\emptyset \in \mathcal{A}$ □ ♡

Theorem 1.2

If A_1 and $A_2 \in \mathcal{A}$, then $A_1 \cap A_2 \in \mathcal{A}$.

Proof \bar{A}_1 and $\bar{A}_2 \in \mathcal{A}$; hence $\bar{A}_1 \cup \bar{A}_2 \in \mathcal{A}$, and $\overline{(\bar{A}_1 \cup \bar{A}_2)} \in \mathcal{A}$, but $\overline{(\bar{A}_1 \cup \bar{A}_2)} = \bar{A}_1 \cap \bar{A}_2 = A_1 \cap A_2$ by De Morgan's law. □ ♡

Theorem 1.3

If $A_1, A_2, \dots, A_n \in \mathcal{A}$ then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

Proof Follows by induction. □ ♡

We will always assume that our collection of events \mathcal{A} is an algebra which partially justifies our use of \mathcal{A} as our notation for it. In practice, one might take that collection of events of interest in a given consideration and enlarge the collection of events of interest in a given consideration and enlarge the collection, if necessary, to include (i) the sure event, (ii) all complements of events already included, and (iii) all finite unions and intersections of events already included, and this will be an algebra \mathcal{A} . Thus, we have not explained why \mathcal{A} can not always be taken to be the collection of all subsets of Ω . Such an explanation will be given when we define probability in the next subsection.

1.5 Probability space

Here, we give the axiomatic definition of probability. Although this formal definition of probability will not in itself allow us to achieve our goal of assigning actual probabilities to events consisting

of certain outcomes of random experiments, it is another in a series of definitions that will ultimately lead to the goal. Since probability, as well as forthcoming concepts, is defined as a particular function.

A. Classical Definition.

Definition 1.4

Consider an experiment with equally likely outcomes. Then the classical definition of probability of event A , denoted $P(A)$, is defined by

$$P(A) = \frac{|A|}{|\Omega|}$$

If A and B are disjoint, i.e. $A \cap B = \emptyset$, then $|A \cup B| = |A| + |B|$. Hence, in this case

$$P(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{|A| + |B|}{|\Omega|} = \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} = P(A) + P(B)$$

We also have

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

$$P(\bar{A}) = \frac{|\bar{A}|}{|\Omega|} = \frac{|\Omega| - |A|}{|\Omega|} = 1 - \frac{|A|}{|\Omega|} = 1 - P(A)$$



B. Relative Frequency Definition.

Suppose that the random experiment is repeated n times. If event A occurs $n(A)$ times, then the probability of event A , denoted $P(A)$, is defined as

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$$

where $n(A)/n$ is called the relative frequency of event A . Note that this limit may not exist, and in addition, there are many situations in which the concepts of repeatability may not be valid. It is clear that for any event A , the relative frequency of A will have the following properties:

1. $0 \leq n(A) \leq n$, where $n(A)/n = 0$ if A occurs in none of the n repeated trials and $n(A)/n = 1$ if A occurs in all of the n repeated trials.
2. If A and B are mutually exclusive events, then

$$n(A \cup B) = n(A) + n(B)$$

and

$$\frac{n(A \cup B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n}$$

$$P(A \cup B) = \lim_{n \rightarrow \infty} \frac{n(A \cup B)}{n} = \lim_{n \rightarrow \infty} \frac{n(A)}{n} + \lim_{n \rightarrow \infty} \frac{n(B)}{n} = P(A) + P(B)$$

C. Axiomatic Definition. Consider a probability space (S, F, P) . Let A be an event in F . Then in the axiomatic definition, the probability $P(A)$ of the event A is a real number assigned to A which

satisfies the following three axioms:

$$\text{Axiom 1 : } P(A) \geq 0$$

$$\text{Axiom 2 : } P(S) = 1$$

Axiom 3 : If A_1, A_2, \dots is a sequence of mutually exclusive events in Ω

$$\text{and if } A_1 \cup A_2 \cup \dots \cup A = \bigcup_{i=1}^{\infty} A_i \in \Omega,$$

$$\text{then } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

These axioms are certainly motivated by the definition of classical and frequency probability. This definition of probability is a mathematical definition; it tells us which set functions can be called probability functions; it does not tell us what value the probability function $P(\cdot)$ assigns to a given event A . We will have to model our random experiment in some way in order to obtain values for the probability of events.

$P(A)$ is read : "the probability of event A " or "the probability that event A occurs", which means the probability that any outcome in A occurs.

We are able to deduce a number of properties of our functions $P(\cdot)$ from its definition and three axioms. We list these as the statements. Elementary Properties of Probability: By using the above axioms, the following useful properties of probability can be obtained:

1. $P(\bar{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. $P(A) \leq P(B)$ if $A \subset B$.
4. $P(A) \leq 1$.
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
6. $P(A \setminus B) = \underline{A}P(A \cap B)$.
7. If A_1, A_2, \dots, A_n are n arbitrary events in \mathcal{A} , then

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) \\ &\quad - \dots (-1)^{n-1} P(A_1 \cap A_2 \cup \dots \cup A_n) \end{aligned}$$

where the sum of the second term is over all distinct pairs of events, that of the third term is over all distinct triples of events, and so forth

8. If A_1, A_2, \dots, A_n is a finite sequence of mutually exclusive events in \mathcal{A} ($A_i \cap A_j = \emptyset$ for $i \neq j$), then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

and a similar equality holds for any subcollection of the events.

Note that property 4 can be easily derived from axiom 2 and property 3. Since $A \subset \Omega$, we have

$$P(A) \leq P(\Omega)$$

Thus, combining with axiom 1, we obtain

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

Property 5 implies that

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

since $P(A \cap B) \geq 0$ by axiom 1. Property 6 implies that

$$P(A \setminus B) = P(A) - P(B) \text{ if } B \subset A$$

since $A \cap B = B$ if $B \subset A$.

1.6 Equally Likely Events

1.6.1 Finite Sample Space

Consider a finite sample space Ω with n finite elements

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$$

where ω_i are elementary events. Let $P(\omega_i) = p_i$. Then

1. $0 \leq p_i \leq 1, \quad i = 1, 2, \dots, n.$
2. $\sum_{i=1}^n p_i = p_1 + \dots + p_n = 1.$
3. If $A = \bigcup_{i \in I} \omega_i$ where I is a collection of subscripts, then

$$P(A) = \sum_{\omega \in A} P(\omega_i) = \sum_{i \in I} p_i$$

1.6.2 Equally Likely Events

When all elementary events $\omega_i (i = 1, \dots, n)$ are equally likely, that is,

$$p_1 = p_2 = \dots = p_n$$

then we have

$$p_i = \frac{1}{n}, \quad i = 1, \dots, n.$$

and

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

where $n(A)$ is the number of the number of outcomes belonging to event A and n is the number of sample points in Ω .

1.7 Conditional Probability

Definition 1.5

The conditional probability of an event A given event B , denoted by $P(A \setminus B)$, is defined as


$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)} \quad P(B) > 0,$$

where $P(A \cap B)$ is the joint probability of A and B . Similarly,

$$P(B \setminus A) = \frac{P(A \cup B)}{P(A)} \quad P(A) > 0$$

is the conditional probability of an event B given event A . From these equations, we have

$$P(A \cap B) = P(A \setminus B)P(B) = P(B \setminus A)P(A) \quad (1.1)$$

Equation (1.1) is often quite useful in computing the joint probability of events. 

1.8 Bayes' Rule

From equation(1.1) we can obtain the following Bayes' Rule :

$$P(A \setminus B) = \frac{P(B \setminus A)P(A)}{P(B)} \quad (1.2)$$

1.9 Total Probability

The events A_1, A_2, \dots, A_n are all mutually exclusive and exhaustive if

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \forall i \neq j \quad (1.3)$$

Let B be any event in Ω . Then

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B \setminus A_i)P(A_i) \quad (1.4)$$

which is known as the total probability of event B . Let $A = A_i$ then by using Eq (1.4) we obtain

$$P(A_i \setminus B) = \frac{P(B \setminus A_i)P(A_i)}{\sum_{i=1}^n P(B \setminus A_i)P(A_i)} \quad (1.5)$$

Note that the terms on the right-hand side are all conditioned on events A_i , while the term on the left is conditioned on B . Equation (1.5) is sometimes referred to as Bayes' theorem.

1.10 Independent Events

Two events A and B are said to be (statistically) independent if and only if

$$P(A \cap B) = P(A)P(B) \quad (1.6)$$

It follows immediately that if A and B are independent, then

$$P(A \setminus B) = P(A) \quad \text{and} \quad P(B \setminus A) = P(B) \quad (1.7)$$

If two events A and B are independent, then it can be shown that A and \bar{B} are also independent; that is

$$P(A \cap \bar{B}) = P(A)P(\bar{B}) \quad (1.8)$$

Then

$$P(A \setminus \bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = P(A) \quad (1.9)$$

If A is independent of B , then the probability of A 's occurrence is unchanged by information as to whether or not B has occurred. Three events A, B, C are said to be independent if and only if

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

We may also extend the definition of independence to more than three events. The events A_1, A_2, \dots, A_n are independent if and only if for every subset $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ ($2 \leq k \leq n$) of events,

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}) \quad (1.10)$$

Finally, we define an infinite set of events to be independent if and only if every finite subset of these events is independent.

To distinguish between the mutual exclusiveness (or disjointness) and independence of a collection of events, we summarize as follows:

(a). $\{A_i, i = 1, \dots, n\}$ is a sequence of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (1.11)$$

(b). If $\{A_i, i = 1, 2, \dots, n\}$ is a sequence of independent events, then


$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i) \quad (1.12)$$

and a similar equality holds for any subcollection of the events.

1.11 Chapter Summary

This chapter has developed the basic concepts of probability, random experiments, and events. Several examples are solved and applications of probability have been provided in the fields of communications and reliability engineering. Finally, it introduced the concepts of permutation and combination that will be used in later chapters.

1.12 Exercises

 **Exercise 1.1** The sample space of an experiment is the real line express as

$$S = \{\nu; -\infty < \nu < \infty\}$$

(a). Consider the events

$$\begin{aligned} A_1 &= \left\{ \nu; 0 \leq \nu < \frac{1}{2} \right\} \\ A_2 &= \left\{ \nu; \frac{1}{2} \leq \nu < \frac{3}{4} \right\} \\ A_i &= \left\{ \nu; 1 - \frac{1}{2^{i-1}} \leq \nu < 1 - \frac{1}{2^i} \right\} \end{aligned}$$

Determine the events

$$\bigcup_{i=1}^n A_i \quad \text{and} \quad \bigcap_{i=1}^n A_i$$

(b). Consider the events

$$\begin{aligned} B_1 &= \left\{ \nu; \nu \leq \frac{1}{2} \right\} \\ B_2 &= \left\{ \nu; \nu \leq \frac{1}{4} \right\} \\ &\vdots \\ B_i &= \left\{ \nu; \nu \leq \frac{1}{2^i} \right\} \end{aligned}$$

Determine the events

$$\bigcup_{i=1}^{\infty} B_i \quad \text{and} \quad \bigcap_{i=1}^{\infty} B_i$$

Solution

(a). It is clear that


$$\bigcup_{i=1}^{\infty} A_i = \{\nu; 0 \leq \nu < 1\}$$

Noting that the A_i 's are mutually exclusive, we have

$$\bigcap_{i=1}^{\infty} A_i = \emptyset$$

(b). Noting that $B_1 \supset B_2 \cdots \supset B_i \subset \cdots$, we have

$$\bigcup_{i=1}^{\infty} B_i = B_1 = \left\{ \nu; \nu \leq \frac{1}{2} \right\} \quad \text{and} \quad \bigcap_{i=1}^{\infty} B_i = \{\nu; \nu \leq 0\}$$

 **Exercise 1.2** Let A and B be events in an event space F . Show that if A and B are independent, then so are (a) A and \bar{B} , (b) \bar{A} and B , and (c) \bar{A} and \bar{B} .

Solution

(a). We have

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

Since A and B are independent, we obtain

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] = P(A)P(\bar{B}) \end{aligned}$$

Thus, by definition, A and \bar{B} are independent.

(b). We get the result by interchanging A and B . Indeed, we get :


$$P(B \cap \bar{A}) = P(B)P(\bar{A})$$

which indicates that \bar{A} and B are independent.

(c). We have

$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P(\overline{A \cup B}) \\ &= 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \quad (\text{because } A \text{ and } B \text{ are independent}) \\ &= 1 - P(A)[1 - P(B)] - P(B) \\ &= 1 - P(A)P(\bar{B}) - P(B) \\ &= P(A)P(\bar{B}) + P(\bar{B}) \\ &= P(\bar{A})P(\bar{B}) \end{aligned}$$

Hence, \bar{A} and \bar{B} are independent.

 **Exercise 1.3** A student buys 1000 integrated circuits (ICs) from supplier A, 2000 ICs from supplier B, and 3000 ICs from supplier C. He tested the ICs and found that the conditional probability of an IC being defective depends on the supplier from whom it was bought. Specifically, given that an IC came from supplier A, the probability that it is defective is 0.05; given that an IC came from supplier B, the probability that it is defective is 0.10; and given that an IC came from supplier C, the probability that it is defective is 0.10. If the ICs from the three suppliers are mixed together and one is selected at random, what is the probability that it is defective?

Solution Let $P(A)$, $P(B)$, and $P(C)$ denote the probability that a randomly selected IC came from supplier A, B, and C, respectively. Also, let $P(D \setminus A)$ denote the conditional probability that an IC is defective, given that it came from supplier A; $P(D \setminus B)$ denote the conditional probability that an IC is defective, given that it came from supplier B; and $P(D \setminus C)$ denote the conditional probability that an IC is defective, given that it came from supplier C. Then the following are true:

$$\begin{aligned} P(D \setminus A) &= 0.05 \\ P(D \setminus B) &= 0.10 \\ P(D \setminus C) &= 0.10 \\ P(A) &= \frac{1000}{1000 + \frac{2000}{2000} + 3000} = \frac{1}{6} \\ P(B) &= \frac{2000}{1000 + \frac{2000}{2000} + 3000} = \frac{1}{3} \\ P(C) &= \frac{3000}{1000 + \frac{2000}{2000} + 3000} = \frac{1}{2} \end{aligned}$$

Let $P(D)$ denote the unconditional probability that a randomly selected IC is defective. The, from principal of total probability,

$$\begin{aligned} P(D) &= P(D \setminus A)P(A) + P(D \setminus B)P(B) + P(D \setminus C)P(C) \\ &= 0.05 \times \frac{1}{6} + 0.10 \times \frac{1}{3} + 0.10 \times \frac{1}{2} \\ &= 0.09167 \end{aligned}$$

✚ **Exercise 1.4** Consider the experiment of tossing a fair coin repeatedly and counting the number of tosses required until the first head appears.

- Find the sample space of the experiment.
- Find the probability that the first head appears on the k th toss.
- Verify that $P(S) = 1$.

Solution

(a). The sample space of this experiment is

$$S = \{a_1, a_2, a_3, \dots\} = \{a_k : k = 1, 2, 3, \dots\}$$

where a_k is the elementary event that the first head appears on the k th toss.

(b). Since a fair coin is tossed, we assume that a head and a tail are equally likely to appear.

Then $P(H) = P(T) = \frac{1}{2}$. Let

$$P(a_k) = p_k \quad k = 1, 2, \dots$$

Let

$$P(a_k) = p_k = \frac{1}{2^k}.$$

Since there are 2^k equally likely ways of tossing a fair coin k times, only one of which consists of $(k - 1)$ tails following a head we observe that

$$P(a_k) = p_k = \frac{1}{2^k} \quad k = 1, 2, 3, \dots$$

(c). Using the power series summation formula, we have

$$P(S) = P(S) = \sum_{k=1}^{\infty} P(a_k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

✚ **Exercise 1.5** Consider the experiment of the previous Exercise.

- Find the probability that the first head appears on an even-numbered toss.
- Find the probability that the first head appears on an odd-numbered toss.

Solution

(a). Let A be the event "the first head appears on an even-numbered toss." Then, we have

$$P(A) = p_2 + p_4 + p_6 \dots = \sum_{m=1}^{\infty} p_{2m} = \sum_{m=1}^{\infty} \frac{1}{2^{2m}} = \sum_{m=1}^{\infty} \left(\frac{1}{4}\right)^m = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}$$


(b). Let B be the event "the first head appears on an odd-numbered toss." Then it is obvious

that $B = \bar{A}$. So, we get

$$P(B) = P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}$$

As a check, notice that

$$P(B) = p_1 + p_3 + p_5 + \cdots = \sum_{m=0}^{\infty} p_{2m+1} = \sum_{m=0}^{\infty} \frac{1}{2^{2m+1}} = \frac{1}{2} \left[\frac{1}{1 - \frac{1}{4}} \right] = \frac{2}{3}$$

 **Exercise 1.6** A number is selected at random from $\{1, 2, \dots, 100\}$. Given that the number selected is divisible by 2, find the probability that it is divisible by 3 or 5.

Solution Let

$$A_2 = \{\text{event that the number is divisible by 2}\}$$

$$A_3 = \{\text{event that the number is divisible by 3}\}$$

$$A_5 = \{\text{event that the number is divisible by 5}\}$$

Then the desired probability is

$$\begin{aligned} P(A_3 \cup A_5) &= \frac{P[(A_3 \cup A_5) \cap A_2]}{P(A_2)} \\ &= \frac{P[(A_3 \cap A_2) \cup (A_5 \cap A_2)]}{P(A_2)} \\ &= \frac{P(A_3 \cap A_2) + P(A_5 \cap A_2) - P(A_3 \cap A_5 \cap A_2)}{P(A_2)} \end{aligned}$$

Now

$$A_3 \cap A_2 = \text{event that the number is divisible by 6}$$

$$A_5 \cap A_2 = \text{event that the number is divisible by 10}$$


$$A_3 \cap A_5 \cap A_2 = \text{event that the number is divisible by 30}$$

and

$$P(A_3 \cap A_2) = \frac{16}{100} \quad P(A_5 \cap A_2) = \frac{10}{100} \quad P(A_3 \cap A_5 \cap A_2) = \frac{3}{100}$$

thus,

$$P(A_3 \cup A_5 \mid A_2) = \frac{\frac{16}{100} + \frac{10}{100} - \frac{3}{100}}{\frac{50}{100}} = \frac{23}{50} = 0.46$$

 **Exercise 1.7** A company producing electric relays has three manufacturing plants producing 50, 30, and 20 percent, respectively, of its product. Suppose that the probabilities that a relay manufactured by these plants is defective are 0.02, 0.05, and 0.01, respectively.

- If a relay is selected at random from the output of the company, what is the probability that it is defective?
- If a relay selected at random is found to be defective, what is the probability that it was manufactured by plant 2?

Solution

(a). Let B be the event that the relay is defective, and let A_i be the event that the relay is

fig1.18

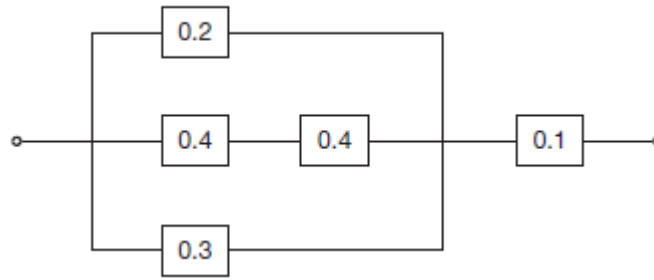


Figure 1.2

manufactured by plant i ($i = 1, 2, 3$). The desired probability is $P(B)$. we have

$$\begin{aligned} P(B) &= \sum_{i=1}^3 P(B \setminus A_i)P(A_i) \\ &= 0,02 \cdot 0,5 + 0,05 \cdot 0,3 + 0,01 \cdot 0,2 = 0,027 \end{aligned}$$

(b). The desired probability is $P(A_2 \setminus B)$ given by

$$P(A_2 \setminus B) = \frac{P(B \setminus A_2)P(A_2)}{P(B)} = \frac{0,05 \cdot 0,3}{0,027} = 0,556$$

Supplementary exercises

✚ **Exercise 1.8** The relay network shown in Fig. 1.2. operates if and only if there is a closed path of relays from left to right. Assume that relays fail independently and that the probability of failure of each relay is as shown. What is the probability that the relay network operates?

✚ **Exercise 1.9** Suppose that the sample space Ω consists of the positive integers from 1 to 10 inclusive. Let $A = \{2, 3, 4\}$, $B = \{3, 4, 5\}$, $C = \{5, 6, 7\}$. List the members of the following sets:

$$(a) \bar{A} \cap B \quad (b) \bar{A} \cup B \quad (c) \overline{\bar{A} \cap \bar{B}} \quad (d) \overline{A \cap (\bar{B} \cup \bar{C})} \quad (e) \overline{A \cap (B \cup C)}$$

✚ **Exercise 1.10** Which of the following relationships are true in general, i.e. for any three events A, B and C ?


$$\begin{aligned} (a) & (A \cup B) \cap (A \cup C) = A \cup (B \cap C) & (d) & \overline{A \cup \bar{B}} \cap C = \bar{A} \cap \bar{B} \cap \bar{C} \\ (b) & A \cup B = (A \cap \bar{B}) \cup B & (e) & (A \cap B) \cap (\bar{B} \cap C) = 0 \\ (c) & \bar{A} \cap B = A \cup B. \end{aligned}$$

✚ **Exercise 1.11** An installation consists of two boilers and one engine. Let A be the event that the engine is in good condition, B_i the event that the i th boiler ($i = 1, 2$) is in good condition, and C the event that the installation can operate. If the installation requires that at least one boiler and the engine are in good condition before it can operate, express C and \bar{C} in terms of A and B_1 and B_2 .


✚ **Exercise 1.12** Let A, B and C be three events associated with a random experiments. Express

the following verbal statements in set notations:

- (a) at least one of the event occurs;
- (b) exactly on of the events occurs;
- (c) exactly two events occur;
- (d) not more than two events occurs simultaneously.

 **Exercise1.13** Suppose that the space consists of all points (x,y) both of whose co-ordinates are integers and which lie inside or on the boundary of the square bounded by the line $x = 0, y = 0, x = 6$ and $y = 6$. List the members of the following sets:

$$\begin{aligned} (a) \quad A &= \{(x, y) \mid x^2 + y^2 \leq 6\}; & (d) \quad B \cap C \\ (b) \quad B &= \{(x, y) \mid y \leq x^2\} & (e) \quad (B \cup A) \cap \bar{C} \\ (c) \quad C &= \{(x, y) \mid x \leq y^2\}. \end{aligned}$$

 **Exercise1.14** Find the sample space for the experiment of tossing a coin repeatedly and of counting the number of tosses required until the first head appears.

Clearly all possible outcomes for this experiment are the terms of the sequence $1, 2, 3, \dots$. Thus:

$$S = \{1, 2, 3, \dots\}$$

Note that there are an infinite number of outcomes.

We define the events

$$\begin{aligned} A &= \{k : k \text{ is odd}\} \\ B &= \{k : 4 \leq k \leq 7\} \\ C &= \{k : 1 \leq k \leq 10\} \end{aligned}$$

where k is the number of tosses required until the first H (head) appears. Determine the events $\bar{A}, \bar{B}, \bar{C}, A \cup B, A \cup C, B \cup C$ and $\bar{A} \cap B$.

Chapter Random variables

Introduction

□ *Random variables*

□ *Expectation*

□ *Density*

□ *Variance*

In this chapter, we introduce the concept of random variable. This concept is a some outcome from a chance process. The main purpose is to calculate density and distribution functions, that make it both convenient and easy to compute the probabilities and expected values of various events. Also, we focus on the study of some transformation and combination of random variables.

2.1 Definitions

We start by defining a random variable.

Definition 2.1. Random Variable

For a given probability space (Ω, \mathcal{A}, P) , a random variable, denoted by X or $X(\cdot)$, is a function with domain Ω and counterdomain the real line. The function X must be such that the set A , defined by $A_t = \{\omega : X(\omega) \leq r\}$ belongs to \mathcal{A} for every real number r .



Remark

- The terminology used here is traditional. Clearly a random variable is not a variable at all in the usual sense, and it is a function.
- The use of words "random" and "variable" is inappropriate since their use cannot be reasonably justified. The expression "random variable" is a misnomer that has gained such widespread use that it would be foolish for us to try rename it.
- The sample space Ω is termed the domain of the random variable X , and the collection of all numbers (values of $X(\cdot)$) is termed the range of the random variable X . Thus the range of X is a certain subset of the set of all real numbers.
- Two or more different sample points might give the same value of $X(\cdot)$, but two different numbers in the range cannot be assigned to the same sample point.

2.1.1 Events Defined by Random Variables

If X is a random variable and x is a fixed real number, we can define the event ($X = x$) as

$$(X = x) = (\omega, X(\omega) = x)$$

Similarly, for fixed numbers x , x_1 and x_2 , we can define the following events:

$$\begin{aligned}(X \leq x) &= (\omega, X(\omega) \leq x) \\(X > x) &= (\omega, X(\omega) > x) \\(x_1 < X \leq x_2) &= (\omega, x_1 < X(\omega) \leq x_2)\end{aligned}$$

These events have probabilities that are denoted by

$$\begin{aligned}P(X = x) &= P(\omega, X(\omega) = x) \\P(X \leq x) &= P(\zeta, X(\omega) \leq x) \\P(X > x) &= P(\omega, X(\omega) > x) \\P(x_1 < X \leq x_2) &= P(\omega, x_1 < X(\omega) \leq x_2)\end{aligned}$$

Example2.1 Consider the experiments of tossing two dice. $\Omega = \{(i, j) : i = 1, \dots, 6, j = 1, \dots, 6\}$. We can define several random variables like


- the sum $X(\omega) = i + j$ if $\omega = (i, j)$.
- the difference between the upturned faces $Y(\omega) = |i - j|$ if $\omega = (i, j)$. It is clear that X can take on the values $2, 3, \dots, 12$ and Y can take its values in $\{0, 1, \dots, 5\}$.

2.2 Distribution Functions

Definition 2.2. Cumulative distribution function

The cumulative distribution function (cdf) of a random variable X , denoted by $F_X(\cdot)$, is the function defined by

$$F_X(x) = P(\omega, X(\omega) \leq x), \quad -\infty < x < +\infty \quad (2.1)$$

Most of the information about a random experiment described by the random variable X is determined by the behavior of $F_X(x)$. 

Example2.2 Consider again the experiment of tossing a single coin. Assume that the coin is fair. Let X denote the number of heads. Then

$$F_x(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \end{cases}$$

2.2.1 Properties of $F_X(x)$

Several properties of $F_X(x)$ follow directly from its definition (2.1).

