Lecture Notes for Probability Theory - STAT 215

Elaborated by Wissem Jedidi
Department of STAT & OR, College of Science, King Saud University

January 2026

The goal of these lecture notes is to familiarize students with elementary probability theory. In addition, the material is designed to prepare students for probability-based professional certification examinations, such as Exam P of the Society of Actuaries (SOA). To this end, the theory is presented in Part I and is divided into three main topics:

- **Topic 1**: In Chapter 1, we review elementary principles of combinatorics. Chapter 2 introduces the probability measure as a function on events.
- **Topic 2**: This topic is presented in Chapter 3 and explores the concept of random variables on the real line (univariate random variables).
- **Topic 3**: This topic is presented in Chapter 4 and completes Topic 2 by studying random variables in higher dimensions (multivariate random variables), with a focus on the bivariate case.

In Part II (Chapters 5, 6, and 7), the necessary mathematical background is included solely to assist students and make the lecture notes self-contained; these chapters serve only as prerequisites.

These lecture notes are divided into chapters, sections, and subsections. For example, 3.2.4 refers to Chapter 3, Section 2, Subsection 4. Thus, "according to 3.4.6" refers to Chapter 3, Section 4, Subsection 6. Within the same chapter, statements are numbered continuously using parentheses. For example, "see (3.5)" refers to the fifth numbered equation in Chapter 3.

Contents

Part I

Probability Theory

Chapter 1

Principle of Combinatorics

1.1 Power set and rules

We start with two crucial rules.

1.1.1 Product rule

Suppose that a procedure can be broken down into a sequence of two tasks. If there are n_1 ways to do the first task and for each of these ways of doing the first task, there are n_2 ways to do the second task, and so on for the e n_k-th task then there are

$$n_1 \times n_2 \times \ldots \times n_k \text{ possibilities.}$$

Example: The number of license plates of three letters (a to z) followed by three numbers (0 to 9) is $26^3 \times 10$.

Example: Choosing one-by-one with replacement r balls from an urn containing n numbered balls gives

$$n^r \text{ possibilities.}$$

1.1.2 Sum rule

If a task can be done either in one of n_i, $i = 1, \ldots k$ ways, where none of the set of n_i ways is the same as any of the set of n_j ways $(\leq i \neq j \leq k)$, then there are

$$n_1 + n_2 + \ldots n_k \text{ possibilities.}$$

Example: The number of passwords that must be six to eight characters long, where each character is an uppercase letter or a digit, and must contain at least one digit: Let P6, P7, and P8 be the numbers of passwords of length 6, 7, and 8, respectively. Then the total number of passwords is

$$P6 + P7 + P8 = (36^6 - 26^6) + (36^7 - 26^7) + (36^8 - 26^8) = 2,684,483,063,360.$$

1.1.3 Power set and its cardinal

A set Ω is finite if it contains, say, n distinct elements $\omega_1, \ldots, \omega_n$. If n is infinite, we say that Ω is countable. Ω is *at most countable* if it is finite or countable. The cardinal n of Ω is often denoted $|\Omega|$. In passing, note that

$$\text{counting the cardinal of set } \Omega \Longleftrightarrow \text{finding in how many ways one can form distinct elements of } \Omega. \tag{1.11}$$

The power set $\mathcal{P}(\Omega)$ is the set of all the subsets of Ω. Its cardinal is

$$|\mathcal{P}(\Omega)| = 2^{|\Omega|}. \tag{1.12}$$

For instance, the number of subsets of $\Omega = a, b, c, d$ is $16 = 2^4$. To show (1.12), it suffices to understand that (1.11) allows us to show that

$$\begin{aligned} \text{forming a distinct element of } \mathcal{P}(\Omega) &\iff \text{forming a subset of } \Omega \\ &\iff \text{for each element } \omega_j \text{ of } \Omega, \text{ one of two choices:} \\ &\qquad \text{to keep it in the subset or not,} \end{aligned}$$

and one concludes with the product rule $2^n = 2 \times 2 \times \ldots \times 2$ (n times) is the number of choices for forming a distinct element of $\mathcal{P}(\Omega)$.

1.2 Permutations

(1) The **factorial** of a number n, denoted by $n!$, is defined as:

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1, \quad \text{for } n \geq 1,$$

with $0! = 1$ by convention.

(2) A permutation of a set of distinct objects is any rearrangement of them (ordered list). Let $r \leq n$. An r-permutation of a set of n distinct objects is any permutation of any r of these n objects. Making an r-permutation is equivalent to one of the following:

(i) choosing a subset of size r from n distinct objects one-by-one without replacement (i.e., after the first object is chosen, the next object is chosen from the remaining $n-1$, the next after that from the remaining $n-2$, etc.);

(ii) choosing one-by-one without replacement r balls from an urn containing n numbered balls;

(iii) dispatching r numbered black balls into n numbered boxes, each box contains at most 1 object.

The number of choices of r-permutation is given by the combinatorial number

$$P(n, r) := \frac{n!}{(n-r)!} = n(n-1)\ldots(n-r+1),$$

sometimes denoted ${}^nP_r$ or $P_{n,r}$.

Example 1.2.1. The number of ways of choosing an ordered subset of size 2 from the set of 3 letters $\{a, b, c\}$ without replacement is:

$$P(3, 2) = \frac{3!}{(3-2)!} = 6,$$

which are:

$$ab, ac, ba, bc, ca, cb.$$

n-permutations: this is the special case of $r = n$. Given n distinct objects, the number of different ways in which the objects may be ordered (permuted) is

$$P(n, n) = n!.$$

Example: the set of 3 letters $\{a, b, \}$ can be ordered in $3! = 6$ ways:

$$abc, acb, bac, bca, cab, cba.$$

1.3 Combinations

(1) Let $0 \leq r \leq n$. An r-combination of a set of n distinct objects is any (unordered) subset that contains exactly r of these objects. Making an r-combination of a set of n distinct objects is equivalent to one of the following:

(i) choosing an r-permutation and forget the order, i.e. choosing a subset of size r without replacement and without regard to order;

(ii) choosing in one hand r balls from an urn containing n numbered balls ;

(iii) dispatching r black balls into numbered n boxes, each box could contain all balls.

The number of choices of r-combination is given by the binomial coefficient

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

and is read as "*n choose r*." This number is sometimes denoted $C(n, r)$ or ${}^nC_r$ or $C_{n,r}$. Procedure (1) is then a special case with $n_1 = r$ and $n_2 = n - r$.

Remarks: $\binom{n}{0} = \binom{n}{n} = 1$, and that if n is an integer and r is a non-negative integer less than n, then:

$$\binom{n}{r} = \binom{n}{n-r} = \binom{n-1}{r-1} + \binom{n-1}{r}.$$

Example: Using the set $\{a, b, c\}$, the number of ways of choosing a subset of size 2 is:

$$\binom{3}{2} = \frac{3!}{2!\,1!} = 3,$$

which are:

$$\{a, b\}, \{a, c\}, \{b, c\}.$$

Example 1.3.1. A purse contains one quarter, one dime, one nickel, and one penny. Two coins are chosen randomly without regard to order. The total number of ways is:

$$\binom{4}{2} = 6,$$

which is the cardinal of the set:

$$\big\{\{Q,D\},\{Q,N\},\{Q,P\},\{D,N\},\{D,P\},\{N,P\}\big\}.$$

Binomial theorem Note that in the power expansion of $(x+y)^n$, x, $y \in \mathbb{R}$, the coefficient of $x^r y^{n-r}$ is $\binom{n}{r}$:

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}, \tag{1.31}$$

Retrieving (1.12)**:** Let Ω be a set with $|\Omega| = n$. Using the Binomial formula (1.31), one has

$$2^n = \sum_{r=0}^{n} \binom{n}{r} = \text{number of ways to form a subset of } \Omega \text{ with a cardinal at most } n \; = |\mathcal{P}(\Omega)|.$$

Chapter 2

General Probability

2.1 Sample spaces

In probability theory, the fundamental building blocks are *sample spaces*, *sample points*, and *events*. These concepts provide the foundation for understanding random experiments and their outcomes.

2.1.1 Sample point and sample space

A *sample point* is the simplest possible outcome of a random experiment. The *sample space*, usually denoted Ω, is the set of all possible sample points associated with a given experiment. For example, when tossing a six-faced die, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

2.1.2 Mutually exclusive and exhaustive outcomes

Mutually exclusive outcomes: Two or more outcomes are said to be *mutually exclusive* if they cannot occur simultaneously. These are also referred to as *disjoint outcomes*.

Exhaustive outcomes: Outcomes are said to be *exhaustive* if, collectively, they represent the entire sample space. In other words, at least one of these outcomes must occur whenever the experiment is performed.

2.2 Events

An *event* is any collection of sample points or any subset of the sample space. Events are central to the study of probability, as they represent the conditions or outcomes we are interested in.

2.2.1 Operations on events

Union of events: The union of two events A and B, denoted $A \cup B$, consists of all sample points that are in either A or B. This concept extends to multiple events: the union of $A_1, A_2, \ldots, A_n$, denoted

$$\bigcup_{i=1}^{n} A_i \quad \text{or} \quad A_1 \cup A_2 \cup \cdots \cup A_n,$$

consists of all sample points in at least one of the events A_i. For an infinite collection of events, this concept generalizes accordingly.

Intersection of events: The intersection of two events A and B, denoted $A \cap B$, consists of all sample points that are in both A and B. Similarly, the intersection of $A_1, A_2, \ldots, A_n$, denoted

$$\bigcap_{i=1}^{n} A_i \quad \text{or} \quad A_1 \cap A_2 \cap \cdots \cap A_n,$$

consists of all sample points in every one of the events A_i.

Mutually exclusive events: Events $A_1, A_2, \ldots, A_n$ are *mutually exclusive* if they have no sample points in common, or equivalently, if their two by two intersections are empty:

$$A_i \cap A_j = \emptyset, \quad \forall i \neq j.$$

Exhaustive events: Events $A_1, A_2, \ldots, A_n$ are said to be *exhaustive* if their union equals the entire sample space:

$$\bigcup_{i=1}^{n} A_i = \Omega.$$

Complement of an event: The complement of an event A, denoted A^c, is also an event that consists of all sample points in the sample space that are not in A. Formally, $A^c = \Omega \setminus A$.

Subevent: If event B contains all the sample points in event A, then A is a subevent of B, denoted $A \subseteq B$. The occurrence of event A implies that event B has occurred.

Partition with events: Events $E_1, E_2, \ldots, E_n$ form a partition of event A if:

$$A = \bigcup_{i=1}^{n} E_i \quad \text{and} \quad E_i \cap E_j = \emptyset, \ \ \forall\ i \neq j.$$

Indicator function for event A: The function:

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases} \quad \text{where } x \text{ denotes a sample point,}$$

is the indicator function for event A.

Example 2.2.1. Suppose that an "experiment" consists of tossing a six-faced die. The sample space of outcomes consists of the set $\Omega = \{1, 2, 3, 4, 5, 6\}$, each number being a sample point representing the number of spots that can turn up when the die is tossed.

- The outcomes 1 and 2 (or more formally, the events $\{1\}$ and $\{2\}$) are mutually exclusive when tossing a die. The outcomes (sample points) 1 to 6 are exhaustive for the experiment of tossing a die.

- The collection $\{2, 4, 6\}$ represents the event of tossing an even number when tossing a die. If $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$, then:

$$A \cup B = \{1, 2, 3, 4, 6\}, \quad A \cap B = \{2\}.$$

- The events A = "a number less than 4 is tossed" = $\{1, 2, 3\}$ and B = "a 4 is tossed" = $\{4\}$ are mutually exclusive since they have no sample points in common, i.e., $A \cap B = \emptyset$.

• If $A = \{1, 2, 3\}$, then $A^c = \{4, 5, 6\}$.

• If A = "a 2 is tossed" = $\{2\}$ and B = "an even number is tossed" = $\{2, 4, 6\}$, then $A \subseteq B$.

• The events A = "a 2 or 4 is tossed" = $\{2, 4\}$ and B = "a 6 is tossed" = $\{6\}$ form a partition of the event C = "an even number is tossed" = $\{2, 4, 6\}$.

• For the die-tossing experiment, if $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$, then:

$$A^c = \{4, 5, 6\}, \quad B^c = \{1, 3, 5\},$$

and

$$A \cup B = \{1, 2, 3, 4, 6\}, \quad (A \cup B)^c = \{5\}.$$

Thus:

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

2.2.2 Rules for operations on events

Several rules govern operations on events, which form the basis of probability theory:

(a) For any event A,

$$A \cup A^c = \Omega \quad \text{and} \quad A \cap A^c = \varnothing.$$

(b) The set difference $A \setminus B = A \cap B^c$ consists of all sample points in A but not in B. This is sometimes denoted $A \setminus B$.

(c) If $A \subseteq B$, then $A \cup B = B$ and $A \cap B = A$.

(d) **DeMorgan's laws:** For two events A and B,

$$(A \cup B)^c = A^c \cap B^c \quad \text{and} \quad (A \cap B)^c = A^c \cup B^c.$$

The latter generalizes to more than two events.