Proposition 2.1

- (a). $0 \leq F_X(x) \leq 1$.
- (b). $F_X(x)$ is a nondecreasing function of x , that is, if $F_X(x_1) \leq F_X(x_2)$, if $x_1 \leq x_2$.
- (c). $\lim_{x \rightarrow +\infty} F_X(x) = F_X(+\infty) = 1$.
- (d). $\lim_{x \rightarrow -\infty} F_X(x) = F_X(-\infty) = 0$.
- (e). $F_X(x)$ is continuous from the right, that is,
 $\lim_{x \rightarrow a^+} F_X(x) = F_X(a^+) = F_X(a)$, $a^+ = \lim_{0 < \epsilon \rightarrow 0} (a + \epsilon)$.

Proof

(a). Evident.

(b). Note that the event

□

From definition (2.1), we can compute other probabilities, such as $P(a \leq X < b)$, $P(X > a)$ and $P(X < b)$:

$$\begin{aligned} P(a \leq X < b) &= F_X(b) - F_X(a) \\ P(X > a) &= 1 - F_X(a) \\ P(X < b) &= F_X(b^-), \quad b^- = \lim_{b - \epsilon} b - \epsilon \end{aligned}$$

Example 2.3 The cumulative distribution function of the random variable X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x + \frac{1}{2} & \text{if } 0 \leq x \leq \frac{1}{2} \\ 1 & \text{if } x > \frac{1}{2} \end{cases}$$

Compute $P(X > \frac{1}{4})$.

The probability that X is greater than $\frac{1}{4}$ is given by

$$\begin{aligned} P(X > \frac{1}{4}) &= 1 - P(X \leq \frac{1}{4}) \\ &= 1 - F_X(\frac{1}{4}) = 1 - (\frac{1}{4} + \frac{1}{2}) = \frac{1}{4}. \end{aligned}$$

Definition 2.3. Cumulative distribution function

Any function $F(\cdot)$ with domain the real line and codomain the interval $[0,1]$ satisfying the above three properties is defined to be cumulative distribution function.

Remark


- (a). This definition allows us to use the term "cumulative distribution function" without mentioning random variables.
- (b). The cumulative distribution function defined here are univariate; the introduction of bivariate and multivariate cumulative distribution function will be defined in chapter 4.

2.3 Density Functions

Random variables and the cumulative distribution function of a random variable have been defined. The cumulative distribution function described the distribution of values of random variables. For two distinct classes of random variables, the distribution of values can be described more simply by using **density functions**. These two classes, distinguished by the words *discrete* and *continuous*, are considered in the next two subsections.

2.3.1 Discrete Random Variables

Definition 2.4. Discrete random variable

A random variable X will be defined to be discrete if the range of X is countable. If a random variable X is discrete, then the corresponding cumulative distribution function $F_X(\cdot)$ will be defined to be discrete. 



Note By the range of X being countable we mean that there exists a finite or denumerable set of real numbers, say x_1, x_2, \dots such that X takes on values only in that set.


Definition 2.5. Discrete density function

Let X be a discrete random variable with distinct values $x_1, x_2, \dots, x_n, \dots$, then the function denoted by $f_X(\cdot)$ and defined by


$$f_X(x) = \begin{cases} P[X = x_j] & \text{if } x = x_j, j = 1, 2, \dots, n, \dots, \\ 0 & \text{if } x_i \neq x_j \end{cases}$$

is defined to be the discrete density function of X . 

Definition 2.6. Cumulative distribution function

The cumulative distribution function of a random variable X , denoted by $F_X(\cdot)$ is defined to be that function with domain the real line and counterdomain the interval $[0, 1]$ which satisfies $F_X(x) = P[X \leq x] = P[\{\omega : X(\omega) \leq x\}]$ for all every real number. 

Theorem 2.1

Let X be a discrete random variable. $F_X(\cdot)$ can be obtained from $f_X(\cdot)$, and vice versa. 

Proof Denote the mass points of X by x_1, x_2, \dots . Suppose $f_X(\cdot)$ is giving; then $F_X(x) = \sum_{\{j: x_j \leq x\}} f_X(x_j)$. Conversely, suppose $F_X(\cdot)$ is given; then $f_X(x_j) = F_X(x_j) - \lim_{h \rightarrow 0} F_X(x_j - h)$. hence $f_X(x_j)$ can be found for each mass point x_j ; however, $f_X(x) = 0$ for $x \neq x_j, j = 1, 2, \dots$, so $f_X(x)$ is determined for all real numbers. \square

Example 2.4 Consider the experiment of tossing two dice. Let X denote the total of the upturned faces. The mass points of X are $\{2, 3, \dots, 12\}$. $f_X(\cdot)$ is sketched in figure 1. Let Y denote the absolute difference of the upturned faces; then $f_Y(\cdot)$ is given by

y	0	1	2	3	4	5
$f_Y(y)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Proposition 2.2

Any function $f(\cdot)$ with domain the real line and counterdomain $[0, 1]$ is defined to be a discrete density function if for some countable set $x_1, x_2, \dots, x_n, \dots$

- (a). $f(x_j) > 0$ for $j = 1, 2, \dots$.
- (b). $f(x) = 0$ for $x \neq x_j$; for $j = 1, 2, \dots$.
- (c). $\sum f(x_j) = 1$, where the summation is over the points $x_1, x_2, \dots, x_n, \dots$.



This definition allows us to speak of discrete density functions without reference to some random variable. Hence we can talk about properties that a density function might have without referring to a random variable.

Definition 2.7. Probability mass functions

Suppose that the jumps in $F_X(x)$ of a discrete random variable X occur at the points x_1, x_2, \dots , where the sequence may be either finite or countably infinite, and we assume $x_i < x_j$ if $i < j$.

Then

$$F_X(x_i) - F_X(x_{i-1}) = P(X \leq x_i) - P(X \leq x_{i-1}) = P(X = x_i) \quad (2.2)$$

Let

$$p_X(x) = P(X = x) \quad (2.3)$$

The function $p_X(x)$ is called the probability mass function (pmf) of the discrete random variable X .



2.4 Continuous Random Variables

Definition 2.8

Let X be a random variable with cumulative distribution function $F_X(x)$. If $F_X(x)$ is continuous and also has a derivative $\frac{dF_X(x)}{dx}$ which exists everywhere except at possibly a finite number of points and is piecewise continuous, then X is called a continuous random variable. Alternatively, X is a continuous random variable only if its range contains an interval (either finite or infinite) of real numbers. Thus, if X is a continuous random variable, then

$$P(X = x) = 0 \quad (2.4)$$

Note that this is an example of an event with probability 0 that is not necessarily the

impossible event \emptyset .

In most applications, the random variable is either discrete or continuous. But if the cumulative distribution function $F_X(x)$ of a random variable X possesses features of both discrete and continuous random variable's, then the random variable X is called the mixed random variable



2.4.1 Probability Density Function

Let

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.5)$$

The function $f_X(x)$ is called the probability density function (pdf) of the continuous random variable X .

2.4.2 Properties of $f_X(x)$

(a). $f_X(x) \geq 0$.

(b). Since X must assume some value, $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

(c). $f_X(x)$ is piecewise continuous.

(d). $P(a < X \leq b) = \int_a^b f_X(x)dx$, which means that $P(X = a) = \int_a^a f_X(x)dx = 0$.

Thus, the probability that a continuous random variable will assume any fixed value is zero.

The cumulative distribution function $F_X(x)$ of a continuous random variable X can be obtained by

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt \quad (2.6)$$

By Eq. (2.4), if X is a continuous random variable, then

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) = P(a < X < b) \\ &= P(a < X < b) \\ &= \int_a^b f_X(x)dx = F_X(b) - F_X(a) \end{aligned}$$

Example 2.5 Is the following function a legitimate probability density function?

$$f(x) = \begin{cases} \frac{x^2}{9}, & 0 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

For $f(x)$ to be a legitimate probability density function, we need to check to see if $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Thus

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^3 dx = \frac{x^2}{9} = \left[\frac{x^3}{27} \right]_0^3 = 1$$

Therefore, $f(x)$ is a legitimate probability density function.

2.5 Mean And Variance

Two numbers are often used to summarize a probability distribution X . The mean is a measure of the center or middle of the probability distribution, and variance is a measure of the dispersion, or variability in the distribution. The two measures do not uniquely identify a probability distribution. That is, two different distribution can have the same mean and variance. Still, these measures are simple, useful summarize of the probability distribution of X .

2.5.1 Mean

The mean (or expected value) of a rev. X , denoted by μ_X or $E(X)$, is defined by

$$\mu_X = E(X) = \begin{cases} \sum x_k p_X(x_k), & X \text{ discrete} \\ \int_{-\infty}^{+\infty} x_X(x)dx, & X \text{ continuous} \end{cases} \quad (2.7)$$

2.5.2 Moment

The n^{th} moment of a random variable X is defined by

$$E(X^n) = \begin{cases} \sum x_k^n p_X(x_k), & X \text{ discrete} \\ \int_{-\infty}^{+\infty} x^n f_X(x)dx, & X \text{ continuous} \end{cases} \quad (2.8)$$

Note that the mean of X is the first moment of X .

2.5.3 Variance

The variance of a random variable X , denoted by σ_x^2 or $Var(X)$, is defined by

$$\sigma_x^2 = Var(X) = E(X - E(X))^2 \quad (2.9)$$

Thus,

$$\sigma_x^2 = \begin{cases} \sum (x_k - \mu_x)^2 p_X(x_k), & X \text{ discrete} \\ \int_{-\infty}^{+\infty} (x - \mu_x)^2 f_X(x)dx, & X \text{ continuous} \end{cases} \quad (2.10)$$

Note from (2.9) that

$$Var(X) \geq 0 \quad (2.11)$$

The standard deviation of a random variable X , denoted by σ_x , is the positive square root of $Var(X)$.

Expanding the right-hand side of Eq. (2.9), we can obtain the following relation:

$$Var(X) = E(X^2) - [E(X)]^2 \quad (2.12)$$

which is a useful formula for determining the variance.

Remark

- The mean of a discrete random variable X is a weighted of the possible values of X , with weights equal to the probabilities. If $f(x)$ is the probability mass function of a loading on a long, thin beam. $E(X)$ is the point at which the beam balances. Consequently, $E(X)$ describes the "center of the distribution of X in a manner similar to the balance point of a loading.
- The variance of a random variable X is a measure of dispersion or scatter in the possible values for X uses weight $f(x)$ as the multiplier of each possible squared deviation $(x - \mu)^2$.
- The variance of a random variable X can be considered to be the expected value of a specific function of X , namely, $h(X) = (X - \mu)^2$. In general, the expected value of any function $h(X)$ of a discrete random variable is defined in a similar manner.

Now, we give the expectation of an arbitrary random variables X , called the mean of X .

Theorem 2.2. Expected value of a Function

Let $h(\cdot)$ be a function with both domain and couterdomain the real line. If X is a discrete random variable with probability mass function $f(x)$. The expectation or expected value of the function $h(\cdot)$ of the random variable X , denoted $E[h(X)]$, is defined by :

-

$$E[h(X)] = \sum_x h(x)f(x)$$

if X is discrete with mass points $x_1, x_2, \dots, x_j, \dots$ (provided this series is absolutely convergent).

-

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

if X is a continuous with probability density function $f_X(x)$ (provided $\int_{-\infty}^{\infty} |h(x)|f_X(x)dx < \infty$.)



Remark

- Expectation or expected value is not really a very good name since it is not necessarily what you "expect". For example, the expected value of a discrete random variable is not necessarily one of the possible values of the discrete random variable, in which case, you would not "expect" to get the expected value. A better name might be "average value rather

than "expected value".

- Since $E(h(X))$ is defined in term of the density function of X , it could be defined without reference to a random variable.
- If $h(x) = x$, then $E(h(X)) = E(X)$ is the mean of X . If $h(x) = (x - \mu_X)^2$, then $E(h(X)) = \text{Var}(X)$

Below are properties of expected value:

- i)** $\mathbf{E}(c) = c$ for a constant c .
- ii)** $\mathbf{E}(ch(X)) = c\mathbf{E}(h(X))$ for a constant c .
- iii)** $\mathbf{E}[c_1h_1(X) + c_2h_2(X)] = c_1\mathbf{E}(h_1(X)) + c_2\mathbf{E}[h_2(X)]$.

If $h_1(x) \leq h_2(x)$, for all x then $\mathbf{E}(h_1(X)) \leq \mathbf{E}(h_2(X))$

Proof Assume X is continuous. To prove (i), take $h(x) = c$, then

$$\begin{aligned}\mathbf{E}(h(X)) &= \mathbf{E}(c) = \int_{-\infty}^{+\infty} cf_X(x)dx = c \\ \mathbf{E}[ch(X)] &= \int_{-\infty}^{\infty} ch(x)f_X(x)dx = c \int_{-\infty}^{\infty} h(x)f_X(x)dx = c\mathbf{E}[g(X)]\end{aligned}$$

which proves (ii). (iii) is given by

$$\begin{aligned}\mathbf{E}[c_1h_1(X) + c_2h_2(X)] &= \int_{-\infty}^{\infty} [c_1g_1(x) + c_2g_2(x)] f_X(x)dx \\ &= c_1 \int_{-\infty}^{\infty} h_1(x)f_X(x)dx + c_2 \int_{-\infty}^{\infty} h_2(x)f_X(x)dx \\ &= c_1\mathbf{E}[h_1(X)] + c_2\mathbf{E}[h_2(X)]\end{aligned}$$

Finally,

$$0 \leq \mathbf{E}[h_2(X) - h_1(X)] = \mathbf{E}[h_2(X)] - \mathbf{E}[h_1(X)],$$

which gives (iv).



Note Similar proofs could be presented for the discrete random variable case.

□

Theorem 2.3

If X is a random variable, $\text{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$ provided $\mathbf{E}[X^2]$ exists.



Proof (We first note that if $\mathbf{E}[X^2]$ exists then $\mathbf{E}[X]$ exists.) By our definition of variance and $\mathbf{E}[h(X)]$, it follows that $\text{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$. Now $\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X - 2\mathbf{E}[X] + \mathbf{E}[X^2]] = \mathbf{E}[X^2] - 2(\mathbf{E}[X])^2 + (\mathbf{E}[X])^2 = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. □

2.5.4 Chebyshev Inequality

Theorem 2.4

Let X be a random variable and $h(\cdot)$ a nonnegative function with domain the real line; then

$$P[h(X) \geq k] \leq \frac{\mathbf{E}[h(X)]}{k} \quad \text{for every } k > 0$$



Proof Assume that X is a continuous random variable with probability density function the real line; then

$$\begin{aligned} \mathbf{E}[h(X)] &= \int_{-\infty}^{+\infty} h(x)f_X(x)dx = \int_{x:h(x)g(x) \geq x} h(x)f_X(x)dx \\ &+ \int_{-\infty}^{+\infty} h(x)f_X(x)dx \\ &\geq \int_{\{x: xh(x) \geq k\}} h(x)f_X(x)dx \\ &\geq \int_{x:h(x) \geq k} kf_X(x)dx = kP[(h(X) \geq k)] \end{aligned}$$

divided by k , and the result follows. A similar proof holds for X discrete. \square

Proposition 2.3

If X is a random variable with finite variance,

$$P[|X - \mu_X| \geq r\sigma_X] = P[(X - \mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2} \quad \text{for every } r > 0. \quad (2.13)$$



Proof Then $h(x) = (x - \mu_X)^2$ and $k = r^2\sigma_X^2$. \square

Remark If X is a random variable with finite variance,

$$P[|X - \mu_X| < r\sigma_X] \geq 1 - \frac{1}{r^2},$$

which is just a rewriting of (2.13).

2.5.5 Jensen Inequality

Definition 2.9. Convex function

A continuous function $h(\cdot)$ with domain and counterdomain the real line is called "convex" if for every x_0 on the real line, there exists a line which goes through the point $(x_0, h(x_0))$ and lies on or under the graph of the function $h(\cdot)$.



Theorem 2.5. Jensen Inequality

Let X be a random variable with mean $\mathbf{E}[X]$, and let $h(\cdot)$ be a convex function; then $\mathbf{E}[h(X)] \geq h(\mathbf{E}[X])$.



Proof Since $h(x)$ is continuous and convex, there exists a line, say $I(x) = a + bx$, satisfying $I(x) = a + bx \leq h(x)$ and $I(\mathbf{E}[X]) = h(\mathbf{E}[X])$. $I(x)$ is a line given by the definition of continuous

and convex that goes through the point $(\mathbf{E}[X], h(\mathbf{E}(X)))$. Note that $\mathbf{E}[I(x)] = \mathbf{E}[(a + bX)] = a + \mathbf{E}[X] = I(\mathbf{E}[X])$; hence, $h(\mathbf{E}[X]) = I(\mathbf{E}[X]) \leq \mathbf{E}[h(X)]$. \square

2.6 Moments and generating function

The moment (or raw moments) of a random variable or of a distribution are the expectations of the powers of the random variables which has the given distribution.

Definition 2.10. Moments

If X is a random variable, the r^{th} moment of X , is defined as

$$\mu_r = \mathbf{E}[X^r] \quad \text{if the expectation exists.}$$

Note that $\mu_1 = \mathbf{E}[X] = \mu_X$, the mean of X .



Definition 2.11. Central moment

If X is random variable, the r^{th} central moment of X about a is defined as $\mathbf{E}[(X - \mathbf{E}[X])^r]$.

If $a = \mu_X$, we have the r^{th} central moment of X about μ_X , denoted by μ_r , which is

$$\mu_r = \mathbf{E}[(X - \mu_X)^r].$$



Definition 2.12. Quantile

The q^{th} quantile of a random variable X or of its corresponding distribution is denoted by ζ_q and is defined as the smallest number ζ satisfying $F_X(\zeta) \geq q$.



If X is a continuous random variable, the the q^{th} quantile of X is given as the smallest number ζ satisfying $F_X(\zeta) = q$.

Definition 2.13. Median

The median of a random variable X , denoted by $\text{Med}(X)$, satisfying $P[X \leq \text{Med}(X)] \geq \frac{1}{2}$ and $P[X \leq \text{Med}(X)] \leq \frac{1}{2}$



We now define another kind of moment, factorial moment.

Definition 2.14. Factorial moment

If X is a random variable, the r^{th} factorial moment of X is defined as (r is a positive integer):

$$\mathbf{E}[X(X - 1)(X - 2) \cdots (X - r + 1)]$$

For some random variables (usually discrete), factorial moments are easier to calculate than raw moments. However the raw moments can be obtained from the factorial moments and vice versa.



Definition 2.15. Moment generating function

Let X be a random variable with density $f_X(\cdot)$. The expected value of e^{tX} is defined to be the moment generating function of X if the expected value exists for every value of t in some interval $-h < t < h; h > 0$. The moment generating function, denoted by $m_X(t)$ or $m(t)$, is

$$m(t) = \mathbf{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

if the random variable is continuous, and is

$$m(t) = \mathbf{E}[e^{tX}] = \sum_x e^{tx} f_X(x)$$

if the random variable is discrete. 

If a moment generating function exists, then $m(t)$ is continuously differentiable in some neighborhood of the origin. If we differentiate the moment generating function t times with respect to t , we have

$$\frac{d^r}{dt^r} m(t) = \int_{-\infty}^{\infty} x^r e^{xt} f_X(x) dx,$$

and letting $t \rightarrow 0$, we find

$$\frac{d^r}{dt^r} m(0) = \mathbf{E}[X^r] = \mu_r,$$


where the symbol on the left is to be interpreted to mean the r th derivative of $m(t)$ evaluated as $t \rightarrow 0$. Thus the moments of a distribution may be obtained from the moment generating function by differentiation, hence its name.

Example 2.6 Let X be a random variable with probability density function given by $f_X(x) = \lambda e^{-\lambda x} \mathbf{I}_{[0, \infty[}(x)$.

$$\begin{aligned} m_X(t) &= \mathbf{E}[e^{tX}] = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda. \\ m'(t) &= \frac{dm(t)}{dt} = \frac{\lambda}{(\lambda - t)^2} \quad \text{hence } m'(0) = \mathbf{E}[X] = \frac{1}{\lambda}. \\ m''(t) &= \frac{2\lambda}{(\lambda - t)^3}, \quad \text{so } m''(0) = \mathbf{E}[X^2] = \frac{2}{\lambda^2}. \end{aligned}$$

As with moments, there is also a generating function for factorial moments.

Definition 2.16. Factorial moments generating function

Let X be a random variable. The factorial moment generating function is defined as $\mathbf{E}[t^X]$ if this expectation exists. 

The factorial moment generating function is used to generate factorial moment moments in the same way as the raw moments are obtained from $\mathbf{E}[e^{tX}]$ except that t approaches 1 instead of 0. It sometimes simplifies finding moments of discrete distributions.

Example 2.7 Suppose X is a discrete density function given by

$$f_X(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

Then

$$\begin{aligned} \mathbf{E}[t^X] &= \sum_{x=0}^{\infty} \frac{t^x e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)} \\ \frac{d}{dt} \mathbf{E}[t^X] &= \frac{d}{dt} e^{\lambda(t-1)} = \lambda e^{\lambda(t-1)} \quad \text{hence } \left. \frac{d}{dt} \mathbf{E}[t^X] \right|_{t=1} = \lambda. \end{aligned}$$

A moment generating function is used, as its name suggests, to generate moments. That, however, will not be its only use for us. An important use will be in determining distributions.

Theorem 2.6

Let X and Y be two random variables with density $f_X(\cdot)$ and $f_Y(\cdot)$, respectively. Suppose that $m_X(t)$ and $m_Y(t)$ both exist and are equal for all t in the interval $-h < t < h$ for some $h > 0$. Then the two cumulative distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$ are equal. ♡

2.6.1 Characteristic Functions

There are random variables for which the moment generating function does not exist on any real interval with positive length. For example, consider the random variable X that has a Cauchy distribution

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad \forall x$$

You can show that for any nonzero real number t

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi} \frac{1}{1+x^2} dx = \infty$$

Therefore, the moment generating function does not exist for this random variable on any real interval with positive length. If a random variable does not have a well-defined MGF, we can use the characteristic function defined as

$$\phi_X(t) = \mathbf{E} [e^{itX}]$$

where $j = -1$ and ω is a real number. It is worth noting that this is a complex-valued random variable. We have not discussed complex-valued random variables. Nevertheless, you can imagine that a complex random variable can be written as $X = Y + jZ$, where Y and Z are ordinary real-valued random variables. Thus, working with a complex random variable is like working with two real-valued random variables. The advantage of the characteristic function is that it is defined for all real-valued random variables. Specifically, if X is a real-valued random variable, we can write

$$|e^{j\omega X}| = 1$$

Therefore, we conclude

$$\begin{aligned} |\phi_X(t)| &= |\mathbf{E}[e^{j\omega X}]| \\ &\leq \mathbf{E}[|e^{j\omega X}|] \\ &\leq 1 \end{aligned}$$

The characteristic function has similar properties to the MGF. For example, if X and Y are independent

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$$

More generally, if X_1, \dots, X_n are n independent random variables, then

$$\phi_{X_1+\dots+X_n}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) \cdots \phi_{X_n}(t).$$

Example 2.8 Let X be a random variable with probability density function

$$f_X(x) = \lambda e^{-\lambda x} \quad x > 0$$

2.7 Some Special Discrete Distributions


In this section we present and we list some important special discrete distributions.

2.7.1 Discrete Uniform Distribution

Definition 2.17. Discrete uniform distribution

Each member of the family of discrete density function

$$f(x) = f(x, N) = \begin{cases} \frac{1}{N} & \text{for } x = 1, 2, \dots, N \\ 0 & \text{otherwise} \end{cases}$$

where the parameter N ranges over the positive integers, is defined to have a discrete uniform distribution. 

If X has a discrete uniform distribution, then $\mathbf{E}[X] = \frac{N+1}{2}$, $\text{Var}(X) = \frac{(N-1)}{12}$ and $m_X(t) = \sum_{j=1}^N e^{jt} \frac{1}{N}$.

2.7.2 Bernoulli and Binomial Distributions

Definition 2.18. Bernoulli Distribution

A random variable X is defined to have a Bernoulli random variable with parameter p if its probability mass function is given by

$$p_X(k) = P(X = k) = p^k(1 - p)^{1-k}, \quad k = 0, 1 \quad (2.14)$$

where $0 \leq p \leq 1$. By Eq. (??), the cumulative distribution function $F_X(x)$ of the Bernoulli random variable X is given by

$$F_X(x) = \begin{cases} 0, & x < 0; \\ 1 - p, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases} \quad (2.15)$$



Theorem 2.7

The mean, variance and generating function of the Bernoulli random variable X are

$$\begin{aligned} \mu_X &= E(X) = p \\ \sigma_X^2 &= \text{Var}(X) = p(1 - p) \\ m_X(t) &= pe^t + q \end{aligned}$$



A Bernoulli random variable X is associated with some experiment where an outcome can be classified as either a "success" or a "failure," and the probability of a success is p and the probability of a failure is $1 - p$. Such experiments are often called Bernoulli trials

Definition 2.19. Binomial Distribution

A random variable X is called a binomial random variable with parameters (n, p) if its probability mass function is given by

$$p_X(k) = P(X = k) = C_n^p p^k(1 - p)^{n-k}, \quad k = 0, 1, \dots, n \quad (2.16)$$

where $0 \leq p \leq 1$ and

$$C_n^p = \frac{n!}{p!(n - p)!}$$

which is known as the binomial coefficient. The corresponding cumulative distribution function of X

$$F_X(x) = \sum_{k=0}^n C_n^p p^k(1 - p)^{n-k}, \quad n \leq x < n + 1 \quad (2.17)$$



Theorem 2.8

The mean, variance and the generating function of the binomial random variable X are

$$\begin{aligned}\mu_X &= E(X) = np \\ \sigma_X^2 &= Var(X) = np(1-p) \\ m_X(t) &= (q + pe^t)^n.\end{aligned}$$



A binomial random variable X is associated with some experiments in which n independent Bernoulli trials are performed and X represents the number of successes that occur in the n trials. Note that a Bernoulli random variable is just a binomial random variable with parameters $(1, p)$.

2.7.3 Geometric and negative Binomial distribution

An other family of discrete distributions that play important roles in statistics is geometric (or Pascal) and negative binomial distribution.

Definition 2.20. Geometric distribution

A random variable X is defined to have geometrix (or Pascal) distribution if its density of X is given by

$$f_X(x) = f_X(x, p) = \begin{cases} p(1-p) & \text{for } x = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases} = p(1-p)^x \mathbb{I}_{\{0, 1, \dots, \}(x)},$$

where the parameters p satisfies $0 < p \leq 1$. (Define $q = (1-p)$).

**Definition 2.21. Negative Binomial Distribution**

A random variable X with density

$$f_X(x) = \begin{cases} \binom{r+x-1}{x} p^r q^x & \text{for } x = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

where the parameters r and p satisfy $r = 1, 2, 3, \dots$ and $0 < p < 1$ are defined to have a negative binomial distribution.



Remark If in the negative binomial distribution $r = 1$, then the negative binomial density specializes to the geometric density.

Theorem 2.9

If the random variable X has a geometric distribution, then

$$E[X] = \frac{1-p}{p}, \quad V[X] = \frac{1-p}{p^2}, \quad m_X(t) = \frac{p}{1-(1-p)e^t}$$



The geometric distribution is well named since the values that the geometric density assumes are the terms of a geometric series. Also the mode of the geometric density is necessarily 0.

A geometric density possesses one other interesting property, which is given in the following theorem.

Theorem 2.10

If X has the geometric density with parameter p , then

$$P[X \geq i + j \mid X \geq i] = P[X \geq j] \quad \text{for } j = 0, 1, 2, \dots$$



Proof

$$\begin{aligned} P[X \geq i + j \mid X \geq i] &= \frac{P[X \geq i + j]}{P[X \geq i]} \\ &= \frac{\sum_{x=i+j}^{\infty} p(1-p)^x}{\sum_{x=i}^{\infty} p(1-p)^x} = \frac{(1-p)^{i+j}}{(1-p)^i} \\ &= (1-p)^j \\ &= P[X \geq j]. \end{aligned}$$

□

Theorem 2.11

Let X have a negative binomial distribution; then

$$\mathbf{E}[X] = \frac{rp}{p} \quad \text{Var}[X] = \frac{rq}{p^2} \quad \text{and } m_X(t) = \left[\frac{p}{1 - qe^t} \right]$$



2.7.4 Hypergeometric Distribution

Definition 2.22. Hypergeometric distribution

A random variable X is defined to have a hypergeometric distribution if the discrete density function of X is given by

$$f_X(x, M, K, n) = \begin{cases} \binom{K}{x} \binom{M-K}{n-x} & \text{for } x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

where M is a positive integer, K is nonnegative integer that is at most M , and n is a positive integer that is at most M . Any distribution function defined by the density function given above is called a hypergeometric distribution.



If X is a hypergeometric distribution, then

$$\mathbf{E}[X] = n \frac{K}{M} \quad \text{var}[X] = n \frac{K}{M} \frac{M-K}{M} \frac{M-n}{M-1}$$

2.7.5 Poisson Distribution

A random variable X is called a Poisson random variable with parameter λ ($\lambda > 0$) if its probability mass function is given by

$$p_X(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots \quad (2.18)$$

The corresponding cumulative distribution function of X is

$$F_X(x) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}, \quad n \leq x < n + 1 \quad (2.19)$$

Theorem 2.12

Let X be a Poisson distribution random variable; then

$$\mu_X = E(X) = \lambda$$

$$\sigma_X^2 = \text{Var}(X) = \lambda$$

and

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



The Poisson random variable has a tremendous range of applications in diverse areas because it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small enough so that np is of a moderate size.

2.8 Continuous Distributions

In this section several parametric families of univariate probability density function are presented. Sketches of some are included; the mean and the variance (when they exist) of each are given. We can give a few examples.

- The number of telephone calls arriving at a switching center during various intervals of time.
- The number of misprints on a page of a book.
- The number of customers entering a bank during various intervals of time.

2.8.1 Uniform or Rectangular Distribution

A random variable X is called a uniform random variable over (a, b) if its pdf is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise.} \end{cases} \quad (2.20)$$

The corresponding cumulative distribution function of X is

$$F_X(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases} \quad (2.21)$$

Theorem 2.13

The mean and variance of the uniform random variable X are

$$\begin{aligned} \mu_X &= E(X) = \frac{a+b}{2} \\ \sigma_X^2 &= Var(X) = \frac{(b-a)^2}{12} \\ m_X(t) &= \frac{e^{bt} - e^{at}}{(b-a)t} \end{aligned}$$



A uniform random variable X is often used where we have no prior knowledge of the actual pdf and all continuous values in some range seem equally likely.

2.8.2 Exponential Distribution

A random variable X is called an exponential random variable with parameter λ ($\lambda > 0$) if its pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x < 0 \end{cases} \quad (2.22)$$

The corresponding cumulative distribution function of X is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.23)$$

Theorem 2.14

The mean and variance of the exponential random variable X are

$$\begin{aligned} \mu_X &= E(X) = \frac{1}{\lambda} \\ \sigma_X^2 &= Var(X) = \frac{1}{\lambda^2} \\ m_X(t) &= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda. \end{aligned}$$



The most interesting property of the exponential distribution is its "memoryless" property. By this we mean that if the lifetime of an item is exponentially distributed, then an item which has been in use for some hours is as good as a new item with regard to the amount of time remaining until the item fails. The exponential distribution is the only distribution which possesses this property.

2.8.3 Gamma Distribution

A continuous random variable whose density is given by

$$F_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.24)$$

for some $\lambda > 0, \alpha > 0$ is said to be a gamma random variable with parameters α, λ . The quantity $\Gamma(\alpha)$ is called the gamma function and is defined by

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

It is easy to show by induction that for integral α , say, $\alpha = n$, $\Gamma(n) = (n-1)!$.

Remark If in the game density $r = 1$, the gamma density specializes to the exponential density.

Theorem 2.15

If X has a gamma distribution with parameters r and λ , then

$$\mathbf{E}[X] = \frac{r}{\lambda} \quad \text{var}[X] = \frac{r}{\lambda^2} \quad m_X(t) = \left(\frac{\lambda}{\lambda - t} \right)^r \quad \text{for } t < \lambda.$$



2.8.4 Normal (or Gaussian) Distribution

A random variable X is called a normal (or gaussian) random variable if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.25)$$

where the parameters μ and σ satisfy $-\infty < \mu < \infty$ and $\sigma > 0$.

We have used the symbols μ and σ^2 to represent the parameters because these parameters turn out, as we shall see, to be the mean and the variance, respectively, of the distribution.

If the random variable has mean 0 and variance 1, it is called a *standard or normalized* random variable. For a standard normal random variable the subscripts of the density and distribution function notations are dropped; that is,

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

Since $f_X(x)$ is given to be a density function, it implied that

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

The corresponding cumulative distribution function of X is

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt \quad (2.26)$$

This integral cannot be evaluated in a closed form and must be evaluated numerically. It is convenient to use the function $\phi(z)$, defined as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt \quad (2.27)$$

to help us to evaluate the value of $F_X(x)$. Then Eq. (2.26) can be written as

$$F_X(x) = \phi\left(\frac{x - \mu}{\sigma}\right) \quad (2.28)$$

Note that

$$\phi(-z) = 1 - \phi(z) \quad (2.29)$$

Theorem 2.16

If X is a normal random variable,

$$\mu_X = E(X) = \mu$$

$$\sigma_X^2 = \text{Var}(X) = \sigma^2 m_X(t) = e^{+\sigma^2 t^2/2}$$



The following theorem shows that we can find the probability that a normally distributed random variable, with mean μ and variance σ^2 , falls in interval in terms of the standard normal cumulative distribution function, and this standard normal cumulative distribution function is Table 1 of Appendix D.

Theorem 2.17

If $X \sim \mathcal{N}(m, \sigma^2)$, then

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



Remark

$$\Phi(x) = 1 - \Phi(-x).$$

2.8.5 Cauchy distribution

Definition 2.23

A distribution which we shall find useful for illustrative purposes is the Cauchy, which has the density

$$f_X(x, \alpha, \beta) = \frac{1}{\pi\beta(1 + [(x - \alpha)/\beta]^2)}$$

where $-\infty < \alpha < \infty$ and $\beta > 0$.



Although the Cauchy density is **symmetrical** about the parameters α , its mean and higher moments do not exist. **The cumulative distribution function** is

$$\begin{aligned} F_X(x) &= \frac{1}{\pi} \int_{-\infty}^x \frac{dt}{\pi\beta(1 + [(t - \alpha)/\beta]^2)} \\ &= \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x - \alpha}{\beta}. \end{aligned}$$

2.8.6 Lognormal distribution

Let X be a positive random variable, and let a new random variable Y be defined as $Y = \log_e X$. If Y has a normal distribution, then, X is said to have a lognormal distribution. Lognormal Distribution If a random variable X has a density

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}(\log_e x - \mu)^2\right] I_{(0,\infty)}(x)$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.

2.9 Comments

We conclude this chapter by making several comments.

Approximations

Although many approximations of one distribution by another exist, we will give only three here. Others will be given with the central limit theorem.

Proposition 2.4. Binomial by Poisson

Let X be a random variable with Binomial distribution with parameters n and p , as

$$\binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

If the parameter n approaches infinity and p approaches 0 in such a way that np remains constant, say equal to λ , then

$$\binom{n}{x} p^x (1-p)^{n-x} \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}$$

for fixed integer x .

Theorem 2.18

Let random variable X have a Poisson distribution with parameters λ ; then for fixed $a < b$

$$P\left[a < \frac{X - \lambda}{\sqrt{\lambda}} < b\right] = P[\lambda + a\sqrt{\lambda} < X < \lambda + b\sqrt{\lambda}] \rightarrow \Phi(b) - \Phi(a) \quad \text{as } \lambda \rightarrow \infty$$


Theorem 2.19. De Moivre-Laplace theorem

Let a random variable X have a binomial distribution with parameters n and p ; then for fixed $a < b$

$$P\left[a \leq \frac{X - np}{\sqrt{npq}} \leq b\right] = P[np + a\sqrt{npq} \leq X \leq np + b\sqrt{npq}] \rightarrow \Phi(b) - \Phi(a) \quad \text{as } n \rightarrow \infty.$$

Note this theorem is a special case of the central limit theorem.


2.10 Exercises

 **Exercise 2.1** Let X be a random variable with a uniform distribution on $\{0, 1, 2, \dots, a\}$, where $a \in \mathbb{N}$. Assume that $\mathbf{E}[X] = 6$. Determine the value of a .

Solution We have

$$\mathbf{E}[X] = \frac{1}{a+1}(0+1+2+\dots+a) = \frac{a}{2}.$$

We can therefore deduce that $a = 12$.

 **Exercise 2.2** Let X and Y be two independent random variables following the uniform distribution on $\{0, 1, \dots, n\}$.

- Determine $P[X = Y]$.
- Determine $P[X \geq Y]$.
- Determine the law of $X + Y$.

Solution

(a). Note that

$$X = Y \Leftrightarrow \exists k \in \{1, 2, \dots, n\}, X = k \text{ and } Y = k.$$

So, we have

$$P[X = k] = \sum_{k=1}^n P[X = k \cap Y = k] = \sum_{k=1}^n P[X = k]P[Y = k].$$

Since the random variables are independent. We deduce

$$P[X = Y] = \sum_{k=1}^n \frac{1}{n} \times \frac{1}{n} = \frac{1}{n}.$$

(b). A similar method can be used to study all cases where $X \geq Y$. It's even simpler to note that, by symmetry of the roles played by X and Y , we have

$$P[X \geq Y] = P[Y \geq X].$$

But,

$$\begin{aligned} P[(X \geq Y) \cup (Y \geq X)] &= P[X \geq Y] + P[Y \geq X] - P[(X \cap Y) \cap (Y \cap X)] \\ &= 2P[X \geq Y] - P[X = Y]. \end{aligned}$$

Using the result from the previous question, we find

$$1 = 2P[X \geq Y] - \frac{1}{n} \Rightarrow P[X \geq Y] = \frac{1}{2} + \frac{1}{2n}.$$

(c). Note that

$$X + Y = k \Leftrightarrow \exists i \in \{1, \dots, k-1\}, X = i \text{ and } Y = k - i$$

It follows, for all $k = 2, \dots, 2n$ (possible values of $X+Y$), that

$$P[X + Y = k] = \sum_{i=1}^{k-1} P[(X = i) \cap (Y = k - i)] = \sum_{i=1}^{k-1} P[X = i]P[Y = k - i]$$

by independence of X and Y . In reality, we can sum over a smaller set of indices. Indeed,


$P[X = i] \neq 0$ implies that $1 \leq i \leq n$ and $P[Y = k - i] \neq 0$ implies that $1 \leq k - i \leq n$, let $k - n \leq i \leq k - 1$. In this case, there are two main situations,

- If $k \leq n$, then $k - n \leq 0$ and we sum from 1 to $k - 1$ of terms all equal to $1/n^2$. We deduce that in this case

$$P[X + Y = k] = \frac{k - 1}{n^2}.$$

- If $k > n$, then $n \leq k - 1$ and we sum from $k - n$ to n of terms all equal to $1/n^2$. In this case

$$P[X + Y = k] = \frac{2n + 1 - k}{n^2}$$

 **Exercise 2.3** Let X be a random variable taking its values in $\{0, 1, \dots, n\}$. Show that

$$E[X] = \sum_{n=0}^{n-1} P[X > n].$$

Solution The key is to write

$$P[X = n] = P[X > n - 1] - P[X > n]$$

According to the definition of expectation, we have

$$E[X] = \sum_{k=1}^n kP[X > k - 1] - \sum_{k=1}^n kP[X > k].$$


We perform an index shift in the first sum:

$$E[X] = \sum_{k=0}^{n-1} (k + 1)P[X > k] - \sum_{k=1}^n kP[X > n].$$

In the second sum, the term with index n is zero, and we can also add a term with index 0, which will also be zero. The result is

$$E[X] = \sum_{k=0}^{n-1} ((k + 1) - k)P[X > n] = \sum_{k=0}^{n-1} P[X > k],$$

and we have the result.

 **Exercise 2.4** Let X be a real random variable defined on a finite probability space. Show that

$$E[X]^2 \leq E[X^2].$$

Solution Let x_1, \dots, x_n be the values taken by the random variable X , with probabilities, respectively, p_k . We have

$$E[X] = \sum_{k=1}^n x_k p_k = \sum_{k=1}^n (x_k \sqrt{p_k}) \sqrt{p_k}.$$

We apply the Cauchy-Schwarz inequality with $a_k = x_k \sqrt{p_k}$ and $b_k = \sqrt{p_k}$. We deduce that

$$E[X]^2 \leq \left(\sum_{k=1}^n x_k^2 p_k \right) \left(\sum_{k=1}^n p_k \right) = E[X^2].$$

We can also use the fact that $V[X] = E[X]^2 - E[X^2]$ and that the variance is always positive

or zero.

- ✎ **Exercise2.5** Two players, A and B, place respective bets a and b to take part in the game. It is as follows: if the result of a roll is 1 or 2, A wins the bet otherwise it's B. What must be the respective stakes of these players for the game to be fair?

Solution Let X be the random variable for the earnings of player A. X takes the values $-a$ and $a + b$ with probabilities; respectively, $\frac{2}{6}$ and $\frac{4}{6}$.

The expectation of X is

$$\mathbf{E}[X] = \frac{2}{6}(a + b) + \frac{4}{6}(-a)$$

The game is fair if we have $\mathbf{E}[X] = a$.

- ✎ **Exercise2.6** Let f be the function defined on \mathbb{R} by

$$f(x) = \frac{a}{1 + x^2}$$

- Determine the value of a so that f is a probability density.
- Determine the distribution function of X . Does X have an expectation?

Solution

- (a). The function f is continuous on \mathbb{R} and positive. It will be the density of a random variable if it is integrable and if $\int_{-\infty}^{\infty} f(x)dx = 1$. Now, a primitive of $\frac{1}{1 + x^2}$ is $\arctan(x)$. So, we have

$$\int_{-\infty}^{\infty} f(x)dx = a \lim_{x \rightarrow \infty} \arctan(x) - a \lim_{x \rightarrow -\infty} \arctan(x) = a\pi$$

Thus, f is the probability density of a random variable X if and only if $a = \frac{1}{\pi}$.

- (b). In this case the distribution function of X is given by

$$F_X(t) = \int_{-\infty}^t \frac{1}{\pi(1 + x^2)} dx = \frac{1}{\pi} [\arctan(x)]_{-\infty}^t = \frac{1}{\pi} \arctan(t) + \frac{1}{2}.$$

Finally, X has no expectation because the function $xf(x)$ is not integrable on \mathbb{R} .

- ✎ **Exercise2.7** Let f be the function defined on \mathbb{R} by $f(x) = e^x$ if $x < 0$ and 0 if x else.

- Show that the function f is a probability density of a certain random variable, which we will call X .
- Determine the distribution function of X .
- Show that X has an expectation and calculate it.
- Let $Y = 2X + 1$
 - Determine the distribution function of Y .
 - Show that Y is a random variable with density, and determine the density of Y .
 - Repeat the same questions with $Y = X^2$.

Solution

- (a). The function f is positive and continuous. It is therefore the probability density of a

random variable X if $\lim_{x \rightarrow \infty} \int_x^0 f(t) dt = 1$. But,

$$\int_{-\infty}^{\infty} e^t dt = [e^t]_{-\infty}^0 = 1 - e^x \rightarrow 1$$

when $x \rightarrow -\infty$.

(b). X takes its values in $]-\infty, 0]$, so $F_X(t) = 1$ if $t \geq 0$. For $t < 0$, we have

$$F_X(t) = \int_{-\infty}^t e^x dx = e^t.$$

(c). The function xf , which is null on $[0, \infty[$ and continuous on \mathbb{R} except on 0. f is integrable in the neighbourhood of $-\infty$ because it is integrable in front of $1/x^2$ at this point. so f has an expectation. This expectation is calculated by integration by parts. Indeed,

$$\begin{aligned} \mathbf{E}[X] &= \int_{-\infty}^0 xe^x \\ &= [xe^x]_{-\infty}^0 - \int_{-\infty}^0 e^x dx \\ &= -1 \end{aligned}$$

(d). • Let $t \in \mathbb{R}$. We have $Y \leq t \Leftrightarrow 2X + 1 \leq t \Leftrightarrow X \leq (t - 1)/2$. So

$$F_Y(t) = F_X((t - 1)/2) = e^{(t-1)/2} \quad \text{if } t \leq 1 \quad \text{and } F_Y(t) = 1 \text{ if } t \geq 1$$

- The distribution function F_Y is derivative except at 1. The distribution function F_Y is derivable except at 1. From this we deduce that Y has a density given, for all $t \neq 1$, by $f(t) = F'_Y(t) = \frac{1}{2}e^{(t-1)/2}$ and for $t \geq 1$, $f_Y(t) = 0$.
- We repeat the same process. We have

$$Y \leq t \Leftrightarrow X^2 < t \Leftrightarrow -\sqrt{t} \leq X \leq \sqrt{t} \quad \text{for } t \geq 0$$

Since X has values in \mathbb{R} . we get

$$f(t) = \frac{1}{2\sqrt{t}}e^{\sqrt{t}} \quad \text{if } t \geq 0, \quad f(t) = 0.$$

 **Exercise 2.8** Let X be a Gaussian random variable with the distribution $\mathcal{N}(0, 1)$ and $x > 0$.

(a). Prove that

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^2} \right) e^{-x^2/2} \leq P(X > x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

(b). Prove that $P(X > x) \leq e^{-x^2/2}$.

Solution

(a). Calculate the derivative of the right-hand side

$$\begin{aligned} \frac{1}{dx} \left(\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \right) &= -\frac{1}{\sqrt{2\pi}} \frac{1}{x^2} (1 + x^2) e^{-x^2/2} \\ &\leq -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned}$$

and, since $P(X > x) = \int_x^{\infty} e^{-t^2/2} dt$, the upper bound is obtained by integrating between

x and $+\infty$. Similarly, the derivative of the left-hand side is

$$\begin{aligned} \frac{d}{dx} \left(\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^2} \right) e^{-x^2/2} \right) &= -\frac{1}{\sqrt{2\pi}} \left(1 - \frac{3}{x^4} \right) e^{-x^2/2} \\ &\geq -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \end{aligned}$$

the lower bound is obtained by integrating between x and $+\infty$.

- (b). If we pose $f(x) = e^{-x^2/2} - P(X > x)$ then $f'(x) = \left(\frac{1}{\sqrt{2\pi}} - x \right) e^{-x^2/2}$ so f is so f is increasing on $\left[0, \frac{1}{\sqrt{2\pi}}\right]$ and decreasing on $\left[\frac{1}{\sqrt{2\pi}}, +\infty\right]$. Since we have $f(0) = 0$ and $\lim_{x \rightarrow +\infty} f(x) = 0$, this means that f is positive on \mathbb{R}^+ .

-  **Exercise 2.9** A random variable is called a Rayleigh random variable if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$

- (a). Determine the corresponding cumulative density function.
 (b). Sketch $f_X(x)$ and $F_X(x)$ for $\sigma = 1$.

Solution The cdf of X is

$$F_X(x) = \int_0^x \frac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} dt \quad x \geq 0.$$

Let $y = t^2/(2\sigma^2)$. Then $dy = (1/\sigma^2)t dt$, and

$$F_X(x) = \int_0^{x^2/(2\sigma^2)} e^{-y} dy = 1 - e^{-x^2/(2\sigma^2)}$$

-  **Exercise 2.10** Prove or disprove

- (a) If $P\{X > Y\} = 1$, then $\mathbf{E}[X] > \mathbf{E}[Y]$.
 (b) If $\mathbf{E}[X] > \mathbf{E}[Y]$, then $P\{X > Y\} = 1$.
 (c) If $\mathbf{E}[X] > \mathbf{E}[Y]$, then $P\{X > Y\} > 0$.


-  **Exercise 2.11** Prove or disprove


- (a) If $F_X(x) > F_Y(x)$ for all x , then $\mathbf{E}[Y] > \mathbf{E}[X]$.
 (b) If $\mathbf{E}[Y] > \mathbf{E}[X]$, then $F_X(x) > F_Y(x)$ for all x .
 (c) If $F_X(x) = F_Y(x)$ for all x , then $P\{X = Y\} = 1$.
 (d) If $Y = X + 1$, then $F_X(x) = F_Y(x + 1)$ for all x .

-  **Exercise 2.12** If X_1 and X_2 are independent random variables with distribution given by


$$P\{X_i = -1\} = P\{X_i = 1\} = \frac{1}{2} \quad \text{for } i = 1, 2$$


then are X_1 and X_1X_2 independent ?


-  **Exercise 2.13** If X is binomially distributed with parameters n and p , what is the distribution of $Y = n - X$?

-  **Exercise 2.14** In a town with 5000 adults, a sample of 100 is asked their opinion of a proposed municipal project; 60 are found to favor it, and 40 oppose it. If, in fact, the adults of the town were equally divided on proposal, what would be the probability of obtaining a majority of 60

or more favoring it in a sample of 100 ?

 **Exercise2.15** Two players A and B are taking part in a game with respective probabilities of winning each game p and $q = 1 - p$. The winner is the one who first obtains two more wins than the other. What is the probability of each player winning?

 **Exercise2.16** Let X be the life in hours of a radio tube. Assume that X is normally distributed with mean 20 and variance σ^2 . If a purchaser of such radio tubes requires that at least 90 percent of the tubes have lives exceeding 150 hours, what is the largest value σ can be and still have the purchaser satisfied ?

 **Exercise2.17** Suppose that the cumulative distribution function (c.d.f) $F_X(x)$ can be written as a function of $(x - \alpha)/\beta$, where α and $\beta > 0$ are constants ; that is, x , and β appears in $F_X(\cdot)$ only in the indicated form.

- Prove that if α is increased by δx , then so is the mean of X .
- Prove that if β is multiplied by $\lambda (\lambda > 0)$, then standard deviation of X .

 **Exercise2.18**

(a). Make use of characteristic functions to determine the distribution of $X_1 + X_2$, where X_1 and X_2 are independent random variables, satisfying

- $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$
- $X_1 \sim Poi(\lambda_1)$ and $X_2 \sim Poi(\lambda_2)$

(b). Let X_1, X_2, \dots, X_n be *i.i.d.* real valued random variables with distribution $Poi(\lambda_i)$ Determine the characteristic function of the random sum $S = \sum_{i=1}^n X_i$.

 **Exercise2.19** Let X and Y be independent discrete random variables such that

$$\begin{aligned} P(X = 0) = \frac{1}{2} \quad P(X = 1) = \frac{1}{4} \quad P(X = 2) = \frac{1}{4} \\ P(Y = 1) = \frac{1}{3} \quad P(Y = 3) = \frac{2}{3} \end{aligned}$$

Use generating function to find the probability function of $X + Y$.

 **Exercise2.20** Let X and Y be independent discrete random variable such that

$$\begin{aligned} P(X = 0) = \frac{1}{2} \quad P(X = 1) = \frac{1}{4} \quad P(X = 2) = \frac{1}{4} \\ P(Y = 1) = \frac{1}{4} \quad P(Y = 3) = \frac{3}{4} \end{aligned}$$


Find the moment generating function of $X - Y$.

 **Exercise2.21**

- Calculate the characteristic function of an exponential random variable with parameter λ .
- Prove that

$$M_X(t) = \frac{e^{3it-2t^2}}{1+it}$$

is a characteristic function and find $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

 **Exercise2.22** Let X be a non-negative, integer-valued random variable satisfying:

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

$$P(X = 2n) = \frac{1}{2}P(X = 2n - 1) = 2/3P(X = 2n + 1), \quad n \geq 1.$$

Compute its probability generating function.

 **Exercise 2.23**

- (a). Use the moment generating function to compute $\mathbb{E}[X^4]$, where $X \sim \text{Bin}(n, p)$.
- (b). The random variable X has the property that $\mathbb{E}[X^n] = 3^n/(n + 1)$, $n = 1, 2, \dots$. Find the unique distribution of X having these moments.
- (c). Use moment generating function to prove that, if a random variable X has density function

$$f_X(x) = \frac{1}{2}e^{-|x|}$$

then X can be written as $X = Y - Z$, where Y and Z are independent, exponentially distributed random variables.

Chapter Vector Random variables

Introduction

- *Vector random variables*
- *Joint distribution*
- *covariance matrix*
- *Conditional Laws*
- *Independence*

3.1 Introduction

In many applications it is important to study two or more random variables defined on the same sample space. In this chapter, we first consider the case of two random variables, their associated distribution, and some proprieties, such as independence of the r.v.'s. These concepts are then extended to the case of many r.v.'s defined on the same sample space.