(e) For any events $A, B_1, B_2, \dots, B_n$,

$$\begin{aligned} A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) &= (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n), \\ A \cup (B_1 \cap B_2 \cap \cdots \cap B_n) &= (A \cup B_1) \cap (A \cup B_2) \cap \cdots \cap (A \cup B_n). \end{aligned}$$

(f) If $B_1, B_2, \dots, B_n$ are exhaustive events (i.e., $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$), then for any event A:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

As a special case, for any events A and B,

$$A = (A \cap B) \cup (A \cap B^c) \Longrightarrow A \cap B \text{ and } A \cap B^c \text{ form a partition of } A.$$

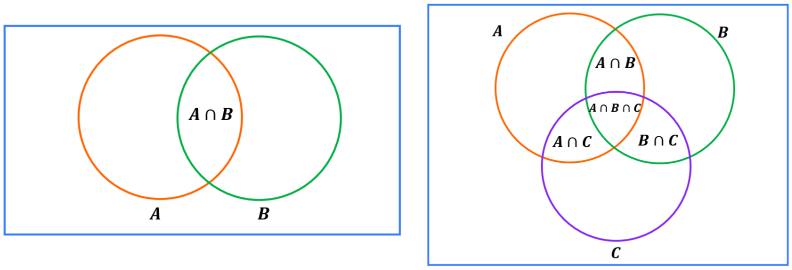


Figure 2.1: Venn diagram for events.

2.3 Probability

2.3.1 Probability function

A *probability function* P assigns a probability to each event E in the sample space Ω. The probability function must satisfy the following axioms:

(i) $0 \leq \mathbb{P}(A) \leq 1$ for every event A,

(ii) $\mathbb{P}(\Omega) = 1$,

(iii) For mutually exclusive events $A_1, A_2, \ldots, A_n \ldots$,

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup \ldots) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

2.3.2 The uniform probability on a finite sample space

If the sample space Ω has a finite number of sample points, say $\Omega = \{\omega_1, \ldots, \omega_n\}$, then the probability of an event A is computed as follows

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}).$$

If furthermore each sample point ω_i is equally likely, then the probability function $\mathbb{P}$ is said to be *uniform*, and

$$\mathbb{P}(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, 2, \ldots, n.$$

Moreover, for any event A, we have

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

Example 2.3.1 (Tossing a fair die)**.** In the experiment of tossing a six-faced die, assume each face has an equal chance of turning up. The sample space is $\Omega = \{1,2,3,4,5,6\}$, and the probability function is uniform:

$$\mathbb{P}(\{\omega\}) = \frac{1}{6}, \quad \omega \in \Omega.$$

The event A of rolling an even number is $A = \{2,4,6\}$. The probability of A is:

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

2.3.3 Additional rules for probability

(i) $\mathbb{P}(\varnothing) = 0$,

(ii) $\mathbb{P}(\Omega) = 1$,

(iii) For any event A, $0 \le \mathbb{P}(A) \le 1$,

(iv) If $A \subseteq B$, then $\mathbb{P}(A) \le \mathbb{P}(B)$.

(v) For any event A, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(vi) For any events A and B, since $\mathbb{P}(A) + \mathbb{P}(B)$ counts $\mathbb{P}(A \cap B)$ twice, then

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c), \\ \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \end{aligned}$$

(vii) For any events A, B and C,

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

The latter extends to more that three events by taking caution to the sign alternation.

(viii) For any events $A_1, A_2, \ldots, A_n$,

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

with equality holding if and only if the events are mutually exclusive.

(ix) **Law of total probability:** For exhaustive events $B_1, B_2, \ldots, B_k$ that are mutually exclusive (forming a partition of the entire sample space), for any event A:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \cdots + \mathbb{P}(A \cap B_k). \tag{2.31}$$

Example 2.3.2. A survey finds that in a city:

$$\begin{aligned} \mathbb{P}(R) &= 0.75 \quad \text{(households with radios)},\\ \mathbb{P}(I) &= 0.65 \quad \text{(households with irons)},\\ \mathbb{P}(T) &= 0.55 \quad \text{(households with toasters)},\\ \mathbb{P}(R \cap I) &= 0.50,\\ \mathbb{P}(R \cap T) &= 0.40,\\ \mathbb{P}(I \cap T) &= 0.30,\\ \mathbb{P}(R \cap I \cap T) &= 0.20. \end{aligned}$$

Find the probability that a household has at least one appliance.
Solution:

$$\mathbb{P}(R \cup I \cup T) = \mathbb{P}(R) + \mathbb{P}(I) + \mathbb{P}(T) - \mathbb{P}(R \cap I) - \mathbb{P}(R \cap T) - \mathbb{P}(I \cap T) + \mathbb{P}(R \cap I \cap T).$$

Substitute the values:

$$\mathbb{P}(R \cup I \cup T) = 0.75 + 0.65 + 0.55 - 0.50 - 0.40 - 0.30 + 0.20 = 0.95.$$

Example 2.3.3. Given $\mathbb{P}(A \cap B) = 0.3$, $\mathbb{P}(A) = 0.6$, $\mathbb{P}(B) = 0.5$, find $\mathbb{P}(A^c \cap B^c)$.
Solution: We have

$$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A \cup B) \text{ and } \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Substituting the values,

$$\mathbb{P}(A \cup B) = 0.6 + 0.5 - 0.3 = 0.8 \Longrightarrow \mathbb{P}(A^c \cap B^c) = 1 - 0.8 = 0.2.$$

2.4 Conditional probability and independence of events

2.4.1 Definition.

If $\mathbb{P}(B) > 0$, the *conditional probability* of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{for an event } A.$$

Note that the function $\mathbb{P}_B = \mathbb{P}(. \mid B)$ is also a probability. Indeed,

(i) For every event A,

$$0 \leq \mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1;$$

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

(iii) For mutually exclusive events $A_1, A_2, \ldots, A_n \ldots$, the events $A_1 \cap B, A_2 \cap B, \ldots, A_n \cap B \ldots$ are also mutually exclusive , then

$$\begin{aligned}\mathbb{P}_B(\cup_i A_i) &= \frac{\mathbb{P}\big(B \cap (\cup_{i=1}^n A_i)\big)}{\mathbb{P}(B)} = \frac{\mathbb{P}\big(\cup_i (B \cap A_i)\big)}{\mathbb{P}(B)} = \sum_i \frac{\mathbb{P}(B \cap A_i)}{\mathbb{P}(B)} = \sum_i \mathbb{P}(A_i \mid B) \\ &= \sum_i \mathbb{P}_B(A_i).\end{aligned}$$

Note that for any events A and B, such that $\mathbb{P}(A)$, $\mathbb{P}(B) > 0$, we have

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

and

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\, \mathbb{P}(B) + \mathbb{P}(A \mid B^c)\, \mathbb{P}(B^c). \tag{2.41}$$

2.4.2 Bayes theorem

Bayes Theorem is a consequence of the law of total probability (2.31): If $B_1, B_2, \ldots, B_n$ form a partition of sample space Ω, then

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\, \mathbb{P}(B_i)}{\sum_{j=1}^n \mathbb{P}(A \mid B_j)\, \mathbb{P}(B_j)}.$$

Example 2.4.1 (Bayes and the sweets)**.** Box A contains 5 dark chocolates and 10 milk chocolates. Box B contains 12 dark chocolates and 12 milk chocolates. Craig picks a box at random and then takes out one chocolate at random. What is the probability that he gets a dark chocolate?
Solution: Let M=“picking milk chocolate”, D=“picking dark chocolate”, A=“picking box A, and B=“picking box B. Assuming the boxes are picked with equal probability, we have:

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(D \mid A) = \frac{5}{15}, \quad \mathbb{P}(D \mid B) = \frac{12}{24}.$$

By formula (2.41), we then have:

$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D \mid A)\mathbb{P}(A) + \mathbb{P}(D \mid B)\mathbb{P}(B) \\ &= \frac{5}{15} \cdot \frac{1}{2} + \frac{12}{24} \cdot \frac{1}{2} = \frac{5}{12}.\end{aligned}$$

2.4.3 Independent events

If events A and B satisfy the relationship

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \text{or} \quad \mathbb{P}(B \mid A) = \mathbb{P}(B),$$

then the events are said to be *independent* or *stochastically independent* or *statistically independent*. The independence of (non-empty) events A and B is equivalent to:

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

2.4.4 Mutually independent events

If events $A_1, A_2, \ldots, A_n$ satisfy the relationship

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\ldots\mathbb{P}(A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i),$$

then they are said to be mutually independent.

2.4.5 Additional rules

(a) If $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \neq 0$, then:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2)\ldots\mathbb{P}(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

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

(c) If $A \subseteq C$, then $\mathbb{P}(A \mid C) = \frac{\mathbb{P}(A)}{\mathbb{P}(C)}$.

(d) If A and B are independent events, then A^c and B are independent events, A and B^c are independent events, and A^c and B^c are independent events.

(e) Since $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\Omega \cap A) = \mathbb{P}(A)$ for any event A, it follows that Ω is independent of any event A.

2.5 Solved problems

Example 2.5.1. Let events A and B be independent. Find the probability, in terms of $\mathbb{P}(A)$ and $\mathbb{P}(B)$, that exactly one of the events A and B occurs.

Solution: Observe that "exactly one of A and B" $= (A \cap B^c) \cup (B \cap A^c)$. Since $A \cap B^c$ and $B \cap A^c$ are mutually exclusive, then

$$\begin{aligned}\mathbb{P}(\text{exactly one of } A \text{ and } B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(B \cap A^c) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).\end{aligned}$$

Now, since A and B are independent, then

$$\mathbb{P}(\text{exactly one of } A \text{ and } B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A)\,\mathbb{P}(B).$$

Example 2.5.2. If

$$\mathbb{P}(A) = \frac{1}{6}, \ \mathbb{P}(B) = \frac{5}{12}, \ \mathbb{P}(A \mid B) + \mathbb{P}(B \mid A) = \frac{7}{10},$$

find $\mathbb{P}(A \cap B)$.

Solution: Let $\mathbb{P}(A \cap B) = x$. We know that

$$\mathbb{P}(A \mid B) + \mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \left(\frac{1}{\mathbb{P}(B)} + \frac{1}{\mathbb{P}(A)}\right)x.$$

Then:

$$\frac{7}{10} = (6 + \frac{12}{5})x = 6 + \frac{42}{5}x \Longrightarrow \mathbb{P}(A \cap B) = x = \frac{35}{420} = \frac{1}{12}.$$

.

Example 2.5.3. Suppose the die-tossing experiment is considered again. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. We define the following events:

- $A = \{$"the number tossed is 1 or 3"$\} = \{1, 3\}$,
- $B = \{$"the number tossed is even"$\} = \{2, 4, 6\}$,
- $C = \{$"the number tossed is a 1 or a 2"$\} = \{1, 2\}$,
- $D = \{$"the number tossed doesn't start with the letters 'f' or 't'"$\} = \{1, 2, 3, 4, 5, 6\}$.

The conditional probability of A given B is:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{2\})}{\mathbb{P}(\{2, 4, 6\})} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Events A and B are not independent, since:

$$\frac{1}{6} = \mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4},$$

or alternatively, events A and B are not independent since $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)$.
Events A and C are not independent since:

$$\mathbb{P}(A \mid C) = 1 \neq \frac{1}{2} = \mathbb{P}(A).$$

Events B and C are independent, since:

$$\mathbb{P}(B \mid C) = \frac{1}{2} = \mathbb{P}(A).$$

(Alternatively, $\mathbb{P}(B \cap C) = \mathbb{P}(\{2\}) = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}(B)\mathbb{P}(C)$).
Similarly, both A and B are independent of D.

Example 2.5.4. Three dice have the following probabilities of throwing a "six": p, q, r respectively. One of the dice is chosen at random and thrown (each is equally likely to be chosen). A "six" appeared. What is the probability that the die chosen was the first one?
Solution: The event "a 6 is thrown" is denoted by "6".

$$\mathbb{P}(\text{die } 1 \mid \text{"6"}) = \frac{\mathbb{P}(\text{"6"} \mid \text{die } 1)\mathbb{P}(\text{die } 1)}{\sum_i \mathbb{P}(\text{"6"} \mid \text{die } i)\mathbb{P}(\text{die } i)} = \frac{p \times \frac{1}{3}}{\mathbb{P}(\text{"6"})}.$$

Since

$$\begin{aligned}\mathbb{P}(\text{"6"}) &= \mathbb{P}(\text{"6"} \mid \text{die } 1)\mathbb{P}(\text{die } 1) + \mathbb{P}(\text{"6"} \mid \text{die } 2)\mathbb{P}(\text{die } 2) + \mathbb{P}(\text{"6"} \mid \text{die } 3)\mathbb{P}(\text{die } 3)\\ &= \frac{p}{3} + \frac{q}{3} + \frac{r}{3} = \frac{p+q+r}{3},\end{aligned}$$

then

$$\mathbb{P}(\text{die } 1 \mid \text{"6"}) = \frac{p}{p+q+r}.$$

Example 2.5.5. Identical twins come from the same egg and hence are of the same sex. Fraternal twins have a 50-50 chance of being the same sex. Among twins, the probability of a fraternal set is p and an identical set is $q = 1 - p$. If the next set of twins are of the same sex, what is the probability that they are identical?
Solution: Let A="the next set of twins are of the same sex", and B="the next sets of twins are identical". We are given:

$$\mathbb{P}(A \mid B) = 1, \quad \mathbb{P}(A \mid B^c) = \frac{1}{2}, \quad \mathbb{P}(B) = q, \quad \mathbb{P}(B^c) = p = 1 - q,$$

and

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Since

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) = q \quad \text{and} \quad \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B^c)\,\mathbb{P}(B^c) = \frac{p}{2},$$

then

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = q + \frac{p}{2} = \frac{1+q}{2},$$

hence

$$\mathbb{P}(B \mid A) = \frac{2q}{1+q}.$$

Remark 2.5.6. In questions involving coin flips or dice tossing, it is understood, unless indicated. Otherwise, successive flips or tosses are independent of one another. In making a random selection of an object from a collection of n objects, it is understood that each object has the same chance of being chosen. In questions involving choosing r objects at random from a total of n objects, or constructing a random permutation of a collection of objects, it is understood that each of the possible choices or permutations is equally likely to occur.