3.2 Bivariate distribution

Let X and Y be two discrete random variables with

$$X(\Omega) = \{x_i, i \in \mathbb{N}\} \text{ and } Y(\Omega) = \{y_j, j \in \mathbb{N}\}.$$

The joint distribution of (X, Y) is given by $(X, Y)(\Omega)$ (or by $X(\Omega)$ and $Y(\Omega)$) and by the probabilities

$$P(X = x, Y = y) = P(\omega, X(\omega) = x \text{ and } Y(\omega) = y), \text{ for any pair } (x, y) \in (X, Y)(\Omega).$$

Remark We must, of course, have $\sum_{x,y} P(X = x, Y = y) = 1$.

3.2.1 Joint Distribution

More generally, if X_1, \dots, X_n are n discrete random variables with values in \mathbb{N} , the joint distribution of the vector (X_1, \dots, X_n) is given by the image set joint law of the vector (X_1, \dots, X_n) is given by the image set $(X_1, \dots, X_n)(\Omega) \subset \mathbb{N}^n$. and by the probabilities $P(X_1 = i_1, \dots, X_n = i_n)$, for any n -uplet $(i_1, \dots, i_n) \in \mathbb{N}^n$.

Example 3.1 Let $p \in]0, 1[$ and $\lambda > 0$ and consider the pair of random variables (X, Y) with

values in $\{0, 1\} \times \mathbb{N}$ whose distribution is given by:

$$P(X = 0, Y = 0) = 1 - p$$

$$P(X = 1, Y = k) = p e^{-\lambda} \lambda^k / k!, \text{ for any } k \in \mathbb{N}$$

$$P(X = j, Y = k) = 0 \text{ else}$$

We have $\sum_{x,y} P(X = x, Y = y) = 1$: we have therefore written the distribution of a discrete random pair.

Example3.2 An urn containing four tokens numbered from 1 to 4 is used, and two tokens are drawn one after the other. two tokens one after the other. The results of the two draws are given as (X, Y) . We have : $P(X = i, Y = i) = 0$ for all i between 1 and 4 and $P(X = i, Y = j) = 1/12$ if $1 \leq i, 4 \leq j$ and $i \neq j$. The probabilities can be written in the form of the following table (where, for example, in the in the second box of the first line, we read $P(X = 1, Y = 2)$):

$X \setminus Y$	1	2	3	4
1	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{12}$
4	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0

Example3.3 [Trinomial distribution.] Given a strictly positive integer n and two real positive parameters positive real p_x and p_y such that $p_x + p_y \leq 1$. The trinomial distribution (n, p_x, p_y) is the distribution for the pair (X, Y) such that $(X, Y)(\Omega) \in \mathbb{N}^2$ and given for all $(i, j) \in \mathbb{N}^2$ such that $i + j \leq n$ by:

$$P(X = i, Y = j) = \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1 - p_x - p_y)^{n-i-j}$$

and $P(X = i, Y = j) = 0$ otherwise.

Example3.4 Show that this is the correct definition of the distribution of a random pair. The trinomial distribution is an extension of the binomial distribution. Imagine an experiment which has three possible outcomes, denoted x, y and z , with probabilities p_x, p_y and $p_z = 1 - p_x - p_y$. Let's repeat this experiment n times (n is fixed) in an independent manner and count the number of appearances of x (number noted X) and of y (noted Z) among these n repetitions. This is an exercise in counting to show that the pair (X, Y) follows a trinomial distribution of parameters trinomial distribution of parameters (n, p_x, p_y) .

3.2.2 Marginal distribution

Definition 3.1

The (two) marginal distributions of the pair (X, Y) are the laws of the random variables X and Y . They are obtained as follows:

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y),$$

$$P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y).$$



Proof We have

$$\{X = x\} = \{X = x, Y \in Y(\Omega)\} = \bigcup_{y \in Y(\Omega)} \{X = x, Y = y\}$$

Since the union is countable and disjoint, it follows:

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y).$$

□

More generally, a vector (X_1, \dots, X_n) with values in \mathbb{Z}^n has n marginal laws one-dimensional marginal laws, but also $n(n-1)$ two-dimensional marginal laws, and so on.

For example, we have

$$P(X_1 = x) = \sum_{(x_2, \dots, x_n) \in \mathbb{Z}^{n-1}} P(X_1 = x, X_2 = x_2, \dots, X_n = x_n).$$

Let's go back to the previous examples:

Example 3.5 Let's determine the distribution of X : X has values in $\{0, 1\}$ and we have:

$$P(X = 0) = \sum_{j \in \mathbb{N}} P(X = 0, Y = j) = 1 - p.$$

In the same way:

$$P(X = 1) = \sum_{j \in \mathbb{N}} P(X = 1, Y = j) = p e^{-\lambda} \lambda^j / j! = p.$$

The random variable X therefore follows a Bernoulli distribution with parameter p . Let's also calculate the distribution of Y :

$$P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = (1-p) + p e^{-\lambda}.$$

And for all $j \geq 1$,

$$P(Y = j) = P(X = 0, Y = j) + P(X = 1, Y = j) = p e^{-\lambda} \lambda^j / j!.$$

3.2.3 Distribution of $f(X, Y)$

We have a pair of discrete random variables (X, Y) whose joint distribution we know, and we would like to know the distribution of the random variable $Z = f(X, Y)$, where $f: X(\Omega) + Y(\Omega) \rightarrow \mathbb{R}$ is a given function. For example, we often need to know the distribution of $X + Y$, or that of $X + Y$, or of $X - Y$. And determining the law distribution of X from that of (X, Y) is the same as considering the function $f(x, y) = x$. On a $Z(\Omega) = f((X, Y)(\Omega))$ and for all $z \in f((X, Y)(\Omega))$, on a $Z(\Omega) = f((X, Y)(\Omega))$, we have

$$P(Z = z) = \sum_{(x,y) \in (X,Y)(\Omega), f(x,y)=z} P(X = x, Y = y).$$

Proposition 3.1

On a $Z(\Omega) = f((X, Y)(\Omega))$ et pour tout $z \in f((X, Y)(\Omega))$, on a

$$P(Z = z) = \sum_{(x,y) \in (X,Y)(\Omega), f(x,y)=z} P(X = x, Y = y)$$



Example 3.6 Let's take example 1 again and consider the function $f(x, y) = xy$. The random variable XY has values in \mathbb{N} and we have

$$P(XY = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = 1 - p - pe^{-\lambda}$$

and for $k \in \mathbb{N}^*$,

$$P(XY = k) = P(X = 1, Y = k) = pe^{-\lambda} \frac{\lambda^k}{k!}$$

An important particular case.

Here we consider the function $f(x, y) = x + y$. The result is:

$$\begin{aligned} P(X + Y = z) &= \sum_{(x,y) \in (X,Y)(\Omega), x+y=z} P(X = x, Y = y) \\ &= \sum_{x \in X(\Omega)} P(X = x, Y = z - x) \\ &= \sum_{y \in Y(\Omega)} P(X = z - y, Y = y) \end{aligned}$$

More generally, if $\bar{X} = (X_1, \dots, X_n)$ is a discrete random vector and

$$f: \bar{X}(\Omega) \rightarrow \mathbb{R}$$

is a given function, we have :

$$P(f(X_1, \dots, X_n) = z) = \sum_{x_1, \dots, x_n, f(x_1, \dots, x_n)=z} P(X_1 = x_1, \dots, X_n = x_n)$$

This implies the following fundamental proposition.

Proposition 3.2. L

Let X and Y be two discrete integrable random variables. Then the random variable $Z = X + Y$ is integrable and we have $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.



Proof We have just seen that Z is a discrete random variable and that its distribution is given by: for all

$$z \in (X + Y)(\Omega),$$

$$P(X + Y = z) = \sum_{x \in X(\Omega)} P(X = x, Y = z - x).$$

We have

$$P(|X + Y = z|) = \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |z| P(X = x, Y = z - x) \right]$$

so

$$P(|X + Y = z|) = \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |x + (z - x)| P(X = x, Y = z - x) \right]$$

Using the triangular inequality, we get:

$$\begin{aligned} P(|X + Y = z|) &\leq \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |x + (z - x)| P(X = x, Y = z - x) \right] \\ &\leq \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |x| P(X = x, Y = z - x) \right] \\ &\quad + \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |z - x| P(X = x, Y = z - x) \right] \end{aligned}$$

Let's start with the first sum:

$$\begin{aligned} \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |x| P(X = x, Y = z - x) \right] &= \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |x| P(X = x, Y = z - x) \right] \\ &= \sum_{x \in X(\Omega)} [|x| P(X = x)] = E(|Y|) \end{aligned}$$

Let's turn to the second sum: sum over $x \in X(\Omega)$ or over x such that $z - x \in Y(\Omega)$ does change the value of this sum. The result is

$$\begin{aligned} \sum_{z \in (X+Y)(\Omega)} \left[\sum_{x \in X(\Omega)} |z - x| P(X = x, Y = z - x) \right] &= \sum_{z \in (X+Y)(\Omega)} \left[\sum_{(z-x) \in Y(\Omega)} |z - x| P(X = x, Y = z - x) \right] \\ &= \sum_{z \in (X+Y)(\Omega)} \left[\sum_{y \in Y(\Omega)} |y| P(X = z - y, Y = y) \right] \\ &= \sum_{y \in Y(\Omega)} \left[|y| \sum_{z \in (X+Y)(\Omega)} P(X = z - y, Y = y) \right] \\ &= \sum_{y \in X(\Omega)} [|y| P(y = y)] = E(|Y|) \end{aligned}$$

We can therefore see that if X and Y are integrable, so is $X + Y$, and by performing the

same calculation without absolute values, we conclude that $E(X + Y) = E(X + E(Y))$. Same calculation without absolute values, we conclude that $E(X + Y) = E(X) + E(Y)$. \square

Example 3.7 Assume that (X, Y) follows a trinomial distribution (n, p_x, p_y) and calculate the distribution of $X + Y$. This random variable has values in \mathbb{N} and we have, for any integer k :

$$P(X + Y = k) = \sum_{j=1}^n P(X = j, Y = k - j).$$

For all $k > n$, each of the terms in this sum is zero, so $P(X + Y = k) = 0$. Now let's fix an integer $k \in \{0, \dots, n\}$. We have :

$$\begin{aligned} P(X + Y = k) &= \sum_{j=1}^n P(X = j, Y = k - j) \\ &= \sum_{j=0}^n \frac{n!}{j!(k-j)!(n-k)!} p_x^j p_y^{k-j} (1 - p_x - p_y)^{n-k} \\ &= \frac{n!}{(n-k)!} (1 - p_x - p_y)^{n-k} \sum_{j=0}^n \frac{1}{j!(k-j)!} p_x^j p_y^{k-j} \\ &= \frac{n!}{k!(n-k)!} (1 - p_x - p_y)^{n-k} (p_x + p_y)^k \end{aligned}$$

The random variable $X + Y$ therefore follows a binomial distribution $Bin(n, p_x + p_y)$.

3.2.4 Conditional law

Consider a pair (X, Y) of discrete random variables, whose joint distribution we know and let y be such that $P(Y = y) > 0$. The conditional distribution of X knowing the event $Y = y$ is given by the fact that it is a law on $X(\Omega)$ and by the conditional probabilities $P(X = x|Y = y)$ for all $x \in X(\Omega)$. It is easy to check that

$$\sum_{x \in X(\Omega)} P(X = x|Y = y) = 1$$

which implies that the conditional law of X knowing $\{Y = y\}$ is the law of a random variable.

Furthermore, the total probability formula implies that

$$P(X = x) = \sum_{y \in Y(\Omega)} P(X = x|Y = y)P(Y = y)$$

This definition of the conditional law can be extended to random vectors: for example for a random triplet (X, Y, Z) , we can study the conditional law of X knowing $\{Y = y\}$, the conditional law of conditional law of X knowing $\{Y = y; \text{ and } Z = z\}$, the conditional law of the pair (X, Y) knowing $Z = z$.

Example 3.8 The law of Y knowing $X = 1$ is given by $Y(\Omega) \in \mathbb{N}$ and, for any positive integer k , we have

$$P(Y = k|X = 1) = \frac{P(Y = k \text{ and } X = 1)}{P(X = 1)} = pe^{-\lambda} \frac{\lambda^k}{k!} \times \frac{1}{p} = e^{-\lambda} \frac{\lambda^k}{k!}$$

The conditional distribution of Y given that $X = 1$ is therefore a Poisson distribution of

parametric λ .

Example 3.9

Example 3.10 The law of Y knowing $X = 1$ is given by $Y(\Omega) \in \mathbb{N}$ and, for any positive integer k , we have

$$P(Y = k|X = 1) = \frac{P(Y = k \text{ and } X = 1)}{P(X = 1)} = pe^{-\lambda}$$

Suppose that (X, Y) follows a trinomial distribution (n, p_x, p_y) and let's calculate the conditional distribution of X knowing $\{Y = k\}$, of X knowing $\{Y = k\}$, for an integer $k \in \{0, \dots, n\}$. Let us first note that if $j < 0$ or $j > n - k$, $P(X = j|Y = k) = 0$. Let us now fix an integer $j \in \{0, \dots, n - k\}$. We have

$$\begin{aligned} P(X = j|Y = k) &= \frac{P(X = j \text{ et } y = k)}{P(Y = k)} \\ &= \frac{n! p_x^j p_y^k (1 - p_x - p_y)^{n-j-k}}{j! k! (n - j - k)!} \frac{k! (n - k)!}{n! p_y^k (1 - p_y)^{n-k}} \\ &= \frac{(n - k)!}{j! (n - k - j)!} \left(\frac{p_x}{1 - p_y} \right)^j \left(\frac{1 - p_x - p_y}{1 - p_y} \right)^{n-k-j} \end{aligned}$$

The conditional distribution of X knowing $\{Y = k\}$ is therefore a binomial distribution $(n - k, p_x/(1 - p_y))$.

3.2.5 Independence of discrete random variables

Definition 3.2. Independence of random variables

Two discrete random variables X and Y are said to be independent if for any x in $X(\Omega)$ and any y in $Y(\Omega)$, the events $X = x$ and $Y = y$ are independent, i.e. :

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

If X and Y are two independent discrete random variables, then for any $y \in Y(\Omega)$ and any $x \in X(\Omega)$ such that $P(Y = y) > 0$ and $P(X = x) > 0$, $P(X = x|Y = y) = P(X = x)$ and $P(Y = y|X = x) = P(Y = y)$. More generally, the n discrete random variables X_1, \dots, X_n are (mutually or n to n) independent if, for any choice of $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n).$$



- Independence n to n leads to independence 2 to 2 (but the converse is false). Write the proof of this result for $n = 3$.
- Events A_1, \dots, A_n are independent if and only if the random variables $1_{A_1}, \dots, 1_{A_n}$ are independent.

If the n discrete random variables X_1, \dots, X_n are independent then we have

$$\overline{X}(\Omega) = X_1(\Omega) \times \dots \times X_n(\Omega).$$

Remark When the random variables X and Y are independent, knowing the marginal distributions allows us to know the joint distribution of the pair (X, Y) , whereas for any random variables, this is not sufficient.

3.2.6 Expectation, covariance matrix

For the sake of clarity, all the results in this paragraph and the next are stated for pairs of discrete random variables, and they easily extend to random vectors. pairs of discrete random variables, and they easily extend to discrete random vectors. random vectors. Consider a discrete random pair (X, Y) .

Definition 3.3

- The expectation of the pair (X, Y) is defined if X and Y are integrable and we have then: $E(X, Y) = (E(X)E(Y))$.
- If X and Y are two random square integrable variables, the covariance of X and Y , or covariance of the pair (X, Y) , is given by
- The expectation of the pair (X, Y) is defined if X and Y are integrable and we have then

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E[(X - E(X))(Y - E(Y))]$$

bullet If X and Y are two random variables of integrable square, the covariance of X and Y ,

- If X and Y are two random variables with integrable squares, the covariance matrix of the pair pair (X, Y) is the matrix

$$C = \begin{pmatrix} V(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & V(Y) \end{pmatrix}$$

More generally, the covariance matrix of a vector (X_1, \dots, X_n) , each of whose components has an integrable square, is a matrix $n \times n$ whose diagonal terms are the variances of X_i and whose (i, j) term is the covariance $\text{cov}(X_i, X_j)$ for all $i \neq j$.



Remark Calculating the expectation of X or Y involves only marginal distributions, but we shall see that it is not necessary to explain these marginal distributions.

Proposition 3.3. I

(X, Y) is a pair of discrete random variables, for any function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that

$$\sum_{x \in X(\Omega), y \in Y(\Omega)} |h(x, y)| P(X = x, Y = y) < \infty$$

the random variable $h(X, Y)$ is integrable and we have

$$E(h(X, Y)) = \sum_{x \in X(\Omega), y \in Y(\Omega)} h(x, y) P(X = x, Y = y)$$



Proof Let $Z = h(X, Y)$. We have an explicit statement of the distribution of Z : this random variable is discrete and for all $z \in (h(X, Y))(\Omega)$,

$$P(Z = z) = \sum_{(x, y) \in (X, Y)(\Omega), h(x, y) = z} P(X = x, Y = y)$$

The random variable Z is therefore integrable if the sum of

$$S = \sum_{z \in (h(X, Y))(\Omega)} \left(|z| \sum_{(x, y) \in (X, Y)(\Omega), h(x, y) = z} P(X = x, Y = y) \right)$$

converges. This sum can be written as

$$\begin{aligned} S &= \sum_{z \in (h(X, Y))(\Omega)} \left(\sum_{(x, y) \in (X, Y)(\Omega), h(x, y) = z} |h(x, y)| P(X = x, Y = y) \right) \\ &= \sum_{(x, y) \in (X, Y)(\Omega)} |h(x, y)| P(X = x, Y = y) \end{aligned}$$

The same procedure applied to the sum without absolute values gives the desired result for desired result for $\mathbf{E}(h(X, Y))$. \square

Application.

This proposition makes it possible to write the expectation of X , Y or XY without specifying the law of these random variables. If the pair (X, Y) is discrete, we have 0

$$\begin{aligned} E(X) &= \sum_{(x, y) \in (X, Y)(\Omega)} x P(X = x, Y = y) \\ E(Y) &= \sum_{(x, y) \in (X, Y)(\Omega)} y P(X = x, Y = y) \\ E(XY) &= \sum_{(x, y) \in (X, Y)(\Omega)} xy P(X = x, Y = y) \end{aligned}$$

when the series converge. Now back to the covariance matrix:

Proposition 3.4

- A covariance matrix C is always a symmetrical and positive matrix (i.e., for all $v \in \mathbb{R}_n$, $\langle v, Cv \rangle \geq 0$).
- If X and Y are two independent and integrable random variables, we have $\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y)$ and so $\text{cov}(X, Y) = 0$. The reciprocal of this result is false.
- If X and Y are two independent random variables and f and g are two functions such that the random variables $f(X)$ and $g(Y)$ are integrable, then $\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y))$. The reciprocal of this result is false.

(d). If the random variables X_1, \dots, X_n are independent and of integrable square, then the covariance matrix of (X_1, \dots, X_n) is diagonal. covariance matrix of (X_1, \dots, X_n) is diagonal.

The converse of this result is false.



- (a). Let $\bar{X} = (X_1, \dots, X_n)$ be a random vector each component of which has an integrable square. integral, let C be its covariance matrix and let $\bar{v} = (v_1, \dots, v_n)$ be a vector of R^n . If we replace each X_i by $X_i - E(X_i)$, we will assume that the random variables are centred (of zero expectation). (of zero expectation). We have

$$\begin{aligned} \langle v, Cv \rangle &= \sum_{i,j} C_{ij} v_i v_j \\ &= \sum_{i,j} E(X_i X_j) v_i v_j \\ &= E \left(\sum_{i,j} v_i X_i v_j X_j \right) \\ &= E \left(\left(\sum_i v_i X_i \right) \times \left(\sum_j v_j X_j \right) \right) \\ &= E \left(\left(\sum_i v_i X_i \right)^2 \right) \end{aligned}$$

So, we deduce $\langle v, Cv \rangle \geq 0$.

- (b). Consider two independent and integrable discrete random variables X and Y . Let us show that XY is integrable and calculate its expectation. We have

$$E(|XY|) = \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| P(X = x, Y = y)$$

As X et Y are independents, $P(X = x, Y = y) = P(X = x)P(Y = y)$ so

$$\begin{aligned} \mathbf{E}(|XY|) &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| P(X = x)P(Y = y) \\ &= \sum_{x \in X(\Omega)} |x| P(X = x) \sum_{y \in Y(\Omega)} |y| P(Y = y) \\ &= \left(\sum_{x \in X(\Omega)} |x| P(X = x) \right) \left(\sum_{y \in Y(\Omega)} |y| P(Y = y) \right) \\ &= \mathbf{E}(|X|) \mathbf{E}(|Y|) \end{aligned}$$

We then use a similar calculation to show that $E(XY) = E(X)E(Y)$, then deduce that $cov(X, Y) = 0$.

- (c). **3** and **4** are easily deduced from **2**.

Although the following proposition, very simple to prove, is very useful:


Proposition 3.5

If X and Y are two random variables with integrable squares, we have

$$V(X + Y) = V(X) + V(Y) + 2\text{cov}(X, Y)$$


In addition, if X and Y are independent, we have

$$V(X + Y) = V(X) + V(Y)$$

In addition, for two random variables with integrable squares, it is possible to obtain a value for $\text{cov}(X, Y)$ from the variances of X and Y : 

Proposition 3.6. Cauchy Schwarz Inequality

Let X and Y two random variables with integrable squares. We have

$$\begin{aligned} |\mathbf{E}(XY)| &\leq \mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)} \\ |\text{cov}(XY)| &\leq \sqrt{\text{var}(X)\text{var}(Y)} \end{aligned}$$


Proof The second point is obtained by applying the first to $X - \mathbf{E}(X)$ and $Y - \mathbf{E}(Y)$. The first point is shown in the same way as in the classical case (scalar product of two vectors): two vectors): we study the polynomial

$$R(\lambda) = \mathbf{E}(X^2)\lambda^2 - 2\mathbf{E}|XY|\lambda + \mathbf{E}(Y^2)$$

This polynomial can be factored into

$$R(\lambda) = \mathbf{E}[(\lambda|X| - |Y|)^2]$$

It therefore has no two distinct real roots, which means that its (reduced) discriminant is negative:

$$(\mathbf{E}|XY|)^2 - \mathbf{E}(X^2)\mathbf{E}(Y^2) \leq 0$$

In the same way, we study the case of equality: the discriminant is zero only if the polynomial has a double real root. In this case, there exists λ_0 such that

$$\mathbf{E}[(\lambda_0|X| + |Y|)^2] = 0,$$

the expectation of a positive random variable can only be zero if the random variable is (almost surely) zero, we then have: $|Y| = \lambda_0|X|$, almost surely.

The proofs without absolute values are similar. □


3.2.7 Generating function of a pair**Definition 3.4. Generating function**

The generating function of a pair of positive discrete random variables is the function

defined on $[0, 1]^2$ by

$$G_{(X,Y)}(s, t) = E(s^X t^Y) = \sum_{x \in X(\Omega), y \in Y(\Omega)} s^x t^y P(X = x, Y = y)$$



 **Note** The generating function determines the law of the pair (X, Y) in the sense that if two pairs of positive random variables have the same generating function, then they follow the same law. same law.

The generating function of (X, Y) allows us to find, for example:

- The generating function of X : $G_X(s) = G_{(X,Y)}(s, 1)$ et de Y ,
- The expectation of X if X is integrable : $\mathbf{E}(X) = \frac{\partial}{\partial s} G_{(X,Y)}(1, 1)$,
- The expectation of X^2 if X : $\mathbf{E}(X^2) = \frac{\partial^2}{\partial s^2} G_{(X,Y)}(1, 1) + \frac{\partial}{\partial s} G_{(X,Y)}(1, 1)$,
- The expectation of XY if X and Y are square integrable: $\mathbf{E}(XY) = \frac{\partial^2}{\partial s \partial t} G_{(X,Y)}(1, 1)$.

Another important use of the generating function is that it allows us to simply calculate the law of the sum of random variables, as soon as we know their joint law. In fact, we have the following result:

Proposition 3.7

Let (X, Y) be a pair of positive random variables whose joint distribution we know. joint distribution. Let $G_{X,Y}$ be the generating function of the pair. Then $G_{X+Y}(s) = G_{(X,Y)}(s, s)$.



Proof We have $G_{X,Y}(s, s) = E(s^X s^Y) = E(s^{X+Y}) = G_{X+Y}(s)$. If X and Y have integer values, it only remains to decompose the function G_{X+Y} into a series to obtain the probabilities $P(X + Y = k) =$. \square

It is easy to see that if X and Y are two independent positive random variables, then $G_{(X,Y)}(s, t) = G_X(s) G_Y(t)$. The reciprocal of this result is also true:

Theorem 3.1

The positive random variables X and Y are independent if and only if

$$G_{(X,Y)}(s, t) = G_X(s) \cdot G_Y(t).$$



Remark

- The generating function of the random vector $\bar{X} = (X_1, \dots, X_n)$ satisfying $\bar{X}(\Omega) \subset \mathbb{N}^n$ is the function $G : [0; 1]^n \rightarrow \mathbb{R}^+$ defined by

$$G(s_1, \dots, s_n) = E(s^{X_1} \times \dots \times s_n^{X_n}).$$

- This function of n variables characterises the law of the n -uplet, making it possible to find the n marginal laws.

3.2.8 Example of using the generating function

Let's calculate the generating function of the trinomial distribution (n, p_x, p_y) . Let (X, Y) be a pair of random variables with a trinomial distribution (n, p_x, p_y) . We have

$$\begin{aligned}
 G_{(X,Y)}(s, t) &= E(s^X t^Y) \\
 &= \sum_{i, j \in \{0, \dots, n\}, i+j \leq n} s^i t^j P(X = i, Y = j) \\
 &= \sum_{i=0}^n \sum_{j=0}^n s^i t^j \frac{n!}{i!j!(n-i-j)!} p_x^i p_y^j (1-p_x-p_y)^{n-i-j} \\
 &= \sum_{i=0}^n \frac{n!}{i!(n-i)!} (s p_x)^i \sum_{j=0}^n \frac{(n-i)!}{j!(n-i-j)!} (t p_y)^j (1-p_x-p_y)^{n-i-j} \\
 &= \sum_{i=0}^n \frac{n!}{i!(n-i)!} (s p_x)^i (1-p_x-(1-t)p_y)^{n-i} \\
 &= (1 - (1-s)p_x - (1-t)p_y)^n
 \end{aligned}$$

By taking $t = 1$ in the above result, we find: $G(X, Y)(s, 1) = (1 - (1-s)p_x)^n$; we recognise recognise the characteristic function of a binomial distribution (n, p_x) which proves that X follows a binomial distribution (n, p_x) . The generating function of (X, Y) can also be used to calculate the generating function of $X + Y$. We have we have

$$E(s^{X+Y}) = E(s^X s^Y) = G_{(X, Y)}(s, s)$$

Here we find: $G_{X+Y}(s) = (1 - (1-s)(p_x + p_y))^n$; we therefore find that $X + Y$ follows a binomial $(n, p_x + p_y)$.