Example 2.5.7. Three people, X, Y, and Z, in order, roll an ordinary die. The first one to roll an even number wins. The game continues until someone rolls an even number. Find the probability that X will win.

Solution: Since X rolls first, fourth, seventh, etc., until the game ends, the probability that X will win is the probability that, in throwing a die, the first even number will occur on the 1st, or 4th, or 7th, or ... throw.

Due to the independence of successive throws, the probability that the first even number occurs on the n-th throw is given by:

$$\mathbb{P}(\text{first even throw on } n\text{-th throw}) = (1-p)^{n-1}p,$$

where $p = \mathbb{P}(\text{throw is even}) = \frac{1}{2}$. Thus, the probability that the first even throw occurs on the 1st, or 4th, or 7th, or ... throw is:

$$\mathbb{P}(\text{first even throw on 1st, 4th, 7th, } \dots) = p + (1-p)^3 p + (1-p)^6 p + \cdots.$$

Substituting $p = \frac{1}{2}$:

$$\mathbb{P}(\text{first even throw on 1st, 4th, 7th, } \dots) = \frac{1}{2} + \left(\frac{1}{2}\right)^3 \frac{1}{2} + \left(\frac{1}{2}\right)^6 \frac{1}{2} + \cdots.$$

This is an infinite geometric series with the first term $a = \frac{1}{2}$ and common ratio $r = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$. The sum of the series is:

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} = \frac{\frac{1}{2}}{1-\frac{1}{8}} = \frac{4}{7}.$$

Thus, the probability that X will win is:

$$\mathbb{P}(X \text{ wins}) = \frac{4}{7}.$$

Example 2.5.8. A calculator has a random number generator button which, when pressed, displays a random digit $0, 1, 2, \dots, 9$. The button is pressed four times. Assuming that the numbers generated are independent of one another, find the probability of obtaining one "0", one "5", and two "9"s in any order.

Solution: There are $10^4 = 10,000$ possible four-digit orderings that can arise, ranging from 0000 to 9999. From the notes on permutations, if we have four digits with one "0", one "5", and two "9"s, the number of orderings is:

$$\frac{4!}{1!\,1!\,2!} = \frac{24}{2} = 12.$$

The probability in question is then:

$$p = \frac{\text{Number of favorable outcomes}}{\text{Total possible outcomes}} = \frac{12}{10,000} = 0.0012.$$

Chapter 3

Univariate Random Variables

3.1 Random variables and probability distributions

A *random variable* (shortly r.v.) is a function on a sample space Ω that assigns a real number $X(\omega)$ to each sample point $\omega \in \Omega$. The set

$$\Omega' := X(\Omega)$$

is seen as new sample space and is called the *state space* of X. The set $(X \in B)$ denotes

$$\{\omega,\ s.t.\ X(\omega) \in B\}$$

which is an event in Ω'. The *probability distribution* of X, denoted $\mathbb{P}_X$ is a probability on Ω' which assigns to each event $B \subset \Omega'$, the value $\mathbb{P}(X \in B)$:

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B) = \mathbb{P}\big(\{\omega,\ s.t. X(\omega) \in B\}\big).$$

3.1.1 Independence of random variables

Two X and Y are said to be independent, and we denote $X \perp\!\!\!\perp Y$, if

$$\mathbb{P}(X \in B,\ Y \in C) = \mathbb{P}(X \in B)\,\mathbb{P}(Y \in C),$$

for $B \subset X(\omega)$ and $C \subset Y(\omega)$. If furthermore X and Y identical probability distributions, we shortly denote X and Y are i.i.d

Often, a random variable is simply equal to the sample point ω, if the sample points are numerical values. For example, the sample space representing the number of spots that turn up when an ordinary die is tossed, the random variable X , which describes the number of spots that turn up are

$$\Omega = \{1,2,3,4,5,6\}, \text{ and } X(\omega) = \omega.$$

3.1.2 Some clarification

A random variable is sometimes described in terms of the outcome of a random experiment (such as tossing a die). It could be also described without explicit reference to the underlying random experiment or sample space (such as the prime rate of interest two years from now). For instance, suppose that a gamble based on the outcome of the toss X of a die pays \$10 if an even number is tossed and pays \$20 if an odd number is tossed. The payoff can be represented by the random variable Y, where:

$$Y = \begin{cases} 10, & \text{if } X \text{ is even,} \\ 20, & \text{if } X \text{ is odd.} \end{cases}$$

Thus,

$$\mathbb{P}(Y \geq 12) = \mathbb{P}(X \in \{1,3,5\}).$$

Note that for a fair die, this probability is $\frac{1}{2}$.

3.2 Discrete random variables and their distribution

The random variable X is *discrete* and is said to have a discrete distribution if it can take on values only from a finite or countable infinite sequence (usually the integers or some subset of the integers).

As an example, consider the following two random variables related to successive tosses of a coin:

- $X = 0$ if the first head occurs on an even-numbered toss, $X = 1$ if the first head occurs on an odd-numbered toss;
- $Y = n$, where n is the number of the toss on which the first head occurs.

Both X and Y are discrete random variables, where X can take on only the values 0 or 1, and Y can take on any positive integer value. Both X and Y are based on the same sample space, the sample points are sequences of tail coin flips ending with a head coin flip:

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\}.$$

For example:

$$\begin{aligned} X(H) &= 0\,(\text{a head on the second flip, an odd-numbered flip}),\ X(TH) = 1, \ldots \\ Y(H) &= 1\,(\text{first head on flip 1}),\ Y(TH) = 2,\ Y(TTH) = 3, \ldots. \end{aligned}$$

The *probability mass function* (p.m.f.) of a discrete random variable X often denoted by $p(x)$ (or by $p_X(x)$ if several random variables are involved) taking is

$$p(x) = \mathbb{P}(X = x); \;\; x \in X(\Omega) = \{x_1, x_2 \ldots\}.$$

It must satisfy:

$$p(x_n) \geq 0 \quad \text{and} \quad \sum_n p(x_n) = 1.$$

Given a set A of real numbers,

$$\mathbb{P}(X \in A) = \sum_{i,\ s.t.\ x_i \in A} p(x_i).$$

3.3 Continuous random variables and their distribution

A *continuous random variable* usually can assume numerical values from an interval of real numbers, perhaps the whole set of real numbers $\mathbb{R}$. For example, the length of time between successive streetcar arrivals at a particular stop could be regarded as a continuous random variable (assuming time measurement can be made perfectly accurate).

A continuous random variable X u has a *probability density function* (p.d.f.), denoted $f(x)$ (or $f_X(x)$ if several random variables are involved) satisfying

(i) $f(x) \geq 0$ for all x;

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Probabilities related to X are found by integrating:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx, \quad -\infty < a < b < \infty.$$

We emphasize that f is continuous except at a finite number of points, then probabilities are defined and calculated as if f was continuous everywhere (the discontinuities are ignored).

A generic example for a p.d.f. is when X is *uniformly* distributed on the interval $(0,1)$, i.e. X has the density function

$$f(x) = 1 \text{ for } 0 < x \leq 1, \quad f(x) = 0 \text{ otherwise.}$$

3.4 Cumulative distribution function (CDF) and survival function

Given a random variable X, the *cumulative distribution function* (also called the distribution function or c.d.f.) of X is defined as:

$$F(x) = \mathbb{P}(X \leq x),$$

which can also be denoted as $F_X(x)$ if several random variables are involved. The *survival function* is the complement of the cumulative distribution function and is defined as:

$$S(x) = 1 - F(x) = \mathbb{P}(X > x).$$

The event $(X > x) = \{\omega,\ s.t. X(\omega > x)\}$ is referred to as a "tail" of the distribution. Also, for a continuous random variable, the hazard rate is defined as:

$$h(x) = \frac{f(x)}{S(x)}.$$

3.5 Key properties of the CDF

- The c.d.f. $F(x)$ is a non-decreasing function such that

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

- $\mathbb{P}(X = a) = F(a) - F(a_-) = F(a) - \lim_{x \to a_-} F(x)$.
- For a discrete random variable, $F(x)$ has a step increase at each point with non-zero probability mass and remains constant between jumps.
- For a continuous random variable with density function $f_X(x)$, the c.d.f. is continuous, differentiable, and satisfies:
$$F(x) = \int_{-\infty}^{x} f(t)\, dt.$$
- At points of non-zero probability mass for a mixed distribution, $F(x)$ will have a jump (see Section ??).

3.6 Examples of distribution functions

1. Discrete Random Variable: Let X be the number turning up when tossing one fair die. The probability function is:

$$\mathbb{P}(X = x) = \begin{cases} \frac{1}{6}, & x \in \{1, 2, 3, 4, 5, 6\}, \\ 0, & \text{otherwise.} \end{cases}$$

The c.d.f. is given by:

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

2. Continuous Random Variable: Let Y be a continuous random variable on the interval $[0, 1]$ with density function:

$$f_Y(x) = \begin{cases} 2x, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

The c.d.f. is:

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

3.7 Some results and related formulas

(i) For a continuous random variable X,

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

when calculating the probability for a continuous random variable, whether or not the endpoints are included.

$$\mathbb{P}(X = a) = 0$$

Non-zero probabilities only exist over an interval, not at a single point. More generally the non-zero set of the p.d.f. $S = \{x \; s.t. f(x) \neq 0\}$ corresponds to the state space of X:

$$X(\Omega) \equiv S.$$

(ii) If X has a mixed distribution, then $\mathbb{P}(X = a)$ will be non-zero for some value(s) of a, and $\mathbb{P}(a \le X \le b)$ will not always be equal to $\mathbb{P}(a < X < b)$ (they will not be equal if X has a nonzero probability mass at either a or b).

(iii) The p.d.f. may be defined piecewise, meaning that $f(x)$ is defined by a different algebraic formula on different intervals.

(iv) A continuous random variable may have two or more different, but equivalent p.d.f.'s, but the difference in the p.d.f.'s would only occur at a finite (or countably infinite) number of points. The c.d.f. of a random variable of any type is always unique to that random variable.

3.8 Solved problems

Example 3.8.1. A die is loaded in such a way that the probability of the face with k dots turning up is proportional to k for $k = 1, 2, 3, 4, 5, 6$. What is the probability, in one roll of the die, that an even number of dots will turn up?

Solution Let X denote the random variable representing the number of dots that appears when the die is rolled once. Then,

$$\mathbb{P}(X = k) = c \cdot k, \quad \text{for } k = 1, 2, 3, 4, 5, 6,$$

where c is the proportionality constant. Since the sum of all probabilities must equal 1, we have:

$$c \cdot (1 + 2 + 3 + 4 + 5 + 6) = 1 \implies c = \frac{1}{21}.$$

$$\begin{aligned} \mathbb{P}(\text{even number of dots turns up}) &= \mathbb{P}(X = 2) + \mathbb{P}(X = 4) + \mathbb{P}(X = 6) \\ &= \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21} = \frac{4}{7}. \end{aligned}$$

Example 3.8.2. An ordinary single die is tossed repeatedly until the first even number turns up. The random variable X is defined to be the number of the toss on which the first even number turns up. Find the probability that X is an even number.
Solution: X is a discrete random variable that can take on an integer value of 1 or more, more precisely, the distribution is given by

$$\mathbb{P}(X = n) = \left(\frac{1}{2}\right)^{n-1} \times \frac{1}{2} = \left(\frac{1}{2}\right)^{n}, \quad n = 1, 2, \ldots$$

(this is the probability of $n-1$ odd tosses followed by an even toss). The probability $\mathbb{P}(X \text{ is even})$ is then

$$\mathbb{P}(X \text{ is even}) = \sum_{k=1}^{\infty} \mathbb{P}(X = 2k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k} = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^{k} = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}.$$

Example 3.8.3. The continuous random variable X has density function:

$$f_X(x) = \begin{cases} k \cdot x, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find $\mathbb{P}(0 \leq X \leq \frac{1}{2})$.
Solution: First, determine the value of k by ensuring the total probability is 1:

$$\int_0^1 k \cdot x \, dx = 1 \implies \frac{k}{2} = 1 \implies k = 2.$$

Thus, the density function is $f_X(x) = 2x$ for $0 \leq x \leq 1$, and

$$\mathbb{P}\left(0 \leq X \leq \frac{1}{2}\right) = \int_0^{1/2} 2x \, dx = \left[x^2\right]_0^{1/2} = \left(\frac{1}{2}\right)^2 - 0 = \frac{1}{4}.$$

Example 3.8.4. Suppose that the continuous random variable X has the cumulative distribution function:

$$F_X(x) = \frac{1}{1 + e^{-x}}, \quad x \in \mathbb{R}.$$

Find X's density function.
Solution: The density function for a continuous random variable is the first derivative of the cumulative distribution function:

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Example 3.8.5. X is a random variable for which:

$$\mathbb{P}(X \le x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

Which of the following statements is true?
A) $\mathbb{P}(X = 1) = 0$ and $\mathbb{P}(X \le 1) = 1$
B) $\mathbb{P}(X = 1) = 1$ and $\mathbb{P}(X \le 1) = 1$
C) $\mathbb{P}(X = 1) = 0$ and $\mathbb{P}(X < 1) = 1$
D) $\mathbb{P}(X = 1) = 1$ and $\mathbb{P}(X < 1) = 0$
E) $\mathbb{P}(X = 1) = 0$ and $\mathbb{P}(X < 1) = 0$.
Solution: Since $F_X(x)$ is continuous, we know $\mathbb{P}(X = 1) = 0$. Thus,

$$\mathbb{P}(X \le 1) = \mathbb{P}(X < 1) + \mathbb{P}(X = 1) = \mathbb{P}(X < 1) = 1.$$

The correct answer is: A.

3.9 Expected value

For a bounded or is a nonnegative random variable X, the expected value is denoted

$$\mathbb{E}[X] \text{ or } \mu_X \text{ or if there is no ambiguity } \mu,$$

and is computed as

$$\mathbb{E}[X] = \begin{cases} \sum_n x_n \, \mathbb{P}(X = x_n), & \text{if } X \text{ is discrete, } X(\Omega) = \{x_1, x_2 \dots\} \\ \int_I x \, f_X(x) \, dx, & \text{if } X \text{ is continuous, } X(\Omega) - I \subset \mathbb{R}. \end{cases} \tag{3.91}$$

The expected value of X is also called the expectation of X, or the mean of X. The expected value is the "average" over the range of values that X can be, or the "center" of the distribution. If X is non bounded or is nonnegative but

$$\mathbb{E}[|X|] \text{ is finite} \Longrightarrow \mathbb{E}[X] \text{ has a meaning}.$$

3.9.1 Expectation of $h(X)$

If h is a bounded or is a nonnegative function, then formula (3.91) becomes

$$\mathbb{E}[h(X)] = \begin{cases} \sum_n h(x_n) \, \mathbb{P}(X = x_n), & \text{if } X \text{ is discrete,} \\ \int h(x) \, f_X(x) \, dx, & \text{if } X \text{ is continuous,} \end{cases}$$

3.9.2 Moments and variance

If k is an integer and $\mathbb{E}[|X|^k] < \infty$, then the k-th moment of X and the k-th central moment of X about the mean μ are:

$$\mathbb{E}[X^k] \text{ and } \mu_k = \mathbb{E}[(X-\mu)^k].$$

The *variance* of X is

$$\sigma_X^2 = \mu_2 = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

and the *standard variation* is σ_X.

3.9.3 Moment generating function

The moment generating function (MGF) of a random variable X is defined as:

$$M_X(t) = \mathbb{E}[e^{tX}] = \begin{cases} \sum e^{tx} f_X(x) & \text{(discrete case)}, \\ \int e^{tx} f_X(x)\, dx & \text{(continuous case)}. \end{cases}$$

It is always true that $M_X(0) = 1$. The moment generating function of X might not exist for all real numbers but usually exists on some interval of real numbers. The function $\ln[M_X(t)]$ is called the *cumulant generating function*.

If $Y = aX + b,\ a, b \in \mathbb{R}$, then:

$$M_Y(t) = e^{bt} \cdot M_X(at).$$

If $X \perp\!\!\!\perp Y$, then

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}\ e^{tY}] = \mathbb{E}[e^{tX}]\,\mathbb{E}[e^{tY}] = M_X(t)\ M_Y(t),$$

see

3.9.4 Percentiles of a distribution

If $0 < p < 1$, then the $100p$-th percentile of the distribution of X is the number η that satisfies both of the following inequalities:

$$\mathbb{P}(X \le \eta_p) \ge p \quad \text{and} \quad \mathbb{P}(X \ge \eta_p) \ge 1 - p.$$

In the continuous case, the latter reduces to $\mathbb{P}(X \le \eta_p) = p$. The 50-th percentile of a distribution is referred to as the *median* of the distribution. For a continuous random variable, it is sufficient to find η_p such that

$$\mathbb{P}(X \le \eta_p) = p.$$

3.9.5 The mode of a distribution

The mode is any point m at which the p.m.f. $p_X(x)$ or the p.d.f. $f_X(x)$ is maximized.

3.9.6 The skewness of a distribution

If the mean of the random variable X is μ and the variance is σ^2, then the *skewness* is defined as:

$$\frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3}.$$

3.9.7 The covariance of two random variables

Let X_1 and X_2 be two r.v.'s such that $\mathbb{E}[X_1^2]$, $\mathbb{E}[X_2^2]$ exist (hence by Jensen inequality in Subsection **??**, $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$ also exist). The *covariance* of X_1 and X_2 is given by

$$cov(X_1, X_2) = \mathbb{E}\big[(X_1 - \mathbb{E}[X_1])(X_1 - \mathbb{E}[X_2])\big] = \mathbb{E}[X_1\, X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2].$$

3.10 Some results and formulas related to the expectation

3.10.1 Existence of mean and variance

The mean of a random variable X might not exist; or might be $+\infty$ or $-\infty$, hence the variance of X might be $+\infty$. For example, the continuous random variable X with the probability density function (p.d.f.):

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

has an expected value:

$$\int_1^\infty x \cdot \frac{1}{x^2}\, dx = +\infty.$$

3.10.2 Linear transformations and expectations

For any constants a_1, a_2 the following hold"

(i) Linearity: if $\mathbb{E}[X_1]$, $\mathbb{E}[X_2]$ exist, then

$$\mathbb{E}[a_1\, X_1 + a_2\, X_2] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2].$$

In particular, for real numbers a_1, a_2,

$$\mathbb{E}[a_1\, X_1 + a_2] = a_1\mathbb{E}[X_1] + a_2.$$

(ii) Quadraticity property: If $\mathbb{E}[X_1^2]$, $\mathbb{E}[X_2^2]$ exists then:

$$V[a_1\, X_1 + a_2\, X_2] = a_1^2\, V[X_1] + a_2^2\, V[X_2] + 2\, a_1\, a_2\, cov(X_1, X_2).$$

3.10.3 The MGF and the moments

If the moment generating function $M_X(t)$ exists in an interval containing $t = 0$, then:

$$M'_X(0) = \mathbb{E}[X] \quad \text{and} \quad M''_X(0) = \mathbb{E}[X^2] \Longrightarrow \text{Var}[X] = M''_X(0) - M'_X(0)^2.$$

If the m.g.f. of $|X|$ is well defined, then the Taylor series expansion of $M_X(t)$ about $t = 0$ is:

$$M_X(t) = 1 + \mathbb{E}[X]\, t + \mathbb{E}[X^2]\, \frac{t^2}{2!} + \mathbb{E}[X^3]\, \frac{t^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]\, t^k}{k!}.$$

If X and Y are random variables, and $M_X(t) = M_Y(t)$ for all values of t in an interval containing 0, then X and Y have identical probability distributions.

3.10.4 Percentiles and median

The median (50th percentile) and other percentiles of a distribution are not always unique. For example, if X is the discrete random variable with:

$$p(x) = 0.25 \quad \text{for } x = 1, 2, 3, 4,$$

then the median of X could be any value from 2 to 3. The convention is to use the midpoint $M = 2.5$.

3.10.5 Standardization

If $\mathbb{E}[X] = \mu$ and $\text{Var}[X] = \sigma^2$, and $Z = \frac{X-\mu}{\sigma}$, then:

$$\mathbb{E}[Z] = 0 \quad \text{and} \quad \text{Var}[Z] = 1.$$

3.11 Independence of two random variables revisited

Random variables X and Y are said to be independent (or stochastically independent) if the cumulative distribution function of the joint distribution $F(x, y)$ can be factored as:

$$F(x, y) = F_X(x) \cdot F_X(y), \quad \text{for all } (x, y).$$

This definition can be extended to a multivariate distribution of more than two variables. If X and Y are independent, then:

$$p(k, l) = p_X(x) \cdot p_X(y) \text{ (discrete case) and } f(x, y) = f_X(x) f_X(y) \text{ (discrete case).}$$

More generally, for any nonnegative bounded functions g and h,

$$\mathbb{E}[g(X) \cdot h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)].$$

In particular, in case of first moments,

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Note that X and Y are independent $\Longrightarrow \text{Cov}[X, Y] = 0$. The converse is untrue in general.

3.12 Solved problems

Example 3.12.1. A fair die is tossed until the first "1" appears. You receive $\frac{1}{2^X}$ dollars if the "1" appears on the X-th toss. Find the expected amount of the payout P you will receive.
Solution: X is a discrete random variable with distribution

$$\mathbb{P}(X = n) = \frac{1}{6}\left(\frac{5}{6}\right)^{n-1}, \quad n = 1, 2, \ldots$$

Using the relation $P = \frac{1}{2^X}$, the expected payout is:

$$\mathbb{E}[P] = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{1}{6}\left(\frac{5}{6}\right)^{n-1} = \frac{1}{12} \sum_{k=0}^{\infty} \left(\frac{5}{12}\right)^k = \frac{1}{12} \frac{1}{1 - \frac{5}{12}} = \frac{1}{7}.$$

Example 3.12.2. Given $\lambda > 0$ the p.d.f.

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

and 0 elsewhere, find the n-th moment of X.
Solution: The n-th moment is obtained by integration by parts repeatedly:

$$\mathbb{E}[X^n] = \int_0^{\infty} x^n\, \lambda\, e^{-\lambda x}\, dx = \frac{n!}{\lambda^n}.$$

Example 3.12.3. A continuous random variable X has density function

$$\begin{cases} 1 - |x|, & \text{if } -1 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $\text{Var}[X]$.
Solution: The density of X is symmetric about 0 (since $f(x) = f(-x)$), so that $\mathbb{E}[X] = 0$. This can be verified directly, but doers not worth it:

$$\mathbb{E}[X] = \int_{-1}^{1} x(1 - |x|)\, dx = \int_{-1}^{0} x(1 + x)\, dx + \int_0^1 x(1 - x)\, dx = -\frac{1}{6} + \frac{1}{6} = 0.$$

Then,

$$\text{Var}[X] \;=\; \mathbb{E}[X^2] = \int_{-1}^{1} x^2(1 - |x|)\, dx = 2\int_0^1 x^2(1 - x)\, dx = \frac{1}{6}.$$

Example 3.12.4. The moment generating function of a r.v. X is given by:

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda,$$

where $\lambda > 0$. Find Var$[X]$.
Solution: First, compute $\mathbb{E}[X]$:

$$\mathbb{E}[X] = M'_X(0) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right)\bigg|_{t=0} = \frac{\lambda}{(\lambda - t)^2}\bigg|_{t=0} = \frac{1}{\lambda}.$$

Next, compute $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = M''_X(0) = \frac{d^2}{dt^2}\left(\frac{\lambda}{\lambda - t}\right)\bigg|_{t=0} = \frac{2\lambda}{(\lambda - t)^3}\bigg|_{t=0} = \frac{2}{\lambda^2}.$$

Thus:

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

Example 3.12.5. A continuous random variable X has p.d.f.:

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

Find the 87.5-th percentile of the distribution.
Solution: The 87.5-th percentile is the value $\eta = \eta_{0.875}$ for which:

$$0.875 = \mathbb{P}(X \le \eta) = \int_{-\infty}^{\eta} f(x)\,dx = \int_{-\infty}^{\eta} \frac{1}{2}e^{-|x|}\,dx.$$

Note that the distribution is symmetric about 0, as $f(-x) = f(x)$. Therefore, the mean and median are both 0. Since $\eta > 0$, we can rewrite the integral:

$$0.875 = \int_{-\infty}^{\eta} \frac{1}{2}e^{-|x|}\,dx \quad = \quad \int_{-\infty}^{0} \frac{1}{2}e^{-|x|}\,dx + \int_{0}^{\eta} \frac{1}{2}e^{-|x|}\,dx = \frac{1}{2} + [\frac{-e^{-x}}{2}]_0^{\eta} = \frac{1}{2} + \frac{1 - e^{-\eta}}{2}.$$

Thus, $e^{-\eta} = 0.25 \Longleftrightarrow = \eta \ln(4)$.

3.13 Frequently used discrete distributions

3.13.1 Discrete uniform distribution $\mathcal{U}(1, 2, \ldots, N)$

We denote $X \overset{\text{d}}{=} \mathcal{U}(1, 2, \ldots, N)$ if the distribution of X is

$$\mathbb{P}(X = k) = \frac{1}{N} \quad \text{for } x = 1, 2, \ldots, N, \quad \text{and } p(k) = 0 \text{ otherwise.}$$

$$\mathbb{E}[X] = \frac{N+1}{2}, \quad \text{Var}(X) = \frac{N^2 - 1}{12}, \quad M_X(t) = \sum_{x=1}^{N} \frac{e^{tx}}{N} = \frac{e^t(e^{Nt} - 1)}{N(e^t - 1)}$$

for any real t.