3.2.9 Sum of two independent Poisson distribution random variables

Let's consider two independent random variables X and Y with a Poisson distribution of parameters $\mu_1 > 0$; and $\mu_2 > 0$ respectively, and let's calculate the distribution of $X + Y$.

First method

Let's start by calculating the generating function of the distribution of X . We have

$$\begin{aligned}
 G_X(s) &= E(s^X) \\
 &= \sum_{n \geq 0} s^n P(X = n) \\
 &= \sum_{n \geq 0} s^n e^{-\mu_1} \frac{\mu_1^n}{n!} \\
 &= e^{-\mu_1} e^{\mu_1 s} \\
 &= e^{-\mu_1(1-s)}
 \end{aligned}$$

In the same way, $G_Y(t) = e^{-\mu_2(1-t)}$. Since X and Y are independent random variables, we can

the generating function of the pair (X, Y) :

$$G_{(X,Y)}(s, t) = e^{-\mu_1(1-s)} e^{-\mu_2(1-t)}$$

The generating function of $X + Y$ is therefore

$$G_{(X+Y)}(s) = G_{(X,Y)}(s, s) = e^{-(\mu_1+\mu_2)(1-s)}$$

Here we recognise the generating function of a Poisson distribution with parameter $\mu_1 + \mu_2$ and we can conclude that $X + Y$ follows a Poisson distribution with parameter $\mu_1 + \mu_2$. conclude that $X + Y$ follows a Poisson distribution with parameter $\mu_1 + \mu_2$.

Warning

The random variable $X - Y$ does not follow a Poisson distribution. Since it takes its values in \mathbb{Z} instead of \mathbb{N} and, consequently, we cannot calculate its generating function

Second method

We obviously have $(X + Y)(\Omega) \subset \mathbb{N}$. Let us therefore fix $n \in \mathbb{N}$ and calculate the probability directly $P(X + Y = n)$. We have

$$\begin{aligned} P(X + Y = n) &= \sum_{k \geq 0} P(X = k, X + Y = n) \\ &= \sum_{k \geq 0} P(X = k, Y = n - k) \\ &= \sum_{k \geq 0} P(X = k)P(Y = n - k) \\ &= \sum_{k \geq 0} e^{-\mu_1} \frac{\mu_1^k}{k!} e^{-\mu_2} \frac{\mu_2^{n-k}}{(n-k)!} \\ &= e^{-(\mu_1+\mu_2)} \frac{1}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \mu_1^k \mu_2^{n-k} \\ &= e^{-(\mu_1+\mu_2)} \frac{(\mu_1 + \mu_2)^n}{n!} \end{aligned}$$

This means that $X + Y$ follows a Poisson distribution with parameter $\mu_1 + \mu_2$.

3.3 Random pairs with density

3.3.1 Density of a pair

Definition 3.5

A pair (X, Y) of real random variables is said to have a density with respect to the Lebesgue measure of \mathbb{R}^2 if there exists a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ such that, for all intervals I and J and for any continuous bounded or positive function h we have $P(X \in$

$$I \text{ et } Y \in J) = \int_{I \times J} f(x, y) dx dy \text{ et } E(h(X, Y)) = \int \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy$$



Remark

- The density f is always a positive function, integrable and whose integral on \mathbb{R}^2 with respect to the with respect to the Lebesgue measure is 1.
- It can be shown that the definition of density is equivalent to: for any open U of \mathbb{R}^2 .

$$P((X, Y) \in U) = \int_U f(x, y) dx dy.$$

Example 3.11

- (a). Let $D = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$ the disk unit and \mathbb{R}^2 et f \mathbb{R}^2 the function defined on by: $f(x, y) = 1_D(x, y)/\pi$. f is indeed a probability density on \mathbb{R}^2 : it is the density of the uniform distribution on D , i.e. the distribution obtained by throwing a point at random uniformly over D .
- (b). Let X be a random variable with an exponential distribution. The distribution of the pair (X, X) has no density with respect to the Lebesgue measure on \mathbb{R}^2 . Indeed, if such a density f existed, it would have to be zero (almost everywhere) outside the line $\{x = y\}$ since $P(X = Y) = 0$. We would then have

$$1 = \int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\{X=Y\} \subset \mathbb{R}^2} f(x, y) dx dy$$

3.3.2 Marginal distribution

The marginal distributions of the pair (X, Y) are the laws of the random variables X and Y . We can the following proposition:

Proposition 3.8

Let (X, Y) be a pair of random variables of (joint) density $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$. Then the two random variables X and Y each have a density, denoted by respectively f_X and f_Y given by:

$$f_X(x) = \int_{y \in \mathbb{R}} f(x, y) dy \text{ et } f_Y(y) = \int_{x \in \mathbb{R}} f(x, y) dx.$$



Proof Let $t \in \mathbb{R}$ and calculate the distribution function of X . We have

$$\begin{aligned} P(X \leq t) &= P(X \in]-\infty, t] \text{ et } Y \in \mathbb{R}) \\ &= \int_{]-\infty, t] \times \mathbb{R}} f(x, y) dx dy \\ &= \int_{-\infty}^t \left(\int f(x, y) dy \right) dx \end{aligned}$$

□

This principle applies to density vectors, except that the marginal laws of a triplet are the laws

of each of the variables that make it up, as well as the laws of pairs of variables.

3.3.3 Independence

Recall the definition of the independence of random variables:

Definition 3.6

The random variables X and Y are independent if, for all intervals I and J , we have

$$P(X \in I \text{ et } Y \in J) = P(X \in I)P(Y \in J).$$


More generally, n random variables X_1, \dots, X_n are (mutually) independent if, for all intervals I_1, \dots, I_n , we have

$$P\left(\bigcap_{k \leq n} \{X_k \in I_k\}\right) = \prod_{k \leq n} P\{X_k \in I_k\}$$



The link between density and independence is clear.

Theorem 3.2

Two random variables (X, Y) of density f and g respectively are independent if and only if the law of the pair has a density and this density is the function $(x, y) \rightarrow f(x)g(y)$. 

3.3.4 Expectation, covariance

Definition 3.7

As with discrete pairs, we define the expectation of a pair of integrable random easy to check that

$$E(X) = \int_{\mathbb{R}^2} x f(x, y) dx dy \text{ et } E(Y) = \int_{\mathbb{R}^2} y f(x, y) dx dy.$$

when these integrals are absolutely convergent. The non-diagonal terms of this matrix will be

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

with

$$E(XY) = \int_{\mathbb{R}^2} xy f(x, y) dx dy$$



Remark

- As in the discrete case, the covariance matrix is a symmetrical and positive matrix (in the sense of bilinear forms).
- To identify the density of a pair, we usually use a test function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ continuous and bounded and we try to write $E(h(X, Y))$ in the form

$$E(h(x, y)) = \int_{\mathbb{R}^2} h(x, y) f(x, y) dx dy$$

The function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ above, if it exists, will be the density of the pair (X, Y) .

The properties seen in the discrete case remain true for pairs with density: in particular, if the pair consists of independent random variables with integrable squares, $cov(X, Y) = 0$ and $var(X + Y) = 0$. and $var(X + Y) = var(X) + var(Y)$, and, as in the discrete case, the covariance of two random variables whose pair admits a density, can be zero without the variables being independent. are independent: take, for example, the example of the pair with uniform distribution on the disc of centre 0 and radius 1.

3.3.5 Sum law

We can check that if the pair (X, Y) has a density, the variable $X + Y$ will also have a density. a density, and we can explain this density.

Proposition 3.9

- If the pair (X, Y) has as its density the function f , then the density of the random variable $X + Y$ is the function g defined by

$$g(z) = \int_{\mathbb{R}} f(x, z - x) dx = \int_{\mathbb{R}} f(y, z - y) dy.$$

- If X and Y are two independent random variables of density f_X and f_Y , the density, denoted g , of $X + Y$ is the convolution product of f_X and f_Y :


$$g(z) = \int_{\mathbb{R}} f_X(x) f_Y(z - x) dx = \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy.$$

Proof Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a piecewise continuous bounded function. We have

$$\begin{aligned} Eh(X + Y) &= \int_{\mathbb{R}^2} h(x + y) f(x, y) dx dy \\ &= \int_{\mathbb{R}} h(z) \int_{\mathbb{R}} f(x, z - x) dx dz \end{aligned}$$

where $x = x$ and $z = x + y$. If we had changed the variables $z = x + y$ and y , we would have obtained the other expression announced. \square

3.4 Exercises

 **Exercise 3.1 [Simultaneous drawing from a ballot box]** Two balls are simultaneously drawn from an urn containing 4 balls indistinguishable by touch and numbered from 1 à 4. Let U be the number of the smallest ball, and V the number of the largest ball. Determine the joint distribution of (U, V) , then the distributions of U and V .

Solution The universal set Ω is $\Omega = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$ and we consider that all the events are equiprobable.

The joint distribution of U and V and the marginal distributions can be summarised using the

following table:

$U \setminus V$	2	3	4	$P(U = u)$
1	1/6	1/6	1/6	1/2
2	0	1/6	1/6	1/3
3	0	0	1/6	1/6
$P(V = v)$	1/6	1/3	1/2	

Exercise 3.2 [Couple of uniform random variables] Let (Ω, P) be a finite probability space and let $X : \Omega \rightarrow E$ and $Y : \Omega \rightarrow F$ be two random variables. Show that the following two assertions are equivalent:

- $(X, Y) \sim \mathcal{U}(E \times F)$.
- $X \sim \mathcal{U}(E)$, $Y \sim \mathcal{U}(F)$ and X and Y are independent.

Solution Let's start by proving **1.** \implies **2.**, which is the easiest way. Let $(x, y) \in E \times F$. Since X and Y are independent, we have

$$P((X, Y) = (x, y)) = P(X = x)P(Y = y) = \frac{1}{\text{card}(E)} \times \frac{1}{\text{card}(F)} = \frac{1}{\text{card}(E \times F)}$$

where we used that $X \sim \mathcal{U}(E)$, $Y \sim \mathcal{U}(F)$ and that $\text{card}(E \times F) = \text{card}(E) \times \text{card}(F)$.

Reciprocally, we suppose **1.** It is easy to find the marginal laws of X and Y . For example, for $x \in E$, we have

$$P(X = x) = \sum_{y \in F} P((X, Y) = (x, y)) = \sum_{y \in F} \frac{1}{\text{card}(E \times F)} = \frac{\text{card}(F)}{\text{card}(E \times F)} = \frac{1}{\text{card}(E)}$$

So we have $X \sim \mathcal{U}(E)$ and similarly $Y \sim \mathcal{U}(F)$. Finally X and Y are independent since

$$P((X, Y) = (x, y)) = P(X = x) \times P(Y = y)$$

from the formula $\text{card}(E \times F) = \text{card}(E) \times \text{card}(F)$.

Exercise 3.3 [Marginal law] Suppose we have n boxes numbered from 1 to n . Box k contains k balls numbered from 1 to k . A box is chosen randomly, and then a ball from this box. Let X be the number of the box and Y the number of the ball.

- Determine the joint distribution of (X, Y) .
- Deduce the distribution of Y .
- Calculate the expectation of Y .

Solution

(a). For all (i, j) in $\{1, \dots, n\}^2$, we have

$$P((X, Y) = (i, j)) = P(X = i)P(Y = j; X = i)$$

. We deduce that if $j > i$, then $P((X, Y) = (i, j)) = 0$ whereas if $j \leq i$,


$$P((X, Y) = (i, j)) = \frac{1}{n} \frac{1}{i}$$

(b). Y takes its values in $\{1, \dots, n\}$ and we have

$$P(Y = k) = \sum_{i=1}^n P((X = i, \mathbb{N}; Y = k)) = \sum_{i=k}^n P((X = i, \mathbb{N}; Y = k)) = \sum_{i=k}^n \frac{1}{n i}$$

(c). To calculate the expectation, we apply the definition and permute the sums.

$$\begin{aligned} \mathbf{E}[Y] &= \frac{1}{n} \sum_{k=1}^n \sum_{i=k}^n \frac{k}{i} \\ &= \frac{1}{n} \sum_{k=1}^n \frac{1}{i} \sum_{i=k}^n k \\ &= \frac{1}{n} \sum_{k=1}^n \frac{1}{i} \frac{i(i+1)}{2} \\ &= \frac{1}{n} \sum_{k=1}^n \frac{i+1}{2} \\ &= \frac{1}{2n} \left(\frac{(n+1)(n+2)}{2} - 1 \right) \\ &= \frac{n+3}{4} \end{aligned}$$

 **Exercise3.4 [Uniform joint law]** Let (X, Y) be a pair of random variables following a uniform distribution on $\{0, \dots, n\}^2$.

(a). Determine the distribution of X , the distribution of Y and the law of $X + Y$.

(b). Are X and Y independent?

Solution

(a). X has values in $\{0, \dots, n\}$. For all k in $\{0, \dots, n\}$,

$$P(X = k) = \sum_{i=1}^n P((X, Y) = (k, i)) = \sum_{i=1}^n \frac{1}{(n+1)^2} = \frac{1}{n+1}$$

. So X follows a uniform distribution on $\{0, \dots, n\}$. By symmetry, the same applies to Y . The random variable $X + Y$ has values in $\{0, \dots, 2n\}$ and for all k in this interval of integers, we have

$$P(X + Y = k) = \sum_{i=1}^k P((X = i, Y = k - i)).$$

If $k \leq n$, this becomes equal to

$$P(X + Y = k) = \sum_{i=1}^k \frac{1}{(n+1)^2} = \frac{k+1}{(n+1)^2}.$$

If $k > n$, the sum can only go up to n (because $i \leq n$) and can only start in $k - n$ (because $k - i \leq n$) and therefore

$$P(X + Y = k) = \sum_{i=k-n}^n \frac{1}{(n+1)^2} = \frac{2n - k + 1}{(n+1)^2}.$$

(b). The variables X and Y are independent, because for any pair (i, j) of $\{0, \dots, n\}$, we have

$$P((X = i), (Y = j)) = P(X = i)P(Y = j) = \frac{1}{(n+1)^2}$$

 **Exercise3.5 [Random vectors and matrices]** Consider a probability space (Ω, \mathcal{B}, P) and two

random variables X and Y defined on Ω and having values in $\{1, \dots, n+1\}$, where n is a natural number greater than or equal to 2. For any pair $(i, j) \in \{1, \dots, n+1\}^2$, we posit

$$a_{i,j} = P(X = i, Y = j)$$

It is assumed that

$$a_{i,j} = \begin{cases} \frac{1}{2n} & \text{si } |i + j - (n + 2)| = 1 \\ 0 & \text{else.} \end{cases}$$

- Check that the family $(a_{i,j})$ defined in this way is a pairwise probability distribution.
- Write the matrix $A \in \mathcal{M}_{n+1}(\mathbb{R})$ whose general term is $a_{i,j}$. Check that A is diagonalizable.
- Determine the probability distributions of X and Y .
- For any pair $(i, j) \in \{1, \dots, n+1\}^2$, we pose

$$b_{i,j} = P(X = i \setminus Y = j)$$

- Determine the matrix $B \in \mathcal{M}_{n+1}(\mathbb{R})$ whose general term is $b_{i,j}$. Show that the vector

$$v = \begin{pmatrix} P(X = 1) \\ \vdots \\ P(X = n + 1) \end{pmatrix}$$

is an eigenvector of B .

Solution

- All scalars $(a_{i,j})$ are positive or zero. Let's count the number of non-zero $a_{i,j}$, which are non-zero. These are the terms for which:

Let $i + j = n + 3$, with $1 \leq i, j \leq n + 1$, or pairs of the form $(n + 3 - i)$ with $1 \leq i, n + 3 - i \leq n + 1$, which gives $2 \leq i \leq n + 1$. There are therefore n such terms.

Let $i + j = n + 1$, with $1 \leq i, j \leq n + 1$, or pairs of the form $(n + 1 - i)$ with $1 \leq i, n + 1 - i \leq n + 1$, which gives $1 \leq i \leq n + 1$. There are therefore n such terms. In

addition, there are $2n$ terms, and as each is worth $\frac{1}{2n}$, the sum is 1.

The matrix of A has the form:

$$\begin{pmatrix} 0 & \dots & \dots & \dots & \frac{1}{2n} & 0 \\ 0 & \dots & \dots & \frac{1}{2n} & 0 & \frac{1}{2n} \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & & & \\ \frac{1}{2n} & \ddots & \ddots & & & \vdots \\ 0 & \frac{1}{2n} & 0 & \dots & & \end{pmatrix}$$

(the $\frac{1}{2n}$ terms are above and below the non-principal diagonal). It is real symmetric,

and therefore diagonalizable.

The distribution of X is given by

$$P(X = i) = \sum_{j=1}^{n+1} P(X = i, Y = j) = \sum_{j=1}^{n+1} a_{i,j}.$$

We find:

- $P(X = 1) = P(X = n + 1) = \frac{1}{2n}$.
- $P(X = i) = \frac{1}{n}$ si $2 \leq i \leq n$.

Mathematically, this means that we have summed up each line of A . By symmetry, Y follows the same law as X .

(b). For all $1 \leq i, j \leq n + 1$,

$$b_{i,j} = \frac{a_{i,j}}{P(Y = j)}.$$

- $b_{1,j} \neq 0$ if and only if $j = n$ and in this case $b_{1,n} = 1/2$.
- $b_{n+1,j} \neq 0$ if and only if $j = 2$ and in this case $b_{n+1,2} = 1/2$.
- for $2 \leq i \leq n$, $b_{i,j} \neq 0$ if and only if $j = n + 3 - i$ ou $j = n + 1 - i$ the

$$\begin{cases} b_{2,n+1} = 1 & b_{i,n+3-i} = 1/2 \\ b_{n,1} = 1/2 & b_{i,n+1-i} = 1/2 \end{cases}$$

It is then easy to check that v is an eigenvector for eigenvalue 1.

🔥 **Exercise 3.6 [Geometric couple]** Let X and Y be two random variables with values in \mathbb{N}^* , such that:

$$P((X = i) \cap (Y = j)) = \frac{a}{2^{i+j}},$$

for all $i, j \in \mathbb{N}^*$.

- (a). Calculate a .
- (b). Determine the marginal laws of X and Y .
- (c). Are X and Y independent?

Solution

(a). It is first necessary that $a \geq 0$. Then we have

$$\sum_{i,j \geq 1} \frac{a}{2^{i+j}} = 1 \iff a \sum_{i=1}^{+\infty} \frac{1}{2^i} \sum_{j=1}^{+\infty} \frac{1}{2^j} = 1 \iff a = 1$$

(b). For $i \in \mathbb{N}^*$, we have

$$P(X = i) = \sum_{j=1}^{+\infty} P((X = i) \cap (Y = j)) = \sum_{j=1}^{+\infty} \frac{1}{2^{i+j}} = \frac{1}{2^i} = \frac{1}{2} \frac{1}{2^{i-1}}$$

. X therefore follows a geometric distribution with parameter $1/2$. By symmetry, the same applies to Y .

(c). We have

$$P(X = i) \times P(Y = j) = \frac{1}{2^{i+j}} = P((X = i) \cap (Y = j))$$

The random variables are independent. The random variables are independent.

Exercise 3.7 [Geometric couple] Let X and Y be two independent random variables following the same geometric distribution with parameter $p \in]0, 1[$. Let $Z = \min(X, Y)$ and $q = 1 - p$. Let n also be a strictly positive integer.

- Calculate $P(X \geq n)$.
- Calculate $P(Z \geq n)$. Deduce $P(Z = n)$. What is the distribution of Z ?
- Are the variables X and Z independent?

Solution

(a). The event $X \geq n$ is written as a (countable and disjoint) union of elementary events $X = k$, $k \geq n$. We therefore have

$$P(X \geq n) = \sum_{k \geq n} P(X = k) = \sum_{k \geq n} q^{k-1} p = \frac{q^{n-1} p}{1 - q} = q^{n-1}.$$

(b). We have $Z \geq n$ if and only if $X \geq n$ and $Y \geq n$. These two events are independent, and therefore

$$P(Z \geq n) = P(X \geq n)P(Y \geq n) = q^{2n-2}.$$

Now,

$$P(Z = n) = P(Z \geq n) - P(Z \geq n + 1) = q^{2n-2} - q^{2n} = q^{2n-2}(1 - q^2).$$

Thus, Z follows a geometric distribution with parameter $1 - q^2$.

(c). Note that the events $(X = n) \cap (Z \geq n)$ and $(X = n) \cap (Z = n)$ are equal and equal to $(X = n) \cap (Y \geq n)$. If X and Z were independent, then we would have

$$P((X = n) \cap (Z \geq n)) = P(X = n)P(Z \geq n)$$

and

$$P((X = n) \cap (Z = n)) = P(X = n)P(Z = n)$$

In particular, we should have $P(Z = n) = P(Z \geq n)$, which is not the case. X and Z are not independent random variables.

Exercise 3.8 Post Office Counter There are two counters in a post office. Each person arriving at the post office chooses the first counter with probability p , or the second counter with probability $q = 1 - p$. The people make their choices independently. In one hour, the number of people arriving at the post office follows a Poisson distribution $\mathcal{P}(m)$. We denote by Y the number of people who chose the first counter.

- Express the conditional probability of $Y = k$ knowing that $X = n$.
- Deduce the joint distribution of (X, Y) .
- Determine the distribution of Y . We will find that Y follows a Poisson distribution with parameter mp .

Solution

(a). For each person, the choice of the first counter is made with probability p . The choices are

independent of each other, and $Y = k \mid X = n$ counts the number of "successes" when the test is performed n times. We can see the theoretical pattern of a binomial random variable. We therefore have:


$$P(Y = k \mid X = n) = \begin{cases} C_n^k p^k q^{n-k} & \text{si } 0 \leq k \leq n \\ 0 & \text{si } k > n. \end{cases}$$

(b). On a

$$P(Y = k, X = n) = P(X = n)P(Y = k \mid X = n) = \begin{cases} e^{-m} \frac{m^n}{n!} C_n^k p^k q^{n-k} & \text{if } k \leq n \\ 0 & \text{else.} \end{cases}$$

We need to do the summation! Taking into account that the first terms are zero, we have:

$$\begin{aligned} P(X = k) &= \sum_{n=k}^{+\infty} P(Y = k, X = n) \\ &= e^{-m} \left(\frac{p}{q}\right)^k \frac{1}{k!} \sum_{n=k}^{+\infty} \frac{(mq)^n}{(n-k)!} \\ &= e^{-m} \left(\frac{p}{q}\right)^k \frac{1}{k!} (mq)^k \sum_{n=k}^{+\infty} \frac{(mq)^{n-k}}{(n-k)!} \\ &= e^{-m} \left(\frac{p}{q}\right)^k \frac{1}{k!} (mq)^k \sum_{n=0}^{+\infty} \frac{(mq)^n}{n!} \\ &= e^{-m} \left(\frac{p}{q}\right)^k \frac{1}{k!} (mq)^k e^{mq} \\ &= e^{-mp} \frac{(mp)^k}{k!} \end{aligned}$$

 **Exercise 3.9 [Births]** It is assumed that the number N of children in a family follows a Poisson distribution with parameter $\lambda > 0$. It is assumed that at each birth, the probability that the child is a girl is $p \in]0, 1[$ and the probability that it is a boy is $q = 1 - p$. It is also assumed that the sexes of successive births are independent. Let X be the random variable corresponding to the number of girls per family, and Y that of the number of boys.

(a). Determine the joint distribution of (N, X) .

(b). Deduce the distribution of X and that of Y .

Solution

(a). The probability $P(X = k; \mid; N = n)$ is equal $C_n^k p^k q^{n-k}$ if $k \leq n$ (with a fixed number of births, the number of girls follows a binomial distribution with parameters n and p) and 0 if $k > n$ (or $k < 0$). We deduce that

$$P(N = n, X = k) = P(X = k \mid N = n)P(N = n) = C_n^k p^k q^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}.$$

(b). We deduce the marginal law of X from the law of the couple:

$$\begin{aligned}
 P(X = k) &= \sum_{n \geq k} P(X = k, N = n) \\
 &= \sum_{n \geq k} \frac{e^{-\lambda} \lambda^n p^k q^{n-k}}{k!(n-k)!} \\
 &= \frac{e^{-\lambda} \lambda^k p^k}{k!} \sum_{n=k}^{+\infty} \frac{\lambda^{n-k} q^{n-k}}{(n-k)!} \\
 &= \frac{e^{-\lambda} \lambda^k p^k}{k!} \sum_{m=0}^{+\infty} \frac{\lambda^m q^m}{m!} \\
 &= \frac{e^{-\lambda} \lambda^k p^k}{k!} e^{\lambda q} \\
 &= \frac{e^{-\lambda p} (\lambda p)^k}{k!}
 \end{aligned}$$

Thus, X follows a Poisson distribution with parameter λp . Similarly, Y follows a Poisson distribution with parameter λq .

🔥 **Exercise 3.10 [Archery]** Theo is archery a circular target of radius 1. It is assumed that Theo is sufficiently clumsy for the point of impact M of coordinates (X, Y) is uniformly distributed over the target. Let $D = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$.

- (a). What is the density of the pair (X, Y) ?
 (b). Determine the marginal laws of X and Y .
 (c). Are the random variables X and Y independent?

Solution

(a). According the statement, we have

$$p_{(X,Y)}(x, y) = \frac{1}{\pi} \mathbf{1}_D(x, y)$$

the constant $\frac{1}{\pi}$ used to normalise this density (its integral is 1).

(b). Apply the formula from the course:

$$p_X(x) = \int_{\mathbb{R}} p_{(X,Y)}(x, y) dy = \frac{2}{\pi} \sqrt{1-x^2} \mathbf{1}_{]-1,1[}(x).$$

By symmetry of the role played by X and Y , we also have

$$p_Y(y) = \int_{\mathbb{R}} p_{(X,Y)}(x, y) dx = \frac{2}{\pi} \sqrt{1-y^2} \mathbf{1}_{]-1,1[}(y).$$

(c). Let $I = J = [3/4, 1]$. It is clear that $P(X \in I) > 0$ and $P(Y \in J) > 0$. On the other hand, since $(3/4)^2 + (3/4)^2 > 1$, we also have

$$P(X \in I, Y \in J) = 0$$

So

$$P(X \in I, Y \in J) \neq P(X \in I)P(Y \in J)$$

and so the random variables X and Y are not independent.

🔥 **Exercise 3.11 [Triangle]** Let T be the interior of a triangle in the plane bounded by the points

$O(0, 0)$, $I(1, 0)$ et $J(0, 1)$ and let (X, Y) be a pair of random variables with uniform distribution over the triangle T .

- Give the density of the pair (X, Y) .
- Calculate the marginal laws of X and Y .
- Are the random variables X and Y independent?
- Calculate the covariance of the pair (X, Y) . What is your opinion?

Solution

- Since (X, Y) follows a uniform distribution on the triangle T whose area is $1/2$, the density of the pair (X, Y) is given by

$$p_{(X,Y)}(x, y) = 2 \times \mathbf{I}_T(x, y).$$

- We find the marginal density by applying the formula from the course (by integration). Note that X has values in $[0, 1]$, and therefore $p_X(x) = 0$ if $x \notin [0, 1]$. If $x \in [0, 1]$, we deduce

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{+\infty} p_{X,Y}(x, y) dy \\ &= \int_0^{1-x} 2 dy \\ &= 2(1-x). \end{aligned}$$

We can therefore deduce that $p_X(x) = 2(1-x) \mathbf{I}_{[0,1]}(x)$. By symmetry, the density of Y , p_Y , is equal to that of X , p_X .

- Let $I = J = [1/2, +\infty[$. It is clear that $P(X \in I) > 0$ and $P(Y \in J) > 0$. On the other hand, the set $\{(x, y) \in \mathbb{R}^2, x > 1/2 \text{ et } y > 1/2\}$ is disjoint from T , and therefore

$$P(X \in I, Y \in J) = 0 \neq P(X \in I)P(Y \in J).$$

The two random variables are not independent.

- On the one hand we have

$$E(X) = E(Y) = \int_{\mathbb{R}} x; p_X(x) dx = \int_0^1 2x(1-x) dx = \frac{1}{3}$$


. On the other hand,

$$\begin{aligned} \mathbf{E}(XY) &= \int_{\mathbb{R}^2} xy p_{(X,Y)}(x, y) dx dy \\ &= \int_{x=0}^1 \int_{y=0}^{1-x} x(1-x)^2 dx \\ &= \frac{1}{12} \end{aligned}$$

We deduce that

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{12} - \frac{1}{9} = -\frac{1}{36} \neq 0$$

The covariance is non-zero, so we have the result from the previous question telling us that the random variables are not independent.

 **Exercise 3.12 [Exponential]** Let X and Y be two independent random variables following

exponential laws with respective parameters λ and μ . Determine $P(X > Y)$.

Solution Since the variables (X, Y) are independent, the density of the pair (X, Y) is

$$f(x, y) = \mathbf{1}_{]0, +\infty[^2}(x, y) \lambda e^{-\lambda x}; \mu; e^{-\mu y}.$$

We deduce that

$$\begin{aligned} P(X > Y) &= E(\mathbf{I}_{x>y}) \\ &= \int_{y=0}^{+\infty} \int_{x=y}^{+\infty} \lambda e^{-\lambda x} dx \mu e^{-\mu y} dy \\ &= \int_{y=0}^{+\infty} \mu e^{-(\lambda+\mu)y} dy \\ &= \frac{\mu}{\lambda + \mu} \end{aligned}$$

In particular, if $\lambda = \mu$, we get $1/2$.

Exercise 3.13 [Product of Pareto's laws] The random variable X is said to follow a Pareto distribution with parameter $\alpha > 0$ if,

$$\forall x \geq 1, P(X > x) = x^{-\alpha}.$$

- (a). Show that this property effectively characterises the distribution of X . Show that X follows a density distribution, and specify this density. What values of X is its expectation finite?
- (b). Let X, Y be two independent random variables following a Pareto distribution with parameter α . Let dP_Y be the distribution of Y . Show that if $t \geq 1$, then

$$P(XY > t) = \int_1^{+\infty} P\left(X > \frac{t}{y}\right) dP_Y(y).$$

- (c). Deduce that, for all $t \geq 1$, $P(XY > t) = t^{-\alpha} (1 + \alpha \ln t)$.

Solution

- (a). We note that $P(X > 1) = 1$, et donc pour $t \leq 1$, on a $P(X > t) = 1$. The given formula can therefore be used to fully characterise the distribution function F_X , which has the value

$$F_X(t) = \mathbf{I}_{[1, +\infty[}(t)(1 - t^{-\alpha}).$$

- (b). This function is derivable, except possibly at 1. Its derivative is the density

$$f_X(t) = \mathbf{I}_{[1, +\infty[}(t) \alpha t^{-\alpha-1}.$$

- (c). The function $t \times t^{-\alpha-1}$ is integrable in the neighbourhood of $+\infty$ if and only if $-\alpha - 1 < -1$, c'est-à-dire $\alpha > 1$.
- (d). Since (X, Y) are independent, the law of the pair (X, Y) is equal to the (tensor) product of the laws. In other words, by the transfer theorem:

$$P(XY > t) = \int_{\mathbb{R}^2} \mathbf{I}_{xy>t}(x, y) dP_{(X,Y)}(x, y) = \int_{\mathbb{R}^2} \mathbf{I}_{xy>t}(x, y) dP_X(x) dP_Y(y).$$

We can integrate over $[1, +\infty[^2$ instead of \mathbb{R}^2 , then use Fubini's theorem to obtain

$$P(XY > t) = \int_1^{+\infty} \left(\int_1^{+\infty} \mathbf{I}_{xy>t}(x, y) dP_X(x) \right) dP_Y(y).$$

But

$$\int_1^{+\infty} \mathbf{I}_{xy > t}(x, y) dP_X(x) = \int_1^{+\infty} \mathbf{I}_{x \geq t/y}(x) dP_X(x) = P\left(X > \frac{t}{y}\right).$$

(e). Using the previous result, and noting that $P(X > t/y) = 1$ if $y \geq t$, we have

$$P(XY > t) = \int_1^t t^{-\alpha} y^{\alpha} \alpha y^{-\alpha-1} dy + \int_t^{+\infty} \alpha y^{-\alpha-1} dy.$$

By integrating, we get

$$P(XY > t) = \alpha t^{-\alpha} \int_1^t y^{-1} dy + t^{-\alpha} = \alpha t^{-\alpha} \ln(t) + t^{-\alpha} = t^{-\alpha}(1 + \alpha \ln t).$$

🔥 **Exercise 3.14** A student is bored during his probability course and spends his time looking out of the window at the leaves falling from a tree. We assume that the number of leaves falling at the end of the course is a random variable X that follows a Poisson distribution with parameter $\lambda > 0$. This means that for all $k \in \mathbb{N}$,

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

(a). I. Explain why the assumptions in the statement allow us to say that for any $\lambda > 0$,

$$e^{\lambda} = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}.$$

II. Calculate the expectation and variance of X .

(b). Each time a leaf falls to the ground, the student throws a coin which gives tails with probability p and heads with probability $q = 1 - p$, $p \in]0, 1[$. The numbers of heads and tails obtained are F and P respectively.

I. For a fixed $k \in \mathbb{N}$, explain in a simple way why the distribution of F knowing $X = k$ is a binomial distribution, the parameters of which will be specified. Deduce the expression of $P(F = a; X = k)$. For $(k, a) \in \mathbb{N}$, calculate the quantity $P(X = k, F = a)$.

III. From this, deduce the distribution of F , as well as its expectation.

IV. Give, without calculation, the distribution of P .

V. Show that P and F are independent.

VI. Calculate $E[PF]$ and $\text{Var}[P + F]$.

Solution

(a). I. We are told that X is a random variable with values in \mathbb{N} such that

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

We also know that

$$\sum_{k=0}^{+\infty} P(X = k) = 1.$$

From these two relationships, we can easily deduce the required result.

$$e^{\lambda} = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}.$$

II. These are arguments to know. They are

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{k=0}^{+\infty} k P(X = k) \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda.
\end{aligned}$$

In the same way, if we write $k^2 = k(k-1) + k$

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{+\infty} k^2 P(x = k) \\
&= e^{-\lambda} \sum_{k=2}^{+\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\
&= \lambda^2 + \lambda
\end{aligned}$$

We deduce that

$$V(X) = E(X^2) - (E(X))^2 = \lambda.$$

(b). I. When X is set equal to k , F counts the number of successes (obtaining heads) of the repetition of k independent random experiments, each with a probability q of success. So the distribution of F knowing $X = k$ is a binomial distribution $\mathcal{B}(k, q)$. We deduce that

$$P(F = a \mid X = k) = \begin{cases} C_a^k q^a p^{k-a} & \text{et } 0 \leq a \leq k \\ 0 & \text{sinon.} \end{cases}$$

II. We have

$$\begin{aligned}
P(F = a, X = k) &= P(X = k)P(F = a \mid X = k) \\
&= \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!} C_a^k q^a p^{k-a} & \text{and } a \leq k \\ 0 & \text{else.} \end{cases}
\end{aligned}$$

III. According to the total probability formula,

$$\begin{aligned}
 P(F = a) &= \sum_{k=a}^{+\infty} P(F = a, X = k) \\
 &= e^{-\lambda} \left(\frac{q}{p}\right)^a \frac{1}{a!} \sum_{k=a}^{+\infty} \frac{(\lambda p)^k}{(k-a)!} \\
 &= e^{-\lambda} \left(\frac{q}{p}\right)^a \frac{1}{a!} (\lambda p)^a \sum_{k=a}^{+\infty} \frac{(\lambda p)^{k-a}}{(k-a)!} \\
 &= e^{-\lambda} \left(\frac{q}{p}\right)^a \frac{1}{a!} (\lambda p)^a \sum_{k=0}^{+\infty} \frac{(\lambda p)^k}{k!} \\
 &= e^{-\lambda} \left(\frac{q}{p}\right)^a \frac{1}{a!} (\lambda p)^a e^{\lambda p} \\
 &= e^{-\lambda q} \frac{(\lambda q)^a}{a!}.
 \end{aligned}$$

Thus, F follows a Poisson distribution with parameter $q\lambda$.

IV. By symmetry, P follows a Poisson distribution with parameter $p\lambda$.

(c). The event $(F = a) \cap (P = b)$ is the event $(F = a) \cap (X = a + b)$ and therefore

$$\begin{aligned}
 P(F = a \text{ et } P = b) &= P(F = a \text{ et } X = a + b) \\
 &= P(F = a \mid X = a + b)P(X = a + b) \\
 &= e^{-\lambda} \frac{\lambda^{a+b}}{(a+b)!} \binom{a+b}{a} q^a p^b \\
 &= e^{-\lambda} \frac{\lambda^{a+b}}{a!b!} q^a p^b
 \end{aligned}$$

On the other hand

$$P(F = a) \times P(P = b) = e^{-\lambda q} \frac{(\lambda q)^a}{a!} e^{-\lambda p} \frac{(\lambda p)^b}{b!}.$$

These two quantities are equal, so P and F are independent.

(d). By independence of P and F , we have

$$E(PF) = E(P)E(F) = \lambda^2 pq$$

and

$$V(P + F) = \text{Var}(P)\text{Var}(F) = \lambda,$$

which could also be found by simply noting that $P + F = X$.

Chapter Convergence of random variables

Introduction

- Weak laws of large numbers
- Almost sure convergence
- Convergence in probability
- Convergence in \mathbb{L}^r

4.1 Introduction

Probabilities entrain many notions of convergence. The simplest of these is the pointwise convergence of a sequence of random variables X_n to a limit X . So, in probability theory, one uses various modes of convergence of random variables, many of which are crucial for applications. This part of probability is often called « large sample theory », « limit theory » or « asymptotic theory ». This material is extremely important for statistical inference. The basic question is this: what can we say about the limiting behavior of a sequence of random variables X_1, X_2, X_3, \dots ? Since statistics is all about gathering data, we will naturally be interested in what happens as we gather more and more data, hence our interest in this question.

In this chapter we shall consider some of the most important of them: Convergence in Probability, Convergence in \mathbb{L}^r , convergence in Distribution and convergence with probability one (almost sure convergence).

Note that the definitions are stated in terms of scalar random variables, but extend naturally to vector random variables.

4.2 Convergence in Probability

We first define convergence in probability and the related order concepts which, as we shall see, are closely analogous to their deterministic counter-parts. With these tools we can then develop convergence in probability analogues of Taylor expansions which will be used later to drive the large-sample asymptotic distributions.

Let $\{a_n\}$ be a sequence of strictly positive real numbers and let $\{X_n; n \geq 1\}$ be a sequence of random variables all defined on the same probability space.

Definition 4.1. Convergence in Probability to Zero

We say that X_n converges in probability to zero, written $X_n = o_p(1)$ or $X_n \xrightarrow{p} 0$ if for every $\epsilon > 0$

$$P(|X_n| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$



Remark The definition of convergence in probability says nothing about the convergence of random variable X_n to X in the sense in which it is understood in real analysis. Thus, $X_n \xrightarrow{p} X$ does not imply that, given $\epsilon > 0$, we can find an N such that $|X_n - X| < \epsilon$, for $n \geq N$. The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the parameter being estimated. Convergence in probability is also the type of convergence established by the weak law of large numbers.

Definition 4.2. Bounded in Probability

We say that the sequence $\{X_n\}$ is bounded in probability (or tight), $X_n = O_p(1)$, if for every $\epsilon > 0$ there exists $\delta(\epsilon) \in [0, \infty[$ such that

$$P(|X_n| > \delta(\epsilon)) < \epsilon \quad \text{for all } n$$



The relation between these two concepts is clarified by the following equivalent characterization of convergence in probability to zero, viz. $X_n = o_p(1)$ if and only if for every $\epsilon > 0$ there exists a sequence $\delta(\epsilon) \downarrow 0$ such that

$$P(|X_n| > \delta(\epsilon)) < \epsilon \quad \text{for all } n$$

The definition should also be compared with their non-random counterparts, viz. $x_n = o(1)$ if $x_n \rightarrow 0$ and $x_n = O(1)$ if $\{x_n\}$ is bounded.

Definition 4.3. Convergence in probability and Order in Probability

- (a). X_n converges in probability to the random variable X , written $X_n \xrightarrow{p} X$, if and only if $X_n - X = o_p(1)$.
- (b). $X_n = o_p(a_n)$ if and only if $a_n^{-1}X_n = o_p(1)$.
- (c). $X_n = O_p(a_n)$ if and only if $a_n^{-1}X_n = O_p(1)$.



Notice that if we drop the subscripts p in Definition 4.3 we have the usual definition of $o(\cdot)$ and $O(\cdot)$ for non-random sequences. In fact, most of the rules governing the manipulation of $o(\cdot)$ and $O(\cdot)$ carry over to $o_p(\cdot)$ and $O_p(\cdot)$. In particular we have the following results.

Proposition 4.1

- (a). If $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$, we have

$$\begin{aligned} X_n Y_n &= o_p(a_n b_n), \\ X_n + Y_n &= o_p(\max(a_n, b_n)), \end{aligned}$$

and

$$|X_n|^r = o_p(a_n^r) \quad \text{for } r > 0.$$

(b). If $X_n = o_p(a_n)$ and $Y_n = O_p(b_n)$, we have

$$X_n Y_n = o_p(a_n b_n).$$

Moreover

(c). The statements (1) remains valid if o_p is everywhere replaced by O_p .

Proof

(a). If $|X_n Y_n|/(a_n b_n) > \epsilon$ then either $|Y_n|/b_n \leq 1$ and $|X_n|/a_n > \epsilon$ or $|Y_n|/b_n > 1$ and $|X_n Y_n|/(a_n b_n) > \epsilon$. Hence

$$\begin{aligned} P(|X_n Y_n|/(a_n b_n) > \epsilon) &\leq P(|X_n|/a_n > \epsilon) + P(|Y_n|/b_n > 1) \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

If $|X_n + Y_n|/\max(a_n, b_n) > \epsilon$ then either $|X_n|/a_n > \epsilon/2$ or $|Y_n|/b_n > \epsilon/2$.

Hence,

$$\begin{aligned} P(|X_n + Y_n|/\max(a_n, b_n) > \epsilon) &\leq P(|X_n|/a_n > \epsilon/2) + P(|Y_n|/b_n > \epsilon/2) \\ &\rightarrow 0 \quad n \rightarrow \infty. \end{aligned}$$

For the last part of (1) we simply observe that

$$P(|X_n|^r/a_n^r > \epsilon) = P(|X_n|/a_n > \epsilon^{1/r}) \rightarrow 0 \quad n \rightarrow \infty.$$

(b). Parts (2) and (3) are left as exercise for readers. □

The definitions 4.1-4.3 extend in a natural way to sequence of random vectors. Suppose now that $\{\mathbf{X}_n, n = 1, 2, \dots\}$ is a sequence of random vectors, all defined on the same probability space and such that X_n has k components $X_{n1}, \dots, X_{nk}, n = 1, 2, \dots$

Definition 4.4. Order in Probability for Random Vectors

- (a). $\mathbf{X}_n = o_p(a_n)$ if and only if $X_{nj} = o_p(a_n), j = 1, 2, \dots, k$.
- (b). $\mathbf{X}_n = O_p(a_n)$ if and only if $X_{nj} = O_p(a_n), j = 1, 2, \dots, k$.
- (c). \mathbf{X}_n converge in probability to the random vector \mathbf{X} , written $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, if and only if $\mathbf{X}_n - \mathbf{X} = o_p(1)$.

Convergence in probability of \mathbf{X}_n to \mathbf{X} can also be conveniently characterized in terms

of the Euclidian distance $|\mathbf{X}_n - \mathbf{X}| = \left[\sum_{j=1}^k (X_{nj} - X_j)^2 \right]^{1/2}$.

Proposition 4.2

$\mathbf{X}_n - \mathbf{X} = o_p(1)$ if and only if $|X_n - X_j| = o_p(1)$.

Proof If $\mathbf{X}_n - \mathbf{X} = o_p(1)$ then each $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|X_{nj} - X_j|^2 > \epsilon/k) = 0$ for each $j = 1, \dots, k$. But

$$P\left(\sum_{j=1}^k |X_{nj} - X_j|^2 > \epsilon\right) \leq \sum_{j=1}^k P(|X_{nj} - X_j|^2 > \epsilon/k) \quad (4.1)$$

Since $\sum_{j=1}^k |X_{nj} - X_j|^2 > \epsilon$ implies that at least one summand exceeds ϵ/k . Since the right side of (4.1) converges to zero so too does the left side and hence $|\mathbf{X}_n - \mathbf{X}|^2 = o_p(1)$. By proposition 4.1 this implies that $|\mathbf{X}_n - \mathbf{X}| = o_p(1)$.

Conversely if $|\mathbf{X}_n - \mathbf{X}| = o_p(1)$ we have $|X_{ni} - X_i|^2 \leq |\mathbf{X}_n - \mathbf{X}|^2 = o_p(1)$ whence $P(|X_{ni} - X_i| > \epsilon) \leq P(|\mathbf{X}_n - \mathbf{X}|^2 > \epsilon^2) \rightarrow 0$. \square

4.3 Convergence in r^{th} mean, $r > 0$

Definition 4.5. Convergence in r^{th} Mean, $r > 0$

Let r be fixed. The sequence of random variables $\{X_n\}$ is said to converge in r^{th} mean to X , written $X_n \xrightarrow{r} X$ if

$$\mathbb{E}|X_n - X|^r \rightarrow 0$$

as $n \rightarrow \infty$

Example 4.1 Let $\{X_n\}_{n \geq 1}$ be a sequence of random variables such that for some real numbers $(a_n)_{n \geq 1}$, we have

$$P(X_n = a_n) = p_n, \quad P(X_n = 0) = 1 - p_n.$$

Then $X_n \xrightarrow{L^r} 0$ iff $\mathbb{E}[|X_n|^r] \equiv |a_n|^r p_n \rightarrow 0$ as $n \rightarrow \infty$.

Example 4.2 Let $\{X_n\}$ be a sequence of random variables such that,

$$\begin{aligned} P[X_n = 1] &= \frac{1}{n} \quad \text{and} \\ P[X_n = 0] &= 1 - \frac{1}{n} \end{aligned}$$

Then show that X_n converges in mean square to X , where X is a random variable degenerate at zero. Also, show that X_n converges to in probability

Here we have,

$$\begin{aligned} E[X_n^2] &= 1^2 \times P[X_n = 1] + 0^2 \times P[X_n = 0] \\ &= 1 \times \frac{1}{n} = \frac{1}{n}. \end{aligned}$$

So,

$$\begin{aligned} E[X_n^2] &\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \\ E[|X_n^2 - 0|] &\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \\ &\rightarrow X_n^2 \xrightarrow{2} 0 \Rightarrow X_n \xrightarrow{2} X \quad \text{where, } X = 0 \end{aligned}$$

Thus, the random variable X is degenerate at 0. Now,

$$\begin{aligned} P[X_n = 1] &= \frac{1}{n} \quad \text{and} \\ P[X_n = 0] &= 1 - \frac{1}{n} \quad \text{implies} \\ P[|X_n| > 0] &= \frac{1}{n} \\ \Rightarrow P[|X_n| > \epsilon] &= \frac{1}{n} \quad \text{if } 0 < \epsilon < 1 \\ &= 0 \quad \text{if } \epsilon \geq 1 \end{aligned}$$

Thus,

$$\begin{aligned} \lim_{n \rightarrow \infty} P[|X_n| > \epsilon] &= 0 \\ \Rightarrow P[|X_n - 0| > \epsilon] &= 0 \\ \Rightarrow X_n &\xrightarrow{P} 0. \end{aligned}$$

The following result is the L^2 weak law of large numbers (L^2 – WLLN)

Theorem 4.1

Let $X_j, j \geq 1$, be a sequence of uncorrelated random variables with $\mathbb{E}(X_j) = \mu$ and

$\text{Var}(X_j) \leq C < \infty$. Denote $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{1}{n} S_n \xrightarrow{L^2} \mu \quad \text{as } n \rightarrow \infty$$

Proposition 4.3. Chebychev's inequality

If $\mathbb{E}[|X|^r] < \infty, r \geq 0$ and $\epsilon > 0$, then

$$P[|X| \geq \epsilon] \leq \epsilon^{-r} \mathbf{E}|X|^r.$$

Proof

$$\begin{aligned} P(|X| \geq \epsilon) &= P(\epsilon^{-r} \mathbf{E}[|X|^r] \epsilon^{-r} \geq 1) \\ &= \mathbf{E}[\mathbb{I}_{1,+\infty}(|X|^r \epsilon^{-r})], \quad \text{where } \mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \\ &= \mathbf{E}[|X|^r \epsilon^{-r} \mathbb{I}_{[1, \infty)}(|X|^r \epsilon^{-r})] \\ &\leq \epsilon^{-r} \mathbf{E}[|X|^r]. \end{aligned} \quad \square$$

The following three propositions provide useful connections between the behaviour of moments and order in probability.

Proposition 4.4

If $X_n \xrightarrow{r} X$ then $X_n \xrightarrow{P} X$



Proof By Chebechev's inequality we have for any $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \leq \epsilon^{-r} \mathbf{E}[|X_n - X|^r] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

□

Proposition 4.5

If $a_n > 0$, $n = 1, 2, \dots$, and $\mathbf{E}(X_n^2) = O(a_n^2)$, then $X_n = O_p(a_n)$.



Proof Applying Chebychev's inequality again, we have for any $M > 0$,

$$\begin{aligned} P(a_n^{-1}|X_n| > M) &\leq a_n^{-2} \mathbf{E}[|X_n|^2] / M^2 \\ &\leq C / M^2 \quad \text{where } C = \sup_n (a_n^{-2} \mathbf{E}[|X_n|^2]) < \infty. \end{aligned}$$

Defining $\delta(\epsilon) = 2(C/\epsilon)^{1/2}$ if $C > 0$ and any positive constant if $C = 0$, we see from definition 4.2 that $a_n^{-1}|X_n| = O_p(1)$.

□

Proposition 4.6

If $\mathbf{E}[X_n] \rightarrow \mu$ and $\text{Var}(X_n) \rightarrow 0$ then $X_n \xrightarrow{m.s} \mu$ (and $X_n \xrightarrow{P} \mu$ by proposition 4.4)



Proof

$$\begin{aligned} \mathbf{E}[|X_n - \mu|^2] &= \mathbf{E}[X_n - \mathbf{E}[X_n]]^2 + \mathbf{E}[|X_n - \mathbf{E}[X_n]|^2] \\ &\rightarrow 0 \quad n \rightarrow \infty. \end{aligned}$$

□

4.4 Convergence in Distribution

Convergence in distribution is a very different matter. Indeed, the statements $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{m.s} X$ are meaningful only when the random variables X, X_1, X_2, \dots are all defined on the same probability space. The notions of convergence in distribution however depends only on the distribution functions of X, X_1, X_2, \dots and is meaningful avec if X, X_1, X_2, \dots are defined on different probability spaces. We shall show in Proposition 4.5 that convergence in distribution of a sequence $\{X_n\}$ is implied by convergence in probability. We begin with this definition.

Definition 4.6. Convergence in Distribution


The sequence $\{X_n\}$ of random k -vectors variables with distribution functions $\{F_{X_n}(\cdot)\}$ is said to converge in distribution if there exists a random k vector \mathbf{X} such

$$\lim_{n \rightarrow \infty} F_{\mathbf{X}_n}(x) = F_{\mathbf{X}}(x), \quad \text{for all } x \in \mathbf{C} \quad (4.2)$$

where \mathbf{C} is the set of continuity point of the distribution function $F_{\mathbf{X}}(\cdot)$ of \mathbf{X} .

If 4.2 holds we shall say that \mathbf{X}_n converges in distribution to \mathbf{X} . Such convergence will be denoted by $\mathbf{X}_n \Rightarrow \mathbf{X}$ or $F_{\mathbf{X}_n} \Rightarrow F_{\mathbf{X}}$.



 **Note** If $\mathbf{X}_n \Rightarrow \mathbf{X}$ then the distribution of \mathbf{X}_n can be well approximated for large n by the distribution of X . This observation is extremely useful since $F_{\mathbf{X}}$ is often easier to compute than $F_{\mathbf{X}_n}$ s.

4.5 almost sure convergence

Almost sure convergence - or convergence with probability one - is the probabilistic version of point wise convergence known from elementary real analysis.