3.13.2 Binomial distribution $\mathcal{B}(n,p)$

A single trial of an experiment results in either success with probability $p \in (0,1)$, or failure with probability $1-p=q$. We say that

$$X = \begin{cases} 1, & \text{if the trial is a success,} \\ 0, & \text{if the trial is a failure.} \end{cases}$$

follows a Bernoulli distribution with parameter $p \in [0,1]$, denoted $X\mathcal{B}(p)$, if

$$\mathbb{P}(X=1) = p \quad \text{and} \quad \mathbb{P}(X=0) = 1-p.$$

If n independent trials of the experiment are performed, and S is the number of successes that occur, then S_n is an integer between 0 and n. S is said to have a binomial distribution with parameters n and p and we denote $S \overset{\text{d}}{=} \mathcal{B}(n,p)$, if

$$\mathbb{P}(S=k) = \binom{n}{k} p^k \, (1-p)^{n-k} \quad \text{for } k = 0,1,2,\dots,n,$$

$$\mathbb{E}[S] = np, \quad \text{Var}(S) = np(1-p), \quad M_S(t) = (1-p+pe^t)^n.$$

In the special case of $n=1$ (a single trial), the distribution is referred to as a Bernoulli distribution. If $S \overset{\text{d}}{=} B(n,p)$, then S is the sum of n independent random variables each with distribution $\mathcal{B}(p)$. More generally, from the MGF form, we see that if $X \perp\!\!\!\perp Y$, $X \overset{\text{d}}{=} \mathcal{B}(n,p)$ and $Y \overset{\text{d}}{=} \mathcal{B}(m,p)$, then

$$X+Y \overset{\text{d}}{=} \mathcal{B}(n+m,p).$$

The Binomial distribution generalizes the multinomial distribution which will be seen in Example 4.7.1.

3.13.3 Poisson distribution $\mathcal{P}(\lambda)$

Let $\lambda > 0$. We denote $X \overset{\text{d}}{=} \text{Poisson}(\lambda)$ if

$$\mathbb{P}(X=k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{for } k = 0,1,2,3,\dots,$$

$$\mathbb{E}[X] = \text{Var}(X) = \lambda, \quad M_X(t) = e^{\lambda(e^t-1)}, \; t \in \mathbb{R}.$$

From the MGF form, we see that if $X \perp\!\!\!\perp Y$, $X \overset{\text{d}}{=} \mathcal{P}(\lambda)$ and $Y \overset{\text{d}}{=} \mathcal{P}(\mu)$, then

$$X+Y \overset{\text{d}}{=} \mathcal{P}(\lambda+\mu).$$

The Poisson distribution is often used as a model for counting the number of events of a certain type that occur in a certain period of time. Suppose that X represents the number of customers arriving for service at a bank in a 1-hour period, and that a model for X is the Poisson distribution with parameter λ. Under some reasonable assumptions (such as independence of the numbers arriving in different time intervals), it is possible to show that the number arriving in any time period also has a Poisson distribution with the appropriate parameter that is "scaled" from λ. Suppose that $\lambda = 40$ - meaning that X, the number of bank customers arriving in one hour, has a mean of 40. If Y represents the number of customers arriving in 2 hours, then Y has a Poisson distribution with a parameter of 80 — for any time interval of length t, the number of customers arriving in that time interval has a Poisson distribution with parameter $\lambda t = 40t$ — so the number of customers arriving during a 15-minute period ($t = \frac{1}{4}$ hour) will have a Poisson distribution with parameter $40 \times \frac{1}{4} = 10$.

3.13.4 Geometric distribution $\mathcal{G}(p)$

A single trial of an experiment results in either success with probability $p \in (0,1)$, or failure with probability $1 - p = q$. The experiment is performed with successive independent trials until the first success occurs. If X represents the number of failures until the first success, then X is a discrete random variable that can be 0, 1, 2, 3, ... X is said to have a geometric distribution with parameter p and we denote $X \stackrel{\mathrm{d}}{=} \mathcal{G}(p)$, if

$$\mathbb{P}(X = k) = p\,(1-p)^k, \quad \text{for } k = 0, 1, 2, 3, \ldots,$$

$$\mathbb{E}[X] = \frac{1-p}{p} = \frac{q}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{q}{p^2}, \quad M_X(t) = \frac{pe^t}{1-(1-p)e^t} \quad \text{for } t < -\ln(1-p).$$

The geometric distribution has the lack of memory property:

$$\mathbb{P}(X \geq n + k \mid X \geq n) = \mathbb{P}(X \geq k).$$

Another version of a geometric distribution is the random variable Y, the number of the experiment on which the first success occurs; $Y = X + 1$ and

$$\mathbb{P}(Y = k) = \mathbb{P}(X = k - 1), \;\; k \geq 1.$$

3.13.5 Negative binomial distribution $\mathcal{NB}(r, p)$

Let r is an integer and $p \in (0,1)$. We denote $X \stackrel{\mathrm{d}}{=} \mathcal{NB}(r,p)$ if the probability mass function of X is

$$\mathbb{P}(X = k) = \binom{k+r-1}{r-1} p^r\,(1-p)^k \quad \text{for } k = 0, 1, 2, 3, \ldots,$$

$$\mathbb{E}[X] = \frac{r(1-p)}{p}, \quad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}, \quad M_X(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r, \quad \text{for } t < -\ln(1-p).$$

The negative binomial random variable X can be interpreted as being the number of failures until the r-th success occurs when successive trials of an experiment are performed for which the probability of success in a single particular trial is p. As the geometric distribution, the Pascal distribution denoted $\mathcal{P}_a(r,p)$ corresponds to the r.v.

$$Y = r + X$$

which counts the number of trials needed to see the r-th success, that is, Y is the first trial that gives r-successes. Its p.m.f. is

$$\mathbb{P}(Y=k) = \binom{k-1}{r-1} p^r \, (1-p)^{k-r} \quad \text{for } k = r, r+1, r+2, r+3, \ldots$$

Note that

$$\mathcal{P}_a(1,p) = \mathcal{G}(p)$$

and that

$$M_Y(t) = \mathbb{E}[e^{t(r+X)}] = e^{tr} \, M_X(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r, \quad \text{for } t < -\ln(1-p),$$

and that the distributions $\mathcal{NB}(r,p)$ and $\mathcal{P}_a(r,p)$ are defined even if r is not an integer. From the MGF form, we see that if $X \perp\!\!\!\perp Y$, $X \overset{\text{d}}{=} \mathcal{P}_a(r,p)$ and Y$\overset{\text{d}}{=} \mathcal{P}_a(s,p)$, then

$$X + Y \overset{\text{d}}{=} \mathcal{P}_a(r+,p).$$

This distribution is discussed later in these notes.

3.13.6 Hypergeometric distribution $\mathcal{H}(N,n,K)$

Consider a group of N objects, K of which are of Type I and $M-K$ are of Type II. If n objects are randomly chosen without replacement from the group of N, let X denote the number of Type I objects in the group of n. X is a non-negative integer that satisfies:

$$\max[0, n-(N-K)] \leq X \leq \min[n,K],$$

and has a hypergeometric distribution:

$$X \overset{\text{d}}{=} \mathcal{H}(N,n,K).$$

The p.m.f. is given by:

$$\mathbb{P}(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \qquad \text{for } k = \max[0, n-(N-K)] \text{ to } x = \min[n,K].$$

The expected value and variance are:

$$\mathbb{E}[X] = \frac{nK}{M}, \quad \text{Var}[X] = \frac{nK(N-K)}{N^2(N-1)}.$$

Example 3.13.1. Let X be a discrete random variable that is uniformly distributed on the even integers $x = 0, 2, 4, \ldots, 22$, so that the p.m.f of X is:

$$\mathbb{P}(X = x) = \frac{1}{12}, \quad \text{for each even integer } x \text{ from 0 to 22.}$$

Find $\mathbb{E}[X]$ and $\text{Var}[X]$.
Solution: The discrete uniform distribution described earlier in the notes is on the points $x = 1, 2, \ldots, N$. If we consider the transformation:

$$Y = 1 + \frac{X}{2}, \tag{3.131}$$

then the random variable Y is distributed on the points $Y = 1, 2, \ldots, 12$, with p.m.f:

$$\mathbb{P}(Y = y) = \frac{1}{12}, \quad \text{for } y = 1, 2, \ldots 12.$$

Thus, Y has the discrete uniform distribution described earlier in Subsection 3.13.1, and:

$$\mathbb{E}[Y] = \frac{1+12}{2} = 6.5, \quad \text{Var}[Y] = \frac{12^2-1}{12} = \frac{143}{12}.$$

Using (3.131), we use the rules for expectation and variance to get:

$$\mathbb{E}[X] = 2 \cdot \mathbb{E}[Y] - 2 = 11, \quad \text{Var}[X] = 4 \cdot \text{Var}[Y] = \frac{143}{3}.$$

Example 3.13.2. If X is the number of "6"s that turn up when 72 ordinary dice are independently thrown, find the expected value of X.
Solution: X has a binomial distribution with $n = 72$ and $p = \frac{1}{6}$. Then:

$$\mathbb{E}[X] = np = 72 \cdot \frac{1}{6} = 12, \quad \text{Var}[X] = np(1-p) = 12 \cdot \frac{5}{6} = 10.$$

Example 3.13.3. The number of hits, X, per baseball game has a Poisson distribution. If the probability of a no-hit game is $\mathbb{P}(X = 0) = 10^{-4}$, find the probability of having 4 or more hits in a particular game.
Solution: The probability of having no hits is

$$\mathbb{P}(X = 0) = e^{-\lambda} = \frac{1}{10.000} \Longrightarrow \lambda = 4\ln(10).$$

Thus, the probability of having 4 or more hits is:

$$\mathbb{P}(X \geq 4) = 1 - \sum_{k=0}^{3} \mathbb{P}(X = k) = 1 - e^{-\lambda} \sum_{k=0}^{3} \frac{\lambda^k}{k!} = 1 - \frac{1}{10.000} \sum_{k=0}^{3} \frac{\left(4\ln(10)\right)^k}{k!}.$$

Example 3.13.4. In rolling a fair die repeatedly (and independently on successive rolls), find the probability of getting the third "1" on the t-th roll.

Solution: The negative binomial random variable X with parameters $r = 3$ and $p = \frac{1}{6}$ is the number of failures (rolling 2, 3, 4, 5, or 6) until the third success. The probability that the third success occurs on the k-th roll is the same as the probability of having $k - 3$ failures before the third success. Thus, the probability is:

$$\mathbb{P}(X = k - 3) = \binom{k-1}{2} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^{k-3}.$$

Example 3.13.5. An urn contains 6 blue and 4 red balls. 6 balls are chosen at random and without replacement from the urn. If X is the number of red balls chosen, find the standard deviation of X.

Solution: This is a hypergeometric distribution with $M = 10$, $K = 4$, and $n = 6$. The p.m.f of X is:

$$f(x) = \frac{\binom{4}{x}\binom{6}{6-x}}{\binom{10}{6}}, \quad x = 0, 1, 2, 3, 4.$$

The variance is:

$$\mathrm{Var}[X] = \frac{nK(M-K)(M-n)}{M^2(M-1)} = \frac{6 \cdot 4 \cdot 6 \cdot 4}{10^2 \cdot 9} = 0.8.$$

Thus, the standard deviation is then: $\sigma_X = \sqrt{0.8} \approx 0.894$.

3.14 Frequently used continuous distributions

3.14.1 Continuous uniform distribution $\mathcal{U}(a, b)$

Let $-\infty < a < b < \infty$. A continuous random variable X has the uniform distribution on the interval (a, b), and we denote $X \overset{d}{=} \mathcal{U}(a, b)$, if

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

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathrm{Var}[X] = \frac{(b-a)^2}{12}, \quad M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}, \quad t \neq 0.$$

This is a symmetric distribution about the mean and Median = $\frac{a+b}{2}$.

3.14.2 Normal distribution $\mathcal{N}(\mu, \sigma^2)$

A continuous random variable X has the normal distribution with mean $\mu \in \mathbb{R}$ and standard variation $\sigma > 0$, and we denote $X \overset{\mathrm{d}}{=} \mathcal{N}(\mu, \sigma^2)$, if

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

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2, \quad M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \quad t \in \mathbb{R}.$$

The distribution $\mathcal{N}(0,1)$ is referred to as the standard normal distribution. In this case, $F(x)$ is sometimes denoted $\Phi(x)$.

3.14.3 Properties of the normal distibution

From the standard normal table:

$$\Phi(1) = \mathbb{P}(Z \le 1) = 0.8413.$$

Because of symmetry:

$$1 - \Phi(-1) = \mathbb{P}(Z \ge -1) = 0.8413, \quad \Phi(-1) = 1 - \mathbb{P}(Z \ge -1) = 0.1587.$$

In general, for $a > 0$:

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

Given any normal random variable $X \overset{\mathrm{d}}{=} N(\mu, \sigma^2)$:

$$Z = \frac{X - \mu}{\sigma} \overset{\mathrm{d}}{=} N(0,1).$$

Thus

$$\mathbb{P}(r < X < s) = \mathbb{P}\left(\frac{r-\mu}{\sigma} < Z < \frac{s-\mu}{\sigma}\right) = \Phi\left(\frac{s-\mu}{\sigma}\right) - \Phi\left(\frac{r-\mu}{\sigma}\right).$$

Additionally, if $X' \overset{\mathrm{d}}{=} N(\mu', \sigma'^2)$ and is independent of X, then

$$X + X' \overset{\mathrm{d}}{=} N(\mu + \mu', \sigma^2 + \sigma'^2).$$

3.14.4 Exponential distribution $\mathcal{E}(\lambda)$

A positive continuous random variable X has the exponential distribution with mean $\lambda > 0$ if

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \implies F_X(x) = 1 - e^{-\lambda x}, \ \ x \ge 0, .$$

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

An exponential distribution with mean μ has p.d.f. $f(x) = \frac{1}{\mu}e^{-x/\mu}$.