Definition 4.7

The sequence of random variables $\{X_1, X_2, \dots\}$ is said to converge almost surely to a random variable X if

$$P\left(\left\{w : \lim_{n \rightarrow \infty} X_n(w) = X(w)\right\}\right) = 1$$

In this case we write $X_n \xrightarrow{a.s.} X$ (or $X_n \rightarrow X$ with probability 1).



4.6 Relations between convergence modes

Sometimes, we are led to determine a mode of convergence of a series of random variables and in order to do so it is essential to know the hierarchy of convergence. Indeed, there are several relations among the various modes of convergence, which are summarized in what follows (an arrow denotes implication in the arrow's direction):

Proposition 4.7

- If a sequence of random variables $\{X_n\}$ converges almost surely to a random variable X , then $\{X_n\}$ also converges in probability to X .
- If a sequence of random variables c converges in probability to a random variable X , then $\{X_n\}$ also converges in distribution to X .
- If a sequence of random variables $\{X_n\}$ converges almost surely to a random variable X , then $\{X_n\}$ also converges in distribution to X .
- If a sequence of random variables $\{X_n\}$ converges in mean square to a random variable X , then $\{X_n\}$ also converges in probability to X .



4.7 Exercises

Exercise 4.1 Let $\{X_n\}$ be an IID sequence of continuous random variables having a uniform distribution with support $R_{X_n} = \left[-\frac{1}{n}, \frac{1}{n}\right]$ and probability density function

$$f_{X_n}(x) = \begin{cases} \frac{n}{2} & \text{if } x \in \left[-\frac{1}{n}, \frac{1}{n}\right] \\ 0 & \text{if } x \notin \left[-\frac{1}{n}, \frac{1}{n}\right] \end{cases}$$

Find the probability limit (if it exists) of the sequence $\{X_n\}$.

Solution As n tends to infinity, the probability density tends to become concentrated around the point $x = 0$. Therefore, it seems reasonable to conjecture that the sequence $\{X_n\}$ converges in probability to the constant random variable $X = 0$. To rigorously verify this claim we need to use the formal definition of convergence in probability. For any $\epsilon > 0$,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) &= \lim_{n \rightarrow \infty} P(|X_n - 0| > \epsilon) \\ &= \lim_{n \rightarrow \infty} [1 - P(-\epsilon \leq X_n \leq \epsilon)] \\ &= 1 - \int_{-\epsilon}^{\epsilon} f_{X_n}(x) dx \\ &= 1 - \lim_{n \rightarrow \infty} \int_{\max(-\epsilon, -1/n)}^{\max(\epsilon, 1/n)} \frac{n}{2} dx \\ &= \lim_{n \rightarrow \infty} \int_{-1/n}^{1/n} \frac{n}{2} dx \quad \text{when } n \text{ becomes large } \frac{1}{n} < \epsilon \\ &= 1 - \lim_{n \rightarrow \infty} 1 = 0 \end{aligned}$$

Exercise 4.2 For any natural number $n \geq 0$, consider the function f_n defined by

$$f_n(x) = n^2 x e^{-n^2 x^2/2} \quad \text{if } x > 0.$$

- (a). Show that f_n is the density of random variable.
 (b). Let $\{X_n\}$ be a sequence of random variables such that, for all integer $n \geq 1$, X_n has the density f_n . Show that the sequence $\{X_n\}$ converges in probability to a random variable X that we will specify.

Solution

- (a). Notice that f_n is positive and continuous. Moreover, we have for $x > 0$,

$$\begin{aligned} \int_{-\infty}^x f_n(t) dt &= \int_0^x n^2 t e^{-n^2 t^2/2} dt \\ &= \left[-e^{-n^2 t^2/2} \right]_0^x \\ &= 1 - e^{-n^2 x^2/2}. \end{aligned}$$

when x tends to infinity, we deduce that

$$\int_{-\infty}^{\infty} f(t) dt = 1,$$

which finally proves that f_n is a probability density.

(b). Let $\epsilon > 0$. We have

$$\begin{aligned} P(|X_n| \geq \epsilon) &= \int_{\epsilon}^{\infty} n^2 t e^{-n^2 t^2/2} dt \\ &= \left[-e^{-n^2 t^2/2} \right]_{\epsilon}^{+\infty} \\ &= e^{-n^2 \epsilon^2/2}. \end{aligned}$$

When $n \rightarrow \infty$, this quantity tends to 0. We deduce that $\{X_n\}$ converges in probability to the random variable $X = 0$.

🚩 **Exercise 4.3** $\{X_n; n \in \mathbb{N}\}$ is a sequence of independent random variables with marginal pmfs given by

$$P\left(X_n = \frac{1}{2} \left(1 - \frac{1}{n}\right)\right) = P\left(X_n = \frac{1}{2} \left(1 + \frac{1}{n}\right)\right) = \frac{1}{2}$$

(a). Show that the sequence converges almost surely.

(b). Check if $\{X_n; n \in \mathbb{N}\}$ converges in mean square.

Solution

(a). First, we note that for a fixed $\epsilon > 0$,

$$\begin{aligned} P\left(\left\{\omega : \left|X_n(\omega) - \frac{1}{2}\right| \leq \epsilon\right\}\right) &= P\left(\frac{1}{2} - \epsilon \leq X_n \leq \frac{1}{2} + \epsilon\right) \\ &= \begin{cases} 0, & \text{if } \epsilon < \frac{1}{2n} \text{ (or } n < \frac{1}{2\epsilon}) \\ 1 & \text{o.w.} \end{cases} \end{aligned}$$

Hence, it is immediate that :

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\left\{\omega : \left|X_n(\omega) - \frac{1}{2}\right| \leq \epsilon\right\}\right) &= 0 \\ \Rightarrow X_n &\xrightarrow{p} \frac{1}{2}. \end{aligned}$$

Now, observe that

$$\sum_{n=1}^{\infty} P\left(\left|X_n - \frac{1}{2}\right| > \epsilon\right) = \frac{1}{2\epsilon} < \infty.$$

Hence, $X_n \xrightarrow{\text{a.s.}} \frac{1}{2}$.

(b). We have that

🚩 **Exercise 4.4** Let n be a natural number and let a be a real number. Consider the function f_n defined on \mathbb{R} by

$$f_n(x) = \frac{a \cdot n}{\pi(1 + n^2 x^2)}.$$

(a). Determine the value of a such that f_n be a density of probability of random variable.

(b). Let $\{X_n\}$ be a sequence of random variables such that for each variable X_n have a density f_n . Study the existence of the moments of $\{X_n\}$.

(c). Study the convergence in law of the sequence $\{X_n\}$.

(d). Study the convergence of probability of the sequence $\{X_n\}$.

Solution

- (a). f_n being continuous and positive, it will be a density of probability if and only if

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

Now, performing the change of variables $u = nx$, we have

$$\int_{-\infty}^{\infty} \frac{an}{\pi(1+n^2x^2)}dx = \frac{-\infty}{\infty} \frac{a}{\pi(1+u^2)}du = \frac{a}{\pi} [\arctan(u)]_{-\infty}^{\infty} = \frac{a}{\pi} = a$$

f_n is therefore density of random variables if and only if $a = 1$.

- (b). We have $x f_n(x) \sim_{\text{infy}} \frac{1}{\pi nx}$ whose integral is divergent in the vicinity of $+\infty$, which is a positive function. Thus, the random variable X does not have an expectation, nor any other moment.

- (c). Notice that the cumulative distribution function of X_n defined by

$$F_n(x) = \int_{-\infty}^x f_n(t)dt = \frac{1}{\pi} \left(\arctan(nx) \frac{\pi}{2} \right).$$

If $x < 0$, $\arctan(nx) \rightarrow -\pi/2$, and so $F_n(x) \rightarrow 0$. If $x > 0$, $\arctan(nx) \rightarrow \pi/2$ and so $F_n(x) \rightarrow 1$. Now consider X a random variable

Exercise 4.5 Let A_1, A_2, \dots, A_n be a sequence of independent random variables, with $P[A_i = 1] = P[A_i = \frac{1}{2}] = \frac{1}{2}$ for all i . Let $B_k = A_1.A_2 \dots A_k$.

- (a). Does $\lim_{k \rightarrow \infty} B_k$ exists in the mean square sense? Justify your answer.
 (b). Does $\lim_{k \rightarrow \infty} B_k$ exists in the almost surely sense? Justify your answer.
 (c). Let $S_n = B_1 + B_2 + \dots + B_n$. You can use without proof that the fact that $\lim_{m,n \rightarrow \infty} \mathbf{E}W_m S_n = \frac{35}{3}$, which implies that $\lim_{n \rightarrow \infty} S_n$ exists in the mean square sense. Find the mean and the variance of the limite random variable.
 (d). Does the $\lim_{n \rightarrow \infty} S_n S_n$ exists almost surely? Justify your answers.

Solution

- (a). $\mathbf{E}[(B_k - 0)^0] = \mathbf{E}[(A_1^2)^k] = \frac{5^k}{8} \rightarrow 0$ as $k \rightarrow \infty$. This $\lim_{k \rightarrow \infty} m.s B_k = 0$.
 (b). Each sample path of the sequence B_k is monotone nondecreasing and bounded below by zero, and is hence convergent. Thus $\lim_{k \rightarrow \infty} B_k$ exists a.s. (The limit has to be the same as the mean square limit, so B_k converges to 0 almost surely).
 (c). Meann square convergence implies convergence in the mean. Thus the mean of the limit is

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{E}[S_n] &= \lim_{n \rightarrow \infty} \mathbf{E}\left[\sum_{k=1}^n \mathbf{E}[B_k]\right] \\ &= \sum_{k=1}^{\infty} \left(\frac{3}{4}\right)^k = 3. \end{aligned}$$

- (d). By the propretly given in the course, the second moment of the limit is $\frac{35}{3}$ so the variance of the limit is $\frac{35}{3} - 3^2 = \frac{5}{3}$.

Consider a sequence $X_n \stackrel{d}{=} \mathcal{B}(n, p_n) (n \geq 1)$ of binomial random variables. Suppose that $\lim_{n \rightarrow \infty} np_n = \lambda$ for some constant λ .

- (a). For each $k \geq 0$, compute $\lim_{n \rightarrow \infty} P(X_n = k)$.

(b). If $\lambda > 0$, show that X_n converges weakly to the zero random variable.

Solution

(a). For each $k \geq 1$, we have

$$\begin{aligned} \frac{b_n(k)}{b_n(k-1)} &= \frac{C_n^k p_n^k (1-p_n)^{n-k}}{C_n^{k-1} p_n^{k-1} (1-p_n)^{n-k+1}} \\ &= \frac{n-k+1}{k} \frac{p_n}{1-p_n} \end{aligned}$$

➤ **Exercise 4.6** Let (U_n) be a sequence of independent random variables all following the uniform distribution on $[0, 1]$. Let $M_n = \max(U_1, \dots, U_n)$ and $X_n = n(1 - M_n)$.

(a). What is the distribution function of X_n ?

(b). Study the convergence in law of the sequence (X_n) .

➤ **Exercise 4.7** Let (U_n) be a sequence of independent random variables all following the uniform law on $[0, 1]$. We note $M_n = \max(U_1, \dots, U_n)$ and $X_n = n(1 - M_n)$.

(a). Determine the distribution function of X_n .

(b). Study the convergence in the law of the sequence (X_n)

Solution

(a). We start by determining the distribution function of M_n . Since M_n takes its values in $[0, 1]$, it is clear that if $x \leq 0$, we have $P(M_n \leq x) = 0$ and if $x \geq 1$, we have $P(M_n \leq x) = 1$. As the random variables $(U_i)_{1 \leq i \leq n}$ are independent, we deduce

$$P(M_n \leq x) = \text{Prod}_{i=1}^n P(U_i \leq x) = x^n.$$

To obtain the distribution function of X_n , we remark

$$\begin{aligned} W_n \leq x &\Leftrightarrow M_n \geq 1 - \frac{x}{n} \\ P(X_n \leq x) &= 1 - P\left(M_n \leq 1 - \frac{x}{n}\right) \end{aligned}$$

Moreover,

$$1 - \frac{x}{n} \in [0, 1] \Leftrightarrow x \in [0, n].$$

We deduce that the distribution function of (X_n) is given by

$$F_{X_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - \left(1 - \frac{x}{n}\right)^n & \text{si } 0 \leq x \leq n \\ 1 & \text{if } x \geq n \end{cases}$$

(b). We will study, at x fixed, the limit of F_{X_n} . First, for $x \leq 0$, it is clear that $\lim_{n \rightarrow \infty} F_{X_n}(x) = 0$.

Now, for $x \geq 0$, as soon as n exists large enough, we have $x \leq n$ and therefore

$$F_{X_n}(x) = 1 - \left(1 - \frac{x}{n}\right)^n = e^x.$$

We deduce that $G_{X_n}(x) = 0$ if $t \leq -\ln(n)$, and if $t > -\ln(n)$,

$$G_n(t) = (1 - e^{-t-\ln(n)}) = \left(1 - \frac{e^{-t}}{n}\right)^n$$

To obtain convergence in law, it is necessary and sufficient to prove that for every real t , we

$G_n(x)$ which converges to $F_Y(t)$. But, if n is big enough

$$\begin{aligned} G_n(t) &= \left(1 - \frac{e^{-t}}{n}\right)^n = \exp\left(n \ln\left(1 - \frac{e^{-t}}{n}\right)\right) = \exp\left(n \ln\left(1 - \frac{e^{-t}}{n}\right)\right) \\ &= \exp(-\exp(-t) + o(1)) \rightarrow F_Y(t). \end{aligned}$$

We actually obtain the desired result.

✚ **Exercise 4.8** Let $(X_n)_n$ be a sequence of random variables. We suppose that there exists a sequence $(c_n)_n$ of real numbers such that, for tout $\delta > 0$

$$P(|X_n - c_n| > \delta).$$

We further assume that the sequence $(c_n)_n$ admits a limit c . Show that (X_n) converges to c in probability

✚ **Exercise 4.9** A factory produces parts of which an unknown proportion p is defective, and we want to find an approximate value of p . A sample of n parts is taken. It is assumed that the sample is taken from a very large population, and can therefore be compared to a series of n independent draws. We denote X_n the random variable equal to the number of defective parts and we wish to quantify the fact that X_n/n approaches p .

(a). What is the law of X_n ? Its mean? Its variance?

Statistical Tables

Table 2: Cumulative Poisson Distribution Function

The tabulated value is $P(X \leq x)$, where X has a Poisson distribution with mean λ .

λ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.0	1.2	1.4	1.6	1.8	λ
x															x
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679	0.3012	0.2466	0.2019	0.1653	0
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358	0.6626	0.5918	0.5249	0.4628	1
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197	0.8795	0.8335	0.7834	0.7306	2
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810	0.9662	0.9463	0.9212	0.8913	3
4		1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963	0.9923	0.9857	0.9763	0.9636	4
5				1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994	0.9985	0.9968	0.9940	0.9896	5
6							1.0000	1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9974	6
7										1.0000	1.0000	0.9999	0.9997	0.9994	7
8												1.0000	1.0000	0.9999	8
9														1.0000	9
λ	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.5	5.0	5.5	λ
x															x
0	0.1353	0.1108	0.0907	0.0743	0.0608	0.0498	0.0408	0.0334	0.0273	0.0224	0.0183	0.0111	0.0067	0.0041	0
1	0.4060	0.3546	0.3084	0.2674	0.2311	0.1991	0.1712	0.1468	0.1257	0.1074	0.0916	0.0611	0.0404	0.0266	1
2	0.6767	0.6227	0.5697	0.5184	0.4695	0.4232	0.3799	0.3397	0.3027	0.2689	0.2381	0.1736	0.1247	0.0884	2
3	0.8571	0.8194	0.7787	0.7360	0.6919	0.6472	0.6025	0.5584	0.5152	0.4735	0.4335	0.3423	0.2650	0.2017	3
4	0.9473	0.9275	0.9041	0.8774	0.8477	0.8153	0.7806	0.7442	0.7064	0.6678	0.6288	0.5321	0.4405	0.3575	4
5	0.9834	0.9751	0.9643	0.9510	0.9349	0.9161	0.8946	0.8705	0.8441	0.8156	0.7851	0.7029	0.6160	0.5289	5
6	0.9955	0.9925	0.9884	0.9828	0.9756	0.9665	0.9554	0.9421	0.9267	0.9091	0.8893	0.8311	0.7622	0.6860	6
7	0.9989	0.9980	0.9967	0.9947	0.9919	0.9881	0.9832	0.9769	0.9692	0.9599	0.9489	0.9134	0.8666	0.8095	7
λ	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.5	5.0	5.5	λ
8	0.9998	0.9995	0.9991	0.9985	0.9976	0.9962	0.9943	0.9917	0.9883	0.9840	0.9786	0.9597	0.9319	0.8944	8
9	1.0000	0.9999	0.9998	0.9996	0.9993	0.9989	0.9982	0.9973	0.9960	0.9942	0.9919	0.9829	0.9682	0.9462	9
10		1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992	0.9987	0.9981	0.9972	0.9933	0.9863	0.9747	10
11				1.0000	1.0000	0.9999	0.9999	0.9998	0.9996	0.9994	0.9991	0.9976	0.9945	0.9890	11
12						1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9992	0.9980	0.9955	12
13								1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9983	13
14											1.0000	0.9999	0.9998	0.9994	14
15												1.0000	0.9999	0.9998	15
16													1.0000	0.9999	16
17														1.0000	17

λ	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	11.0	12.0	13.0	14.0	15.0	λ
x															x
0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0
1	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	1
2	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000	2
3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002	3
4	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009	4
5	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028	5
6	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076	6
7	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180	7
8	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374	8
9	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699	9
10	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185	10
11	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848	11
12	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676	12
13	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632	13
14	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657	14
15	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681	15
16	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641	16
17	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489	17
18	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195	18
19		1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752	19
20				1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170	20
21					1.0000	0.9999	0.9998	0.9996	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469	21
22						1.0000	0.9999	0.9999	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673	22
23							1.0000	0.9999	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805	23
24								1.0000	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888	24
25										0.9999	0.9997	0.9990	0.9974	0.9938	25
26										1.0000	0.9999	0.9995	0.9987	0.9967	26
27											0.9999	0.9998	0.9994	0.9983	27
28											1.0000	0.9999	0.9997	0.9991	28
29												1.0000	0.9999	0.9996	29
30													0.9999	0.9998	30
31													1.0000	0.9999	31
32														1.0000	32

Table 3: Normal Distribution Function

The table gives the probability, p , that a normally distributed random variable Z , with mean = 0 and variance = 1, is less than or equal to z .

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.0
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.1
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.2
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.3
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.4
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.5
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	0.6
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	0.7
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	0.8
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	0.9
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	1.0
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	1.1
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	1.2
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	1.3
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	1.4
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	1.5
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	1.6
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	1.7
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	1.8
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	1.9
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	2.1
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	2.2
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	2.3
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	2.4
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	2.5
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	2.6
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	2.7
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	2.8
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	2.9
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900	3.0
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929	3.1
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	3.2
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965	3.3
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	3.4
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983	3.5
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989	3.6
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992	3.7
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995	3.8
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997	3.9

Table 4: Percentage Points of the Normal Distribution

The table gives the values of z satisfying $P(Z \leq z) = p$, where Z is the normally distributed random variable with mean = 0 and variance = 1.

p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	p
0.5	0.0000	0.0251	0.0502	0.0753	0.1004	0.1257	0.1510	0.1764	0.2019	0.2275	0.5
0.6	0.2533	0.2793	0.3055	0.3319	0.3585	0.3853	0.4125	0.4399	0.4677	0.4958	0.6
0.7	0.5244	0.5534	0.5828	0.6128	0.6433	0.6745	0.7063	0.7388	0.7722	0.8064	0.7
0.8	0.8416	0.8779	0.9154	0.9542	0.9945	1.0364	1.0803	1.1264	1.1750	1.2265	0.8
0.9	1.2816	1.3408	1.4051	1.4758	1.5548	1.6449	1.7507	1.8808	2.0537	2.3263	0.9
p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	p
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392	0.95
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663	0.96
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335	0.97
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904	0.98
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902	0.99

Table 5: Percentage Points of the Student’s *t*-distribution

The table gives the values of x satisfying $P(X \leq x) = p$, where X is a random variable having the Student’s t -distribution with ν degrees of freedom.

p	0.9	0.95	0.975	0.99	0.995
ν					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.121	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763

p	0.9	0.95	0.975	0.99	0.995
ν					
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
55	1.297	1.673	2.004	2.396	2.668
60	1.296	1.671	2.000	2.390	2.660
65	1.295	1.669	1.997	2.385	2.654
70	1.294	1.667	1.994	2.381	2.648
75	1.293	1.665	1.992	2.377	2.643
80	1.292	1.664	1.990	2.374	2.639
85	1.292	1.663	1.998	2.371	2.635
90	1.291	1.662	1.987	2.368	2.632
95	1.291	1.661	1.985	2.366	2.629
100	1.290	1.660	1.984	2.364	2.626
125	1.288	1.657	1.979	2.357	2.616
150	1.287	1.655	1.976	2.351	2.609
200	1.286	1.653	1.972	2.345	2.601
∞	1.282	1.645	1.960	2.326	2.576

Table 6: Percentage Points of the χ^2 Distribution

The table gives the values of x satisfying $P(X \leq x) = p$, where X is a random variable having the χ^2 distribution with ν degrees of freedom.

p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995	p
ν											ν
1	0.00004	0.0002	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559	24
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645	27
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672	30
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003	31
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328	32
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648	33
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.996	56.061	58.964	34
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275	35
36	17.887	19.223	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581	36
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892	62.883	37
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181	38
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476	39
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766	40
45	24.311	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957	73.166	45
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490	50
55	31.735	33.570	36.398	38.958	42.060	68.796	73.311	77.380	82.292	85.749	55
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952	60
65	39.383	41.444	44.603	47.450	50.883	79.973	84.821	89.177	94.422	98.105	65
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215	70
75	47.206	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393	110.286	75
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321	80
85	55.170	57.634	61.389	64.749	68.777	102.079	107.522	112.393	118.236	122.325	85
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299	90
95	63.250	65.898	69.925	73.520	77.818	113.038	118.752	123.858	129.973	134.247	95
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169	100

Table 7: Percentage Points of the *F*-distribution

The tables give the values of x satisfying $P(X \leq x) = p$, where X is a random variable having the *F*-distribution with v_1 degrees of freedom in the numerator and v_2 degrees of freedom in the denominator.

***F*-distribution ($p = 0.995$)**

Use for one-tail tests at significance level 0.5% or two-tail tests at significance level 1%.

v_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100	∞	v_1	
v_2																						v_2
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	24630	24836	24960	25044	25148	25211	25337	25464		1
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5	199.5	2
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.52	43.39	43.08	42.78	42.59	42.47	42.31	42.21	42.02	41.83		3
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.82	20.70	20.44	20.17	20.00	19.89	19.75	19.67	19.50	19.32		4
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.49	13.38	13.15	12.90	12.76	12.66	12.53	12.45	12.30	12.14		5
6	18.635	14.544	12.917	12.028	11.464	11.073	10.786	10.566	10.391	10.250	10.133	10.034	9.814	9.589	9.451	9.358	9.241	9.170	9.026	8.879		6
7	16.236	12.404	10.882	10.050	9.522	9.155	8.885	8.678	8.514	8.380	8.270	8.176	7.968	7.754	7.623	7.534	7.422	7.354	7.217	7.076		7
8	14.688	11.042	9.596	8.805	8.302	7.952	7.694	7.496	7.339	7.211	7.104	7.015	6.814	6.608	6.482	6.396	6.288	6.222	6.088	5.951		8
9	13.614	10.107	8.717	7.956	7.471	7.134	6.885	6.693	6.541	6.417	6.314	6.227	6.032	5.832	5.708	5.625	5.519	5.454	5.322	5.188		9
10	12.826	9.427	8.081	7.343	6.872	6.545	6.302	6.116	5.968	5.847	5.746	5.661	5.471	5.274	5.153	5.071	4.966	4.902	4.772	4.639		10
11	12.226	8.912	7.600	6.881	6.422	6.102	5.865	5.682	5.537	5.418	5.320	5.236	5.049	4.855	4.736	4.654	4.551	4.488	4.359	4.226		11
12	11.754	8.510	7.226	6.521	6.071	5.757	5.525	5.345	5.202	5.085	4.988	4.906	4.721	4.530	4.412	4.331	4.228	4.165	4.037	3.904		12
13	11.374	8.186	6.926	6.233	5.791	5.482	5.253	5.076	4.935	4.820	4.724	4.643	4.460	4.270	4.153	4.073	3.970	3.908	3.780	3.647		13
14	11.060	7.922	6.680	5.998	5.562	5.257	5.031	4.857	4.717	4.603	4.508	4.428	4.247	4.059	3.942	3.862	3.760	3.697	3.569	3.436		14
15	10.798	7.701	6.476	5.803	5.372	5.071	4.847	4.674	4.536	4.424	4.329	4.250	4.070	3.883	3.766	3.687	3.585	3.523	3.394	3.260		15
20	9.944	6.986	5.818	5.174	4.762	4.472	4.257	4.090	3.956	3.847	3.756	3.678	3.502	3.318	3.203	3.123	3.022	2.959	2.828	2.690		20
25	9.475	6.598	5.462	4.835	4.433	4.150	3.939	3.776	3.645	3.537	3.447	3.370	3.196	3.013	2.898	2.819	2.716	2.652	2.519	2.377		25
30	9.180	6.355	5.239	4.623	4.228	3.949	3.742	3.580	3.450	3.344	3.255	3.179	3.006	2.823	2.708	2.628	2.524	2.459	2.323	2.176		30
40	8.828	6.066	4.976	4.374	3.986	3.713	3.509	3.350	3.222	3.117	3.028	2.953	2.781	2.598	2.482	2.401	2.296	2.230	2.088	1.932		40
50	8.626	5.902	4.826	4.232	3.849	3.579	3.376	3.219	3.092	2.988	2.900	2.825	2.653	2.470	2.353	2.272	2.164	2.097	1.951	1.786		50
100	8.241	5.589	4.542	3.963	3.589	3.325	3.127	2.972	2.847	2.744	2.657	2.583	2.411	2.227	2.108	2.024	1.912	1.840	1.681	1.485		100
∞	7.879	5.298	4.279	3.715	3.350	3.091	2.897	2.744	2.621	2.519	2.432	2.358	2.187	2.000	1.877	1.789	1.669	1.590	1.402	1.001		∞

***F*-distribution ($p = 0.99$)**

Use for one-tail tests at significance level 1% or two-tail tests at significance level 2%.

v_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100	∞	v_1	
v_2																						v_2
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6157	6209	6240	6261	6287	6303	6334	6366		1
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50		2
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.13	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.24	26.13		3
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.58	13.46		4
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.13	9.02		5
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790	7.718	7.559	7.396	7.296	7.229	7.143	7.091	6.987	6.880		6
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538	6.469	6.314	6.155	6.058	5.992	5.908	5.858	5.755	5.650		7
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734	5.667	5.515	5.359	5.263	5.198	5.116	5.065	4.963	4.859		8
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178	5.111	4.962	4.808	4.713	4.649	4.567	4.517	4.415	4.311		9
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772	4.706	4.558	4.405	4.311	4.247	4.165	4.115	4.014	3.909		10
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462	4.397	4.251	4.099	4.005	3.941	3.860	3.810	3.708	3.602		11
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220	4.155	4.010	3.858	3.765	3.701	3.619	3.569	3.467	3.361		12
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	4.025	3.960	3.815	3.665	3.571	3.507	3.425	3.375	3.272	3.165		13
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864	3.800	3.656	3.505	3.412	3.348	3.266	3.215	3.112	3.004		14
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730	3.666	3.522	3.372	3.278	3.214	3.132	3.081	2.977	2.868		15
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294	3.231	3.088	2.938	2.843	2.778	2.695	2.643	2.535	2.421		20
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056	2.993	2.850	2.699	2.604	2.538	2.453	2.400	2.289	2.169		25
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.906	2.843	2.700	2.549	2.453	2.386	2.299	2.245	2.131	2.006		30
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727	2.665	2.522	2.369	2.271	2.203	2.114	2.058	1.938	1.805		40
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.625	2.562	2.419	2.265	2.167	2.098	2.007	1.949	1.825	1.683		50
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.430	2.368	2.223	2.067	1.965	1.893	1.797	1.735	1.598	1.427		100
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.248	2.185	2.039	1.878	1.773	1.696	1.592	1.523	1.358	1.000		∞

F-distribution ($p = 0.975$)

Use for one-tail tests at significance level 2.5% or two-tail tests at significance level 5%.

v_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100	∞	v_1	
v_2																						v_2
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	973.0	976.7	984.9	993.1	998.1	1001.4	1005.6	1008.1	1013.2	1018.3		1
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50		2
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.37	14.34	14.25	14.17	14.12	14.08	14.04	14.01	13.96	13.90		3
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.66	8.56	8.50	8.46	8.41	8.38	8.32	8.26		4
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.43	6.33	6.27	6.23	6.18	6.14	6.08	6.02		5
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.410	5.366	5.269	5.168	5.107	5.065	5.012	4.980	4.915	4.849		6
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.709	4.666	4.568	4.467	4.405	4.362	4.309	4.276	4.210	4.142		7
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.243	4.200	4.101	3.999	3.937	3.894	3.840	3.807	3.739	3.670		8
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.912	3.868	3.769	3.667	3.604	3.560	3.505	3.472	3.403	3.333		9
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.665	3.621	3.522	3.419	3.355	3.311	3.255	3.221	3.152	3.080		10
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.474	3.430	3.330	3.226	3.162	3.118	3.061	3.027	2.956	2.883		11
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.321	3.277	3.177	3.073	3.008	2.963	2.906	2.871	2.800	2.725		12
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.197	3.153	3.053	2.948	2.882	2.837	2.780	2.744	2.671	2.595		13
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.095	3.050	2.949	2.844	2.778	2.732	2.674	2.638	2.565	2.487		14
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	3.008	2.963	2.862	2.756	2.689	2.644	2.585	2.549	2.474	2.395		15
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.721	2.676	2.573	2.464	2.396	2.349	2.287	2.249	2.170	2.085		20
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.560	2.515	2.411	2.300	2.230	2.182	2.118	2.079	1.996	1.906		25
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.458	2.412	2.307	2.195	2.124	2.074	2.009	1.968	1.882	1.787		30
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.334	2.288	2.182	2.068	1.994	1.943	1.875	1.832	1.741	1.637		40
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.263	2.216	2.109	1.993	1.919	1.866	1.796	1.752	1.656	1.545		50
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.125	2.077	1.968	1.849	1.770	1.715	1.640	1.592	1.483	1.347		100
∞	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114	2.048	1.993	1.945	1.833	1.708	1.626	1.566	1.484	1.428	1.296	1.000		∞

F-distribution ($p = 0.95$)

Use for one-tail tests at significance level 5% or two-tail tests at significance level 10%.

v_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100	∞	v_1	
v_2																						v_2
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	245.9	248.0	249.3	250.1	251.1	251.8	253.0	254.3		1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.50		2
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.55	8.53		3
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.66	5.63		4
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.41	4.36		5
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.938	3.874	3.835	3.808	3.774	3.754	3.712	3.669		6
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.511	3.445	3.404	3.376	3.340	3.319	3.275	3.230		7
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.218	3.150	3.108	3.079	3.043	3.020	2.975	2.928		8
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.006	2.936	2.893	2.864	2.826	2.803	2.756	2.707		9
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.845	2.774	2.730	2.700	2.661	2.637	2.588	2.538		10
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.719	2.646	2.601	2.570	2.531	2.507	2.457	2.404		11
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.617	2.544	2.498	2.466	2.426	2.401	2.350	2.296		12
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.533	2.459	2.412	2.380	2.339	2.314	2.261	2.206		13
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.463	2.388	2.341	2.308	2.266	2.241	2.187	2.131		14
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.403	2.328	2.280	2.247	2.204	2.178	2.123	2.066		15
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.203	2.124	2.074	2.039	1.994	1.966	1.907	1.843		20
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198	2.165	2.089	2.007	1.955	1.919	1.872	1.842	1.779	1.711		25
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126	2.092	2.015	1.932	1.878	1.841	1.792	1.761	1.695	1.622		30
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038	2.003	1.924	1.839	1.783	1.744	1.693	1.660	1.589	1.509		40
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.986	1.952	1.871	1.784	1.727	1.687	1.634	1.599	1.525	1.438		50
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886	1.850	1.768	1.676	1.616	1.573	1.515	1.477	1.392	1.283		100
∞	3.841	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.789	1.752	1.666	1.571	1.506	1.459	1.394	1.350	1.243	1.000		∞

Table 8: Critical Values of the Product Moment Correlation Coefficient

The table gives the critical values, for different significance levels, of the product moment correlation coefficient, r , for varying sample sizes, n .

One tail Two tail	10% 20%	5% 10%	2.5% 5%	1% 2%	0.5% 1%	One tail Two tail
<i>n</i>						<i>n</i>
4	0.8000	0.9000	0.9500	0.9800	0.9900	4
5	0.6870	0.8054	0.8783	0.9343	0.9587	5
6	0.6084	0.7293	0.8114	0.8822	0.9172	6
7	0.5509	0.6694	0.7545	0.8329	0.8745	7
8	0.5067	0.6215	0.7067	0.7887	0.8343	8
9	0.4716	0.5822	0.6664	0.7498	0.7977	9
10	0.4428	0.5494	0.6319	0.7155	0.7646	10
11	0.4187	0.5214	0.6021	0.6851	0.7348	11
12	0.3981	0.4973	0.5760	0.6581	0.7079	12
13	0.3802	0.4762	0.5529	0.6339	0.6835	13
14	0.3646	0.4575	0.5324	0.6120	0.6614	14
15	0.3507	0.4409	0.5140	0.5923	0.6411	15
16	0.3383	0.4259	0.4973	0.5742	0.6226	16
17	0.3271	0.4124	0.4821	0.5577	0.6055	17
18	0.3170	0.4000	0.4683	0.5425	0.5897	18
19	0.3077	0.3887	0.4555	0.5285	0.5751	19
20	0.2992	0.3783	0.4438	0.5155	0.5614	20
21	0.2914	0.3687	0.4329	0.5034	0.5487	21
22	0.2841	0.3598	0.4227	0.4921	0.5368	22
23	0.2774	0.3515	0.4132	0.4815	0.5256	23
24	0.2711	0.3438	0.4044	0.4716	0.5151	24
25	0.2653	0.3365	0.3961	0.4622	0.5052	25
26	0.2598	0.3297	0.3882	0.4534	0.4958	26
27	0.2546	0.3233	0.3809	0.4451	0.4869	27
28	0.2497	0.3172	0.3739	0.4372	0.4785	28
29	0.2451	0.3115	0.3673	0.4297	0.4705	29
30	0.2407	0.3061	0.3610	0.4226	0.4629	30
31	0.2366	0.3009	0.3550	0.4158	0.4556	31
32	0.2327	0.2960	0.3494	0.4093	0.4487	32
33	0.2289	0.2913	0.3440	0.4032	0.4421	33
34	0.2254	0.2869	0.3388	0.3972	0.4357	34
35	0.2220	0.2826	0.3338	0.3916	0.4296	35
36	0.2187	0.2785	0.3291	0.3862	0.4238	36
37	0.2156	0.2746	0.3246	0.3810	0.4182	37
38	0.2126	0.2709	0.3202	0.3760	0.4128	38
39	0.2097	0.2673	0.3160	0.3712	0.4076	39
40	0.2070	0.2638	0.3120	0.3665	0.4026	40
41	0.2043	0.2605	0.3081	0.3621	0.3978	41
42	0.2018	0.2573	0.3044	0.3578	0.3932	42
43	0.1993	0.2542	0.3008	0.3536	0.3887	43
44	0.1970	0.2512	0.2973	0.3496	0.3843	44
45	0.1947	0.2483	0.2940	0.3457	0.3801	45
46	0.1925	0.2455	0.2907	0.3420	0.3761	46
47	0.1903	0.2429	0.2876	0.3384	0.3721	47
48	0.1883	0.2403	0.2845	0.3348	0.3683	48
49	0.1863	0.2377	0.2816	0.3314	0.3646	49
50	0.1843	0.2353	0.2787	0.3281	0.3610	50
60	0.1678	0.2144	0.2542	0.2997	0.3301	60
70	0.1550	0.1982	0.2352	0.2776	0.3060	70
80	0.1448	0.1852	0.2199	0.2597	0.2864	80
90	0.1364	0.1745	0.2072	0.2449	0.2702	90
100	0.1292	0.1654	0.1966	0.2324	0.2565	100

Table 9: Critical Values of Spearman’s Rank Correlation Coefficient

The table gives the critical values, for different significance levels, of Spearman’s rank correlation coefficient, r_s , for varying sample sizes, n . Since r_s is discrete, exact significance levels cannot be obtained in most cases. The critical values given are those with significance levels closest to the stated value.

One tail Two tail	10% 20%	5% 10%	2.5% 5%	1% 2%	0.5% 1%	One tail Two tail
<i>n</i>						<i>n</i>
4	1.0000	1.0000	1.0000	1.0000	1.0000	4
5	0.7000	0.9000	0.9000	1.0000	1.0000	5
6	0.6571	0.7714	0.8286	0.9429	0.9429	6
7	0.5714	0.6786	0.7857	0.8571	0.8929	7
8	0.5476	0.6429	0.7381	0.8095	0.8571	8
9	0.4833	0.6000	0.6833	0.7667	0.8167	9
10	0.4424	0.5636	0.6485	0.7333	0.7818	10
11	0.4182	0.5273	0.6091	0.7000	0.7545	11
12	0.3986	0.5035	0.5874	0.6713	0.7273	12
13	0.3791	0.4780	0.5604	0.6484	0.6978	13
14	0.3670	0.4593	0.5385	0.6220	0.6747	14
15	0.3500	0.4429	0.5179	0.6000	0.6536	15
16	0.3382	0.4265	0.5029	0.5824	0.6324	16
17	0.3271	0.4124	0.4821	0.5577	0.6055	17
18	0.3170	0.4000	0.4683	0.5425	0.5897	18
19	0.3077	0.3887	0.4555	0.5285	0.5751	19
20	0.2992	0.3783	0.4438	0.5155	0.5614	20
21	0.2914	0.3687	0.4329	0.5034	0.5487	21
22	0.2841	0.3598	0.4227	0.4921	0.5368	22
23	0.2774	0.3515	0.4132	0.4815	0.5256	23
24	0.2711	0.3438	0.4044	0.4716	0.5151	24
25	0.2653	0.3365	0.3961	0.4622	0.5052	25
26	0.2598	0.3297	0.3882	0.4534	0.4958	26
27	0.2546	0.3233	0.3809	0.4451	0.4869	27
28	0.2497	0.3172	0.3739	0.4372	0.4785	28
29	0.2451	0.3115	0.3673	0.4297	0.4705	29
30	0.2407	0.3061	0.3610	0.4226	0.4629	30
31	0.2366	0.3009	0.3550	0.4158	0.4556	31
32	0.2327	0.2960	0.3494	0.4093	0.4487	32
33	0.2289	0.2913	0.3440	0.4032	0.4421	33
34	0.2254	0.2869	0.3388	0.3972	0.4357	34
35	0.2220	0.2826	0.3338	0.3916	0.4296	35
36	0.2187	0.2785	0.3291	0.3862	0.4238	36
37	0.2156	0.2746	0.3246	0.3810	0.4182	37
38	0.2126	0.2709	0.3202	0.3760	0.4128	38
39	0.2097	0.2673	0.3160	0.3712	0.4076	39
40	0.2070	0.2638	0.3120	0.3665	0.4026	40
41	0.2043	0.2605	0.3081	0.3621	0.3978	41
42	0.2018	0.2573	0.3044	0.3578	0.3932	42
43	0.1993	0.2542	0.3008	0.3536	0.3887	43
44	0.1970	0.2512	0.2973	0.3496	0.3843	44
45	0.1947	0.2483	0.2940	0.3457	0.3801	45
46	0.1925	0.2455	0.2907	0.3420	0.3761	46
47	0.1903	0.2429	0.2876	0.3384	0.3721	47
48	0.1883	0.2403	0.2845	0.3348	0.3683	48
49	0.1863	0.2377	0.2816	0.3314	0.3646	49
50	0.1843	0.2353	0.2787	0.3281	0.3610	50
60	0.1678	0.2144	0.2542	0.2997	0.3301	60
70	0.1550	0.1982	0.2352	0.2776	0.3060	70
80	0.1448	0.1852	0.2199	0.2597	0.2864	80
90	0.1364	0.1745	0.2072	0.2449	0.2702	90
100	0.1292	0.1654	0.1966	0.2324	0.2565	100

<i>p</i>	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	<i>p</i>
<i>x</i>	<i>n=20</i>																		<i>x</i>
0	0.8179	0.6676	0.5438	0.4420	0.3585	0.2901	0.2342	0.1887	0.1516	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000	0
1	0.9831	0.9401	0.8802	0.8103	0.7358	0.6605	0.5869	0.5169	0.4516	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000	1
2	0.9990	0.9929	0.9790	0.9561	0.9245	0.8850	0.8390	0.7879	0.7334	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002	2
3	1.0000	0.9994	0.9973	0.9926	0.9841	0.9710	0.9529	0.9294	0.9007	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013	3
4		1.0000	0.9997	0.9990	0.9974	0.9944	0.9893	0.9817	0.9710	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059	4
5			1.0000	0.9999	0.9997	0.9991	0.9981	0.9962	0.9932	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207	5
6				1.0000	1.0000	0.9999	0.9997	0.9994	0.9987	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577	6
7						1.0000	1.0000	0.9999	0.9998	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316	7
8								1.0000	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517	8
9										1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119	9
10											1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881	10
11												0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483	11
12												1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684	12
13													1.0000	0.9997	0.9985	0.9935	0.9786	0.9423	13
14														1.0000	0.9997	0.9984	0.9936	0.9793	14
15															1.0000	0.9997	0.9985	0.9941	15
16																1.0000	0.9997	0.9987	16
17																	1.0000	0.9998	17
18																		1.0000	18
<i>x</i>	<i>n=25</i>																		<i>x</i>
0	0.7778	0.6035	0.4670	0.3604	0.2774	0.2129	0.1630	0.1244	0.0946	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0
1	0.9742	0.9114	0.8280	0.7358	0.6424	0.5527	0.4696	0.3947	0.3286	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000	1
2	0.9980	0.9868	0.9620	0.9235	0.8729	0.8129	0.7466	0.6768	0.6063	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000	2
3	0.9999	0.9986	0.9938	0.9835	0.9659	0.9402	0.9064	0.8649	0.8169	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001	3
4	1.0000	0.9999	0.9992	0.9972	0.9928	0.9850	0.9726	0.9549	0.9314	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005	4
5		1.0000	0.9999	0.9996	0.9988	0.9969	0.9935	0.9877	0.9790	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020	5
6			1.0000	1.0000	0.9998	0.9995	0.9987	0.9972	0.9946	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073	6
7					1.0000	0.9999	0.9998	0.9995	0.9989	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216	7
8						1.0000	1.0000	0.9999	0.9998	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539	8
9								1.0000	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148	9
10										1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122	10
11											0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450	11
12											1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000	12
13												0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550	13
14												1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878	14
15													1.0000	0.9995	0.9971	0.9868	0.9560	0.8852	15
16														0.9999	0.9992	0.9957	0.9826	0.9461	16
17														1.0000	0.9998	0.9988	0.9942	0.9784	17
18															1.0000	0.9997	0.9984	0.9927	18
19																0.9999	0.9996	0.9980	19
20																1.0000	0.9999	0.9995	20
21																	1.0000	0.9999	21
22																		1.0000	22

<i>p</i>	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	<i>p</i>
<i>x</i>	<i>n</i> =30																		<i>x</i>
0	0.7397	0.5455	0.4010	0.2939	0.2146	0.1563	0.1134	0.0820	0.0591	0.0424	0.0076	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0
1	0.9639	0.8795	0.7731	0.6612	0.5535	0.4555	0.3694	0.2958	0.2343	0.1837	0.0480	0.0105	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000	1
2	0.9967	0.9783	0.9399	0.8831	0.8122	0.7324	0.6487	0.5654	0.4855	0.4114	0.1514	0.0442	0.0106	0.0021	0.0003	0.0000	0.0000	0.0000	2
3	0.9998	0.9971	0.9881	0.9694	0.9392	0.8974	0.8450	0.7842	0.7175	0.6474	0.3217	0.1227	0.0374	0.0093	0.0019	0.0003	0.0000	0.0000	3
4	1.0000	0.9997	0.9982	0.9937	0.9844	0.9685	0.9447	0.9126	0.8723	0.8245	0.5245	0.2552	0.0979	0.0302	0.0075	0.0015	0.0002	0.0000	4
5		1.0000	0.9998	0.9989	0.9967	0.9921	0.9838	0.9707	0.9519	0.9268	0.7106	0.4275	0.2026	0.0766	0.0233	0.0057	0.0011	0.0002	5
6			1.0000	0.9999	0.9994	0.9983	0.9960	0.9918	0.9848	0.9742	0.8474	0.6070	0.3481	0.1595	0.0586	0.0172	0.0040	0.0007	6
7				1.0000	0.9999	0.9997	0.9992	0.9980	0.9959	0.9922	0.9302	0.7608	0.5143	0.2814	0.1238	0.0435	0.0121	0.0026	7
8					1.0000	1.0000	0.9999	0.9996	0.9990	0.9980	0.9722	0.8713	0.6736	0.4315	0.2247	0.0940	0.0312	0.0081	8
9							1.0000	0.9999	0.9998	0.9995	0.9903	0.9389	0.8034	0.5888	0.3575	0.1763	0.0694	0.0214	9
10								1.0000	0.9999	0.9999	0.9971	0.9744	0.8943	0.7304	0.5078	0.2915	0.1350	0.0494	10
11									1.0000	0.9992	0.9905	0.9493	0.8407	0.6548	0.4311	0.2327	0.1002		11
12										0.9998	0.9969	0.9784	0.9155	0.7802	0.5785	0.3592	0.1808		12
13											1.0000	0.9991	0.9918	0.9599	0.8737	0.7145	0.5025	0.2923	13
14												0.9998	0.9973	0.9831	0.9348	0.8246	0.6448	0.4278	14
15												0.9999	0.9992	0.9936	0.9699	0.9029	0.7691	0.5722	15
16												1.0000	0.9998	0.9979	0.9876	0.9519	0.8644	0.7077	16
17													0.9999	0.9994	0.9955	0.9788	0.9286	0.8192	17
18													1.0000	0.9998	0.9986	0.9917	0.9666	0.8998	18
19														1.0000	0.9996	0.9971	0.9862	0.9506	19
20															0.9999	0.9991	0.9950	0.9786	20
21															1.0000	0.9998	0.9984	0.9919	21
22																1.0000	0.9996	0.9974	22
23																	0.9999	0.9993	23
24																	1.0000	0.9998	24
25																		1.0000	25

<i>p</i>	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	<i>p</i>
<i>x</i>	<i>n</i> =50																		<i>x</i>
0	0.6050	0.3642	0.2181	0.1299	0.0769	0.0453	0.0266	0.0155	0.0090	0.0052	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0
1	0.9106	0.7358	0.5553	0.4005	0.2794	0.1900	0.1265	0.0827	0.0532	0.0338	0.0029	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1
2	0.9862	0.9216	0.8108	0.6767	0.5405	0.4162	0.3108	0.2260	0.1605	0.1117	0.0142	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	2
3	0.9984	0.9822	0.9372	0.8609	0.7604	0.6473	0.5327	0.4253	0.3303	0.2503	0.0460	0.0057	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	3
4	0.9999	0.9968	0.9832	0.9510	0.8964	0.8206	0.7290	0.6290	0.5277	0.4312	0.1121	0.0185	0.0021	0.0002	0.0000	0.0000	0.0000	0.0000	4
5	1.0000	0.9995	0.9963	0.9856	0.9622	0.9224	0.8650	0.7919	0.7072	0.6161	0.2194	0.0480	0.0070	0.0007	0.0001	0.0000	0.0000	0.0000	5
6		0.9999	0.9993	0.9964	0.9882	0.9711	0.9417	0.8981	0.8404	0.7702	0.3613	0.1034	0.0194	0.0025	0.0002	0.0000	0.0000	0.0000	6
7		1.0000	0.9999	0.9992	0.9968	0.9906	0.9780	0.9562	0.9232	0.8779	0.5188	0.1904	0.0453	0.0073	0.0008	0.0001	0.0000	0.0000	7
8			1.0000	0.9999	0.9992	0.9973	0.9927	0.9833	0.9672	0.9421	0.6681	0.3073	0.0916	0.0183	0.0025	0.0002	0.0000	0.0000	8
9				1.0000	0.9998	0.9993	0.9978	0.9944	0.9875	0.9755	0.7911	0.4437	0.1637	0.0402	0.0067	0.0008	0.0001	0.0000	9
10					1.0000	0.9998	0.9994	0.9983	0.9957	0.9906	0.8801	0.5836	0.2622	0.0789	0.0160	0.0022	0.0002	0.0000	10
11						1.0000	0.9999	0.9995	0.9987	0.9968	0.9372	0.7107	0.3816	0.1390	0.0342	0.0057	0.0006	0.0000	11
12							1.0000	0.9999	0.9996	0.9990	0.9699	0.8139	0.5110	0.2229	0.0661	0.0133	0.0018	0.0002	12
13								1.0000	0.9999	0.9997	0.9868	0.8894	0.6370	0.3279	0.1163	0.0280	0.0045	0.0005	13
14									1.0000	0.9999	0.9947	0.9393	0.7481	0.4468	0.1878	0.0540	0.0104	0.0013	14
15										1.0000	0.9981	0.9692	0.8369	0.5692	0.2801	0.0955	0.0220	0.0033	15
16											0.9993	0.9856	0.9017	0.6839	0.3889	0.1561	0.0427	0.0077	16
17											0.9998	0.9937	0.9449	0.7822	0.5060	0.2369	0.0765	0.0164	17
18											0.9999	0.9975	0.9713	0.8594	0.6216	0.3356	0.1273	0.0325	18
19											1.0000	0.9991	0.9861	0.9152	0.7264	0.4465	0.1974	0.0595	19
20												0.9997	0.9937	0.9522	0.8139	0.5610	0.2862	0.1013	20
21												0.9999	0.9974	0.9749	0.8813	0.6701	0.3900	0.1611	21
22												1.0000	0.9990	0.9877	0.9290	0.7660	0.5019	0.2399	22
23													0.9996	0.9944	0.9604	0.8438	0.6134	0.3359	23
24													0.9999	0.9976	0.9793	0.9022	0.7160	0.4439	24
25													1.0000	0.9991	0.9900	0.9427	0.8034	0.5561	25
26														0.9997	0.9955	0.9686	0.8721	0.6641	26
27														0.9999	0.9981	0.9840	0.9220	0.7601	27
28														1.0000	0.9993	0.9924	0.9556	0.8389	28
29															0.9997	0.9966	0.9765	0.8987	29
30															0.9999	0.9986	0.9884	0.9405	30
31															1.0000	0.9995	0.9947	0.9675	31
32																0.9998	0.9978	0.9836	32
33																0.9999	0.9991	0.9923	33
34																1.0000	0.9997	0.9967	34
35																	0.9999	0.9987	35
36																	1.0000	0.9995	36
37																		0.9998	37
38																		1.0000	38

Bibliography

- [1] Brockwell, P. J., & Davis, R. A. (1991). *Time series: theory and methods*. Springer science business media.
- [2] Brockwell, P. J., & Davis, R. A. (Eds.). (2002). *Introduction to time series and forecasting*. New York, NY: Springer New York.
- [3] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how* (Vol. 488). London: springer.
- [4] Devore, J. L. (2008). *Probability and Statistics for Engineering and the Sciences*.
- [5] Durrett, R. (2019). *Probability: theory and examples* (Vol. 49). Cambridge university press.
- [6] Lecoutre, J. P. (2002). *Statistique et probabilités*. Dunod.
- [7] Mood, Alexander M., Graybill, Franklin A., and Boes, Duane C.. 1974. *Introduction to the Theory of Statistics*. 3rd ed.London: Macmillan.
- [8] Parthasarathy, H. (2022). *Advanced Probability and Statistics: Applications to Physics and Engineering*. Taylor & Francis.
- [9] Papoulis, A., & Unnikrishna Pillai, S. (2002). *Probability, random variables and stochastic processes*.
- [10] Rényi, A. (2007). *Probability theory*. Courier Corporation.
- [11] Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.
- [12] Rozanov, Y. A. (2013). *Probability theory: a concise course*. Courier Corporation.
- [13] Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions technip.
- [14] Shu, H. P. (1997). *Probability, Random variables and Random Processes*. Schaum's Outlines.
- [15] Suhov, Y., & Kelbert, M. (2014). *Probability and statistics by example: Volume 1, basic probability and statistics*. Cambridge University Press.
- [16] Tassi, P., & Legait, S. (1990). *Théorie des probabilités en vue des applications statistiques*. Editions Technip.