3.14.5 Properties of the exponential distribution

1. **Lack of memory property for $\mathcal{E}(\lambda)$:** For $x, y > 0$,

$$\mathbb{P}(X > x + y \mid X > x) = \mathbb{P}(X > y). \tag{3.141}$$

2. **Link to Poisson distribution:** Suppose that $X \overset{\mathrm{d}}{=} \mathcal{E}(\lambda)$, and we regard X as the time between successive occurrences of some type of event (e.g., the arrival of a new insurance claim at an insurance office), where time is measured in appropriate units (seconds, minutes, hours, or days, etc.).

   Now, imagine choosing a starting time (say labeled as $t = 0$), and from this point onward, we begin recording times between successive events. Let N represent the number of events (claims) that have occurred when one unit of time has elapsed. Then N will be a random variable related to the times of the occurring events. The distribution of N is Poisson with parameter λ.

3.14.6 Gamma distribution $\mathcal{G}(\alpha, \beta)$

Let $\alpha > 0$ and $\beta > 0$. The distribution of a continuous positive r.v. X has the Gamma distribution $\mathcal{G}(\alpha, \beta)$, if

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

Here, $\Gamma(\alpha)$ is the gamma function, defined for $\alpha > 0$ as:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy,$$

from which it follows that if n is a positive integer, $\Gamma(n) = (n-1)!$.

$$\mathbb{E}[X] = \frac{\alpha}{\beta}, \quad \mathrm{Var}(X) = \frac{\alpha}{\beta^2}, \quad M_X(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha, \quad 0 < t < \beta.$$

The exponential distribution with parameter λ is a special case of the gamma distribution with $\alpha = 1$ and $\beta = \lambda$. From the MGF form, we see that if $X \perp\!\!\!\perp Y$, $X \overset{\mathrm{d}}{=} \mathcal{G}(\alpha_X, \beta)$ and Y$\overset{\mathrm{d}}{=}$ $mathcalG(\alpha_Y, \beta)$, then

$$X + Y \overset{\mathrm{d}}{=} \mathcal{G}(\alpha_X + \alpha_Y, \beta).$$

3.15 Functions and transformations of random variables

3.15.1 Distribution of a function of a discrete random variable

Suppose that X is a discrete random variable with probability mass function $p(x)$. If $u(x)$ is a function of x, and Y is a random variable defined by the equation $Y = u(X)$, then Y is a discrete random variable with probability function

$$g(y) = \sum_{i,\, u(x_i)=y} p(x_i),$$

where given a value of y, find all values of x for which $y = u(x)$ (say $u(x_1) = u(x_2) = \cdots = u(x_n) = y$), and then $g(y)$ is the sum of those $p(x_i)$ probabilities.

3.15.2 Distribution of a function of a continuous random variable

Suppose that X is a continuous random variable with p.d.f. $f(x)$ and c.d.f. $F(x)$, and suppose that $u(x)$ is a one-to-one function (usually u is either strictly increasing, such as $u(x) = x, e^x, \ln x$, or u is strictly decreasing, such as $u(x) = e^{-x}$). As a one-to-one function, u has an inverse function v, so that

$$v(u(x)) = x.$$

Then the random variable $Y = u(X)$ (Y is referred to as a transformation of X) has p.d.f.

$$f_Y(y) = |v'(y)|\, f_X(v(y)).$$

If u is a strictly increasing function, then

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(u(X) \le y) = \mathbb{P}(X \le v(y)) = F_X(v(y)).$$

3.16 Solved problems

Example 3.16.1. Suppose that $X \stackrel{\mathrm{d}}{=} \mathcal{U}(0, a)$, where $a > 0$. Find $\mathbb{P}(X > X^2)$.
Solution: We have

$$\mathbb{P}(X > X^2) = \mathbb{P}(X < 1) = \begin{cases} 1, & a \ge 1, \\ \frac{1}{a}, & a < 1. \end{cases}$$

Example 3.16.2. A random variable T has an exponential distribution such that $\mathbb{P}(T \le 20 = 2\mathbb{P}(T > \ln 2)$. Find $\mathrm{Var}(T)$.
Solution: We have

$$\mathbb{P}(T \le 2) = 1 - e^{-2\lambda} = 2e^{-\lambda \ln 2}.$$

Solving yields $\lambda = \ln 2$, so:

$$\mathrm{Var}(T) = \frac{1}{\lambda^2} = \frac{1}{(\ln 2)^2}.$$

Example 3.16.3. If X is a normal random variable with $\mathbb{P}(X < 500) = 0.5$ and $\mathbb{P}(X > 650) = 0.0227$, find the standard deviation of X.
Solution: Since

$$\mu = 500, \quad \mathbb{P}\left(Z > \frac{650 - 500}{\sigma}\right) = 0.0227.$$

and since, $Z_0.0227 = \frac{150}{\sigma} = 2.00$ (from the standard normal table), then

$$\sigma = \frac{650 - 500}{2.00} = 75.$$

Example 3.16.4. Verification of Exponential Distribution Properties. Show the lack of memory property (3.141) for $X \stackrel{\mathrm{d}}{=} \mathcal{E}(\lambda)$.
Solution: Using $\mathbb{P}(X > u) = e^{-\lambda u}$, $u > 0$, we obtain

$$\mathbb{P}(X > x + y \mid X > x) = \frac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y}.$$

Example 3.16.5. Normal approximation to the binomial distribution. Suppose that X has a binomial distribution based on 100 trials with a probability of success of 0.2 on any given trial. Find the approximate probability $\mathbb{P}(15 \le E \le 25)$.
Solution: The mean and variance of X are:

$$\mathbb{E}[X] = 100(0.2) = 20, \quad \mathrm{Var}(X) = 100(0.2)(0.8) = 16.$$

Using the normal approximation with integer correction, X is approximately normal. We calculate:

$$\mathbb{P}(15 \le X \le 25) = \mathbb{P}(14.5 \le X \le 25.5).$$

Standardizing:

$$\mathbb{P}(14.5 \le X \le 25.5) = \mathbb{P}\left(\frac{14.5 - 20}{\sqrt{16}} \le Z \le \frac{25.5 - 20}{\sqrt{16}}\right) = \mathbb{P}(-1.375 \le Z \le 1.375),$$

where Z is a standard normal random variable. From standard normal tables:

$$\mathbb{P}(-1.375 \le Z \le 1.375) = \Phi(1.375) - \Phi(-1.375) = 2\Phi(1.375) - 1.$$

Using linear interpolation

$$\Phi(1.375) \approx 0.25\Phi(1.3) + 0.75\Phi(1.4) = 0.25(0.9032) + 0.75(0.9192) = 0.9152,$$

we finally obtain

$$\mathbb{P}(15 \leq X \leq 25) \approx 2(0.9152) - 1 \approx 0.8304.$$

Chapter 4

Multivariate Random Variables

4.1 Joint and Marginal

4.1.1 Joint distribution of random variables X and Y:

A joint distribution of two random variables is described as follows.

Discrete Case: If X and Y are discrete random variables, then have joint p.m.f $p(k,l) = \mathbb{P}(X=k, Y=l)$, $(k,l) \in (X,Y)(\Omega)$, which must satisfy

$$0 \le p(k,l) \le 1 \text{ and } \sum_k \sum_l p(k,l) = 1.$$

Thus,

$$\mathbb{P}(X \in A, Y \in B) = \sum_{k \in A} \sum_{l \in B} p(k,l), \quad A \times B \subset (X,Y)(\Omega).$$

Continuous Case: If X and Y are continuous random variables, then they have a joint p.d.f. $f(x,y)$, which must satisfy

$$f(x,y) \ge 0 \text{ and } \int \int f(x,y)\, dy\, dx = 1.$$

Thus,

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f(x,y)\, dy\, dx, \quad A \times B \subset (X,Y)(\Omega).$$

Remark 4.1.1. *It is possible to have a joint distribution where one variable is discrete, and the other is continuous, or where either has a mixed distribution. The joint distribution of two random variables can be extended to a joint distribution of any number of random variables.*

Cumulative Distribution Function: If random variables X and Y have a joint distribution, then the cumulative distribution function is:

$$F(x,y) = \mathbb{P}(X \le x, Y \le y) \begin{cases} \sum_{k \le x} \sum_{l \le y} p(k,l), & \text{in the discrete case,} \\ \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du, & \text{in the continuous case.} \end{cases}$$

We have

$$\begin{aligned} \lim_{x \to -\infty} F(x,y) &= \lim_{y \to -\infty} F(x,y) = 0, \\ \mathbb{P}(x_1 < X \le x_2, y_1 < Y \le y_2) &= F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1), \\ \mathbb{P}(X > x, Y > y) &= F_X(x) + F_Y(y) - F(x,y) - 1. \end{aligned}$$

Note that in the continuous case, we have

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y} = \frac{\partial^2 F(x,y)}{\partial y\, \partial x}.$$

4.1.2 Marginal distributions

If X and Y have a joint p.m.f. $p(k,l)$ (respectively a joint p.d.f. $f(x,l)$, then the marginal distribution of X has a p.m.f. (respectively a p.d.f. given by

$$p_X(k) = \sum_l p(k,l) \quad \text{(in the discrete case)},$$

respectively,

$$f_X(x) = \int f(x,y)\, dy \quad \text{(in the continuous case)}.$$

The cumulative distribution of X is then:

$$F_X(x) = \lim_{y\to\infty} F(x,y).$$

The density function for the marginal distribution of Y is found in a similar way. This concept can be extended to define the marginal distribution of any one (or subcollection) of variables in a multivariate distribution.

4.1.3 Expectation of a function of jointly distributed random variables

If $h(x,y)$ is a function of two variables, and X and Y are jointly distributed random variables, then the expected value of $h(X,Y)$ is defined as:

$$\mathbb{E}[h(X,Y)] = \begin{cases} \sum_k \sum_l h(k,l)\cdot p(k,l), & \text{(discrete case)}, \\ \int\int h(x,y)\cdot f(x,y)\, dy\, dx, & \text{(continuous case)}. \end{cases}$$

4.1.4 Covariance

If random variables X and Y are jointly distributed with joint density/probability function $f(x,y)$, then the covariance between X and Y is:

$$\text{Cov}[X,Y] = E\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

The covariance satisfies the following

- Symmetry:

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X] \quad \text{with} \quad \mathrm{Cov}[X,X] = \mathrm{Var}[X];$$

- Bilinearity:

$$\mathrm{Cov}[X+X',Y+Y'] = \mathrm{Cov}[X,Y] + \mathrm{Cov}[X,Y'] + \mathrm{Cov}[X',Y'] + \mathrm{Cov}[X',Y].$$

4.1.5 Coefficient of correlation

The coefficient of correlation between random variables X and Y is defined as:

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[Y]}} = \rho_{Y,X}.$$

It satisfies $-1 \le \rho_{X,Y} \le 1$, and $|\rho_{X,Y}| = 1$ if and only if $Y = aX + b$ for some constant a, b.

4.1.6 Moment generating function of a joint distribution

Given jointly distributed random variables X and Y, the moment generating function of the joint distribution is:

$$M(s,t) = E\left[e^{sX+tY}\right], \ (s,t) \in \mathcal{D}.$$

where s and t are real numbers in some domain of definition. This concept can be extended to the moment generating function of the joint distribution of any number of random variables. Note that

$$M_X(s) = M(s,0), \ \ M_Y(t) = M(0,t).$$

4.2 Distribution of a sum of random variables

(a) If X and Y are discrete non-negative integer-valued random variables with joint probability mass function $p(k,l)$, then for an integer $n \ge 0$,

$$\mathbb{P}(X_1 + X_2 = n) = \sum_{k,l} p(k, n-l).$$

If furthermore X and Y are independent, then

$$\mathbb{P}(X_1 + X_2 = n) = \sum p_X(k)\cdot p_Y(n-k).$$

(This is the convolution method of finding the distribution of the sum of independent discrete random variables.)

(b) If random variable $(X,Y) \in \mathbb{R}^2$ is jointly continuous with joint p.d.f. $f(x,y)$, then X and Y are also continuous random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively. The density function of $Z = X + Y$ is given by:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)dx = \int_{-\infty}^{\infty} f(x-y, y)dy.$$

If furthermore X and Y are independent, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(x-y)\, f_Y(y)dy.$$

This integral represents the *convolution product* of the density functions of f_X and f_Y.

(c) If $X_1, X_2, \ldots, X_n$ are random variables with finite second moments, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

Recall that if $X_1, X_2, \ldots, X_n$ are mutually independent random variables, then

$$\mathrm{Cov}(X_i, X_j) = 0.$$

4.3 Sums of certain distributions

Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables and $Y = \sum_{i=1}^{n} X_i$. Then we have the following stability by convolution:

Distribution of X_i	**Distribution of Y**
Bernoulli $\mathcal{B}(p)$	Binomial $\mathcal{B}(n, p)$
Geometric $\mathcal{G}(p)$ on $\mathbb{N}$	Negative Binomial $\mathcal{NB}(n, p)$
Geometric $\mathcal{G}(p)$ on $\mathbb{N}^\star$	Negative Binomial $\mathcal{P}_a(n, p)$
Poisson $\mathcal{P}(\lambda_i)$	$\mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right)$
Gamma $\mathcal{G}(\alpha_i, \beta)$	Gamma $\mathcal{G}\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$

4.4 Conditional distribution of Y Given $X = x$

4.4.1 Definition

Suppose that the random variables X and Y have a joint p.m.f. $p(k = l)$ (respectively p.d.f. $f(x, y)$), and the marginal distribution of X is $p_X(k)$ (respectively p.d.f. $f_X(x)$). Then, the conditional distribution of Y given $X = x$ is given by

$$p_{Y|X=k}(l) = \frac{p(k,l)}{p_X(k)}, \text{ if } p_X(k) > 0 \qquad \left(\text{resp. } f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} \text{ if } f_X(x) > 0\right).$$

The same formula holds for X given $Y = y$ by exchanging the role of X and Y.

4.4.2 Properties

(i) The joint density/probability function of X and Y can be written as:

$$p(k,l) = p_{Y|X=k}(l) \cdot p_X(k) = p_{X|Y=l}(k) \cdot p_Y(l), \text{ (discrete case)}$$

$$f(x,y) = f_{Y|X=x}(y) \cdot f_X(x) = f_{X|Y=y}(x) \cdot f_Y(y), \text{ (continuous case).}$$

(ii) The conditional expectation of X given $Y = x$ is:

$$\mathbb{E}[X \mid Y = y] = \sum_x x \cdot p_X(x \mid Y = y), \text{ (discrete case)}$$

$$\mathbb{E}[X \mid Y = y] = \int x \cdot f(x \mid Y = y)\, dx, \text{ (continuous case).}$$

(iii) Note that $\mathbb{E}[X \mid Y = y]$ depends on y:

$$h(y) \quad := \quad \mathbb{E}[X \mid Y = y] = \begin{cases} \sum_x x\, \mathbb{P}(X = x|Y = y), & \text{(discrete case),} \\ \int x\ f_{X|Y=y)}(x)\ dx, & \text{(continuous case).} \end{cases} \tag{4.41}$$

It can be shown that $h(Y) = \mathbb{E}[X \mid Y]$, hence

$$\mathbb{E}[h(Y)] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X].$$

We also have

$$\text{Var}[X] = \mathbb{E}[\text{Var}[X \mid Y]] + \text{Var}[\mathbb{E}[X \mid Y]].$$

(iv) If X and Y are independent, then:

$$p_X(k \mid Y = l) = p_X(k) \quad \text{and} \quad p_Y(l \mid X = k) = p_Y(l) \quad \text{(discrete case).}$$

and

$$f_X(x \mid Y = y) = f_X(x) \quad \text{and} \quad f_Y(y \mid X = x) = f_Y(y) \quad \text{(continuous case).}$$

4.5 Some results and formulas

(a) The marginal MGF are given by:

$$M_X(t,0) = \mathbb{E}[e^{tX}] = M_X(t), \quad M_Y(0,t) = \mathbb{E}[e^{tY}] = M_Y(t).$$

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

$$M_{X,Y}(s,t) = \mathbb{E}[e^{sX+tY}] = M_X(s) \cdot M_Y(t),$$

for (s,t) in some region in $\mathbb{R}^2$.

(b) If X and Y have a joint uniform distribution over a region R (usually R will be a triangle, rectangle or circle in the (x,y) plane), then the conditional distribution of Y given $X = x$ has a uniform distribution on the line segment (or segments) defined by the intersection of the region R with the line $X = x$. The marginal distribution of Y might or might not be uniform.

4.6 Approximation of sums, the law of large numbers and the central limit theorem

4.6.1 Law of large numbers

The following result is central is statistical theory, as it approximates the sample mean.

Theorem 4.6.1 (Law of Large numbers (LLN)). *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables (i.i.d.) with mean μ. As the sample size n becomes large, the sample mean*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

approaches μ:

$$\overline{X}_n \xrightarrow{a.s.} \mu.$$

where $\xrightarrow{a.s.}$ is convergence in the almost sure sense.

4.6.2 Central Limit Theorem

The following result is a justification for the importance of the normal distribution, as it approximates to the distribution of a sum of random variables.

Theorem 4.6.2 (Central Limit Theorem (CLT)). *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. As the sample size n becomes large, the distribution of the sample mean*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

approaches a normal distribution:

$$\overline{X}_n \overset{d}{\approx} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

or more precisely,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{as } n \to \infty,$$

4.7 Solved problems

Example 4.7.1 (The multinomial distribution)**.** The multinomial distribution is defined with parameters n, $p_1, p_2, \ldots, p_k$ (where n is a positive integer and $p_1 + p_2 + \cdots + p_k = 1$ with $p_i \geq 0$ for all $i = 1, 2, \ldots, k$) as follows. Suppose that an experiment has k possible outcomes, with probabilities $p_1, p_2, \ldots, p_k$ respectively. If the experiment is performed n successive times independently, let X_i denote the number of experiments that resulted in outcome i, so that:

$$X_1 + X_2 + \cdots + X_k = n.$$

The joint probability mass function of $X_1, X_2, \ldots, X_k$ is:

$$\mathbb{P}(X_1 = i_1, X_2 = i_2, \ldots, X_k = i_k) = \frac{n!}{i_1! i_2! \cdots i_k!} p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k},$$

where $i_1 + i_2 + \cdots + i_k = n$ and $i_j \geq 0$ for all $j = 1, \ldots k$. Note that the marginal distributions are such that

$$X_i \stackrel{\text{d}}{=} \mathcal{B}(n, p_i) \quad \text{and} \quad \text{Cov}[X_i, X_j] = -np_i p_j \quad \text{for } 1 \leq i \neq j \leq k.$$

Thus,

$$\text{Corr}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i] \cdot \text{Var}[X_j]}} = \frac{-p_i p_j}{\sqrt{p_i(1 - p_i) \cdot p_j(1 - p_j)}}.$$

For example, the toss of a fair die results in one of $k = 6$ outcomes, with probabilities:

$$p_1 = p_2 = \cdots = p_6 = \frac{1}{6}.$$

If the die is tossed n times, then the counts of each face appearing, then with

$$X_i \ = \ \text{of tosses resulting in face } i \text{ turning up},$$

the random variable $X = X_1 + \ldots + X_n$ follows a multinomial distribution.

Example 4.7.2. Let X and Y be two continuous random variables with joint p.d.f.

$$f(x, y) = K\ (x^2 + y^2),$$

defined over the unit square bounded by the points $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$. Find K.
Solution: The (double) integral of the density function over the region of density must equal 1. Thus:

$$1 = \int_0^1 \int_0^1 K\ (x^2 + y^2)\, dx\, dy = K \int_0^1 \left(\int_0^1 (x^2 + y^2)\, dx \right) dy = K \int_0^1 \left(\frac{1}{3} + y^2 \right) dy = \frac{2}{3} K.$$

Therefore, $K = \frac{3}{2}$.

Example 4.7.3. The cumulative distribution function for the joint distribution of the continuous random variables X and Y is given by:

$$F(x,y) = \frac{1}{5}(3x^3\,y + 2x^2y^2), \quad \text{for } 0 \le x \le 1,\ 0 \le y \le 1.$$

Find the joint p.d.f.
Solution: The density function is obtained by differentiating the cumulative distribution function: for $x,\,y \in (0,1)$,

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\,\partial y} = \frac{1}{5}\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}(3x^3 + 2x^2y^2)\right) = \frac{1}{5}\frac{\partial}{\partial x}\left(3x^3 + 4x^2y\right) = \frac{1}{5}\left(9x^2 + 8xy\right).$$

Example 4.7.4. X and Y are discrete random variables that are jointly distributed with the following probability function $f(x,y)$:

$Y\backslash X$	-1	0	1
1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
0	$\frac{1}{9}$	0	$\frac{1}{6}$
-1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{9}$

Find $\mathbb{E}[XY]$.
Solution:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_k\sum_l kl\cdot p(k,l)\\ &= (-1)(1)\frac{1}{18} + (-1)(-1)\frac{1}{6} + (1)(1)\frac{1}{6} + (1)(-1)\frac{1}{9} + 0 = \frac{1}{6}.\end{aligned}$$

Example 4.7.5. Continuous random variables X and Y have a joint distribution with density function

$$f(x,y) = \frac{3}{2}(2 - 2x - y),$$

in the region bounded by $y = 0$, $x = 0$, and $y = 2 - 2x$. Find the p.d.f. $f_X(x)$ for $0 < x < 1$.
Solution: For $0 < x < 1$, we have

$$f_X(x) = \int_0^{2-2x} f(x,y)\,dy = \frac{3}{2}\int_0^{2-2x}(2 - 2x - y)\,dy = 3(1-x)^2.$$

Example 4.7.6. Suppose that X and Y are independent continuous random variables with the following density functions:

$$f_X(x) = 1 \text{ for } 0 < x < 1 \quad \text{and} \quad f_Y(y) = 2y \text{ for } 0 < y < 1.$$

Find $\mathbb{P}(Y < X)$.

Solution: Since X and Y are independent, the density function of the joint distribution of E and Y is

$$f(x,y) = f_X(x) \cdot f_Y(y) = 2y, \;\; 0 < x < 1, \; 0 < y < 1$$

Thus

$$\mathbb{P}(Y < X) = \int_0^1 \int_0^1 2y \, dy \, dx = \frac{1}{3}.$$

Example 4.7.7. Continuous random variables X and Y have a joint distribution with density function

$$f(x,y) = x^2 + \frac{xy}{3} + \text{ for } 0 < x < 1 \text{ and } 0 < y < 2.$$

Find $\mathbb{P}(X > \frac{1}{2} \mid Y > \frac{1}{2})$.

Solution: For $0 < u < 1, \; 0 < v < 2$,

$$\begin{aligned}
\mathbb{P}(X > u, Y > v) &= \int_u^1 \int_v^2 \left[x^2 + \frac{xy}{3}\right] dy \, dx = \int_x^1 \left[x^2 y + \frac{xy^2}{6}\right]_{y=v}^{y=2} dx \\
&= \int_u^1 \left[x^2(2-v) + \frac{x(4-v^2)}{6}\right]_{y=v}^{y=2} dx = \left[\frac{x^3(2-v)}{3} + \frac{x^2(4-v^2)}{12}\right]_{x=u}^{x=1} \\
&= \frac{(1-u^3)(2-v)}{3} + \frac{(1-u^2)(4-v^2)}{12}
\end{aligned}$$

Thus,

$$\mathbb{P}\left(Y > \frac{1}{2}\right) = \mathbb{P}\left(X > 0, Y > \frac{1}{2}\right) = \frac{13}{16} \quad \text{and} \quad \mathbb{P}\left(X > \frac{1}{2}, Y > \frac{1}{2}\right) = \frac{43}{64},$$

hence

$$\mathbb{P}\left(X > \frac{1}{2} \mid Y > \frac{1}{2}\right) = \frac{\mathbb{P}(X > \frac{1}{2}, Y > \frac{1}{2})}{\mathbb{P}(Y > \frac{1}{2}} = \frac{43}{64} \times \frac{16}{13} = \frac{43}{52}$$

Example 4.7.8. X is a continuous random variable with density function

$$f(x) = x + \frac{1}{2} \text{ for } 0 < x < 1.$$

X is also jointly distributed with the continuous random variable Y, and the conditional density function of Y given $X = x, \; 0 < x < 1$, is

$$f_{Y|X=x}(y) = \frac{x+y}{x+\frac{1}{2}} \text{ for } 0 < y < 1.$$

Find $f_X(y)$. **Solution:** The joint density function is given by

$$f(x,y) = f_{Y|X=x}(y) \cdot f_X(x) = \left(\frac{x+y}{x+\frac{1}{2}}\right) \cdot \left(x + \frac{1}{2}\right) = x + y.$$

Then, the marginal density of Y is

$$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 (x+y)\,dx = y + \frac{1}{2}, \ \ 0 < y < 1.$$

Example 4.7.9. Find $\mathrm{Cov}(X,Y)$ for the jointly distributed discrete random variables in Example 4.7.4 above.

Solution: We have found $\mathbb{E}[XY] = \frac{1}{6}$, it remains to compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x \sum_y \mathbb{P}(X = x, Y = y) = (-1)(\frac{1}{18} + \frac{1}{9} + \frac{1}{6}) + 0 + (1)(\frac{1}{6} + \frac{1}{6} + \frac{1}{9}) = -\frac{1}{3} + \frac{4}{9} = \frac{1}{9} \\
\mathbb{E}[Y] &= \sum_x \sum_y y\mathbb{P}(X = x, Y = y) = (1)(\frac{1}{18} + \frac{1}{9} + \frac{1}{6}) + 0 + (-1)(\frac{1}{6} + \frac{1}{9} + \frac{1}{9}) = \frac{1}{3} - \frac{7}{18} = -\frac{1}{18}.
\end{aligned}$$

Thus,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = \frac{1}{6} - \frac{1}{9}\frac{1}{18} = \frac{41}{81}.$$

Example 4.7.10. The coefficient of correlation between random variables E and Y is $\frac{1}{3}$, and $\sigma_X^2 = a$, $\sigma_Y^2 = 4a$. The random variable Z is defined to be $Z = 3X - 4Y$, and it is found that $\sigma_Z^2 = 114$. Find a.

Solution: Since

$$\mathrm{Cov}(X,Y) = \frac{\sigma_X \sigma_Y}{3} = \frac{2a}{3},$$

and since the variance of Z is given by

$$114 = \mathrm{Var}(3X - 4Y) = 9 \cdot \mathrm{Var}(E) + 16 \cdot \mathrm{Var}(Y) - 24 \cdot \mathrm{Cov}(X,Y) = 9a + 64a - 16a = 77a,$$

we deduce requested value is $a = 2$.

Example 4.7.11. Suppose that X and Y are random variables whose joint distribution has moment generating function

$$M(t_1, t_2) = \left(\frac{3 + 2e^{t_1} + 3e^{t_2}}{8}\right)^{10}, \ \text{ for all real } t_1 \text{ and } t_2.$$

Find the covariance between X and Y.

Solution: To find the expectations, we need to compute the partial derivatives of the MGF with respect to t_1 and t_2.

$$\begin{aligned}
\frac{\partial M(t_1,t_2)}{\partial t_1} &= \frac{10}{8^{10}} \times 2e^{t_1}(3 + 2e^{t_1} + 3e^{t_2})^9 \\
\frac{\partial M(t_1,t_2)}{\partial t_2} &= \frac{10}{8^{10}} \times 3e^{t_2}(3 + 2e^{t_1} + 3e^{t_2})^9
\end{aligned}$$

Then using any of the following formula,

$$\frac{\partial^2 M(t_1,t_2)}{\partial t_1 \partial t_2} = \frac{\partial}{\partial t_1}\frac{\partial M(t_1,t_2)}{\partial t_2} = \frac{\partial}{\partial t_2}\frac{\partial M(t_1,t_2)}{\partial t_1},$$

we find

$$\frac{\partial^2 M(t_1,t_2)}{\partial t_1 \partial t_2} = \frac{10}{8^{10}} \times 6e^{t_1+t_2}(3+2e^{t_1}+3e^{t_2})^8.$$

Now, evaluating at $t_1 = 0$ and $t_2 = 0$, we find

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\partial M(t_1,t_2)}{\partial t_1}\Big|_{t_1=0,t_2=0} = \frac{5}{2} \\
\mathbb{E}[Y] &= \frac{\partial M(t_1,t_2)}{\partial t_2}\Big|_{t_1=0,t_2=0} = \frac{15}{4} \\
\mathbb{E}[XY] &= \frac{\partial^2 M(t_1,t_2)}{\partial t_1 \partial t_2}\Big|_{t_1=0,t_2=0} = \frac{135}{16}.
\end{aligned}$$

The covariance is then $\mathrm{Cov}(X,Y) = -\frac{15}{16}$.

Example 4.7.12. Suppose that X has a continuous distribution with probability density function

$$f_X(x) = \begin{cases} 2x & \text{for } x \in (0,1), \\ 0 & \text{elsewhere,} \end{cases}$$

Suppose that Y is a continuous random variable such that the conditional distribution of Y given $X = x$ is uniform on the interval $(0,x)$. Find the mean and variance of Y.
Solution: We first compute the mean and the variance of X:

$$\begin{aligned}
\mathbb{E}[X] &= \int x f_X(x)\,dx = 2\int_0^1 x^2\,dx = \frac{2}{3}. \\
\mathbb{E}[X^2] &= \int_0^1 x^2 f_X(x)\,dx = 2\int_0^1 x^3\,dx = \frac{1}{2} \\
\mathrm{Var}(X) &= \frac{1}{18}.
\end{aligned}$$

Since $Y|X = x$ has the uniform distribution on the interval $(0,x)$, with conditional p.d.f.

$$f_{Y|X=x}(y) = \begin{cases} \frac{1}{x} & \text{for } 0 \le y \le x, \\ 0 & \text{elsewhere,} \end{cases}$$

hence with mean and variance

$$\mathbb{E}[Y|X = x] = \frac{0+x}{2} = \frac{x}{2}, \quad \mathrm{Var}(Y|X = x) = \frac{x^2}{12}.$$

we deduce that the conditional mean and variance

$$\mathbb{E}[Y|X] = \frac{X}{2}, \quad \text{Var}(Y|X) = \frac{X^2}{12},$$

which give

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}\big[[\mathbb{E}[Y|X]\big] = \frac{\mathbb{E}[X]}{2} = \frac{1}{3} \\
\text{Var}(\mathbb{E}[Y|X]) &= \frac{\text{Var}(X)}{4} = \frac{1}{72} \\
\mathbb{E}\big[\text{Var}(Y|X)\big] &= \frac{\mathbb{E}[X^2]}{12} = \frac{1}{24} \\
\text{Var}(Y) &= \text{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\text{Var}(Y|X)] = \frac{1}{24} + \frac{1}{72} = \frac{1}{18}.
\end{aligned}$$

Example 4.7.13. The random variable X has an exponential distribution with a mean of 1. The random variable Y is defined to be $Y = e^{-bX}$. Find $f_Y(y)$, the p.d.f. of Y.
Solution: For $y \in (0,1)$, we have

$$\mathbb{P}(Y \leq y) = \mathbb{P}(e^{-bX} \leq y) = \mathbb{P}(X \geq -\frac{\ln y}{b}) = 1 - \mathbb{P}(X < -\frac{\ln y}{b}) = 1 - e^{\frac{\ln y}{b}} = 1 - y^{\frac{1}{b}}.$$

Differentiating, we get,

$$f_Y(y) = b\, y^{1+\frac{1}{b}}\ , \quad 0 < y \leq 1.$$

Example 4.7.14. The random variable X has an exponential distribution with a mean of 1. The random variable Y is defined to be $Y = 2\ln(X)$. Find $f_Y(y)$ the p.d.f. of Y.
Solution: We have

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(2\ln(X) \leq y) = \mathbb{P}\left(X \leq e^{y/2}\right) = 1 - e^{-e^{y/2}}, \ \ y \in \mathbb{R}.$$

Then

$$f_Y(y) = F_Y'(y) = \frac{1}{2}e^{y/2}e^{-e^{y/2}}, \ \ y \in \mathbb{R}.$$

Alternatively, we choose $y = u(x) = 2\ln(x))$, $x > 0$, to get $Y == u(x)$. The inverse v of u is given by $x = v(y) = e^{y/2}$, $y \in \mathbb{R}$. It follows that

$$f_Y(y) = |v'(y)|\ f_X(v(y)) = \frac{1}{2}e^{y/2}e^{-e^{y/2}}, \ \ y \in \mathbb{R}.$$

Example 4.7.15. Suppose that X and Y are independent discrete integer-valued random variables with X uniformly distributed on the integers 1 to 5 , and Y having the following probability mass function

$$f_Y(0) = 0.3, \ f_Y(1) = 0.5, \ f_Y(3) = 0.2.$$

Let $Z = X + Y$, find $\mathbb{P}(Z = 5)$.
Solution: Using the fact that $f_X(k) = 0.2$ for all $k \in \{1, \dots, 5\}$, and the convolution method for independent discrete random variables, we have

$$f_Z(5) = \sum_{k=1}^{k} f_X(k) f_Y(5-k) = (0.2)(0) + (0.2)(0.2) + (0.2)(0) + (0.2)(0.5) + (0.2)(0.2) = 0.2.$$

Example 4.7.16. X_1 and X_2 are independent exponential random variables each with a mean of 1. Find $\mathbb{P}(X_1 + X_2 < 1)$.
Solution: Using the convolution method, the density function of $Y := X_1 + X_2$ is

$$f_Y(y) = \int_0^y f_{X_1}(t) f_{X_2}(y-t) dt = \int_0^y e^{-t} e^{-(y-t)} dt = y e^{-y}, \;\; y > 0,$$

hence Y has the gamma distribution $\mathcal{G}(2, 1)$. So that

$$\mathbb{P}(X_1 + X_2 < 1) = \int_0^1 y \, e^{-y} \, dy = 1 - 2e^{-1}.$$

Example 4.7.17. Independent random variables X and Y and Z are identically distributed. Let $W = X + Y$. The moment generating function of W is

$$M_W(t) = (0.7 + 0.3e^t)^6, \;\; t \in \mathbb{R}.$$

Find the moment generating function of $S := X + Y + Z$.
Solution: For independent random variables, the moment generating function of the sum is the product of the moment generating functions. Since X and Y are i.i.d, they have the same moment generating function. Thus,

$$M_W(t) = M_X(t) M_Y(t) = \left(M_X(t)\right)^2, \;\; M_X(t) \geq 0 \Longrightarrow M_X(t) = \sqrt{M_W(t)} = (0.7 + 0.3e^t)^3.$$

Similarly, X, Y and Z are i.i.d, we obtain

$$M_S(t) = (M_X(t))^3 = (0.7 + 0.3e^t)^9.$$

Alternatively, note that the moment generating function of the binomial distribution $Bin(n, p)$ is

$$M(t) = (1 - p + pe^t)^n.$$

Thus, S has a binomial distribution $\mathcal{B}(9, 0.3)$, and each of X, Y, and Z has a binomial distribution, so the sum of these independent binomial distributions is $\mathcal{B}(3, 0.3)$.

Example 4.7.18. The birth weight of males is normally distributed with mean 6 pounds, 10 ounces, standard deviation 1 pound. For females, the mean weight is 7 pounds, 2 ounces with standard deviation 12 ounces. Given two independent male/female births, find the probability that the baby boy outweighs the baby girl.

Solution: Let random variables X and Y denote the boy's weight and girl's weight, respectively. Then, $W = X - Y$ has a normal distribution with mean

$$6\frac{10}{16} - 7\frac{2}{16} = -\frac{1}{2} \quad \text{and variance} \quad \sigma_X^2 + \sigma_Y^2 = 1 + \frac{9}{16} = \frac{25}{16}.$$

Then, the probability that the boy outweighs the girl is

$$\mathbb{P}(X > Y) = \mathbb{P}(W > 0) = \mathbb{P}\left(\frac{W - (-\frac{1}{2})}{\sqrt{\frac{25}{16}}} > \frac{\frac{1}{2}}{\frac{5}{4}}\right) = \mathbb{P}(Z > \frac{2}{5}) = \mathbb{P}(Z > 0.4).$$

Since Z has been standardized, this is the probability that a standard normal variable exceeds 0.4, which is approximately 0.3446.

Example 4.7.19. If the number of typographical errors per page typed by a certain typist follows a Poisson distribution with a mean λ, find the probability that the total number of errors in 10 randomly selected pages is 10.

Solution: The 10 randomly selected pages have independent distributions of errors per page. The sum of n independent Poisson random variables with parameter λ has a Poisson distribution with parameter $n\lambda$. Thus, the total number of errors in the 10 randomly selected pages follows a Poisson distribution with parameter 10λ. The probability of 10 errors in the 10 pages is

$$\frac{e^{-10\lambda}(10\lambda)^{10}}{10!}.$$

Part II

Mathematical Tools for Probability Theory

Chapter 5

Sequences, Limits and Series

5.1 Sequences

Finite sequences are also called strings, denoted by

$$a_1,\ a_2,\ a_3, \dots a_n.$$

A sequence is defined as a function from a subset of $\mathbb{N}$ to a set S. We use the notation a_n to denote the image of the integer n. We call a_n a term of the sequence and we use the notation $(a_n)_{n\in\mathbb{N}}$ to describe the sequence. It is convenient to describe a sequence with a formula.

Example 5.1.1.

n	1	2	3	4	5	$\cdots$
a_n	2	4	6	8	10	$\cdots$

This sequence can be specified as $a_n = 2n$.

Example 5.1.2. What are the formulas that describe the following sequences $a_1, a_2, a_3, \dots$?

1) $1, 3, 5, 7, 9, \dots$;
2) $-1, 1, -1, 1, -1, \dots$;
3) $2, 5, 10, 17, 26, \dots$;
4) $0, 0.25, 0.5, 0.75, 1, 1.25, \dots$;
5) $1, 3, 9, 27, 81, 243, \dots$.
6) $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \dots$.

Solution:

1) $1, 3, 5, 7, 9, \dots, \quad a_n = 2n - 1,\ n \geq 1.$
2) $-1, 1, -1, 1, -1, \dots, \quad a_n = (-1)^n,\ n \geq 1.$
3) $2, 5, 10, 17, 26, \dots, \quad a_n = n^2 + 1,\ n \geq 1.$
4) $0, 0.25, 0.5, 0.75, 1, 1.25, \dots, \quad a_n = 0.25n,\ n \geq 0.$
5) $1, 3, 9, 27, 81, 243, \dots, \quad a_n = 3^n,\ n \geq 0.$
6) This is the famous Fibonacci sequence $a_{n+2} = a_{n+1} + a_n$. With some analysis, it can be shown that is has the closed-form expression:

$$a_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}, \quad \text{where } \varphi = \frac{1+\sqrt{5}}{2}. \tag{5.11}$$

5.1.1 Arithmetic and geometric sequences

An arithmetic sequence satisfies

$$a_{n+1} - a_n = r,\ r \in \mathbb{R},\ n \in \mathbb{N}. \quad \text{Its form is } a_n = a_0 + nr.$$

A geometric non-null sequence satisfies

$$a_{n+1}/a_n = r,\ r \in \mathbb{R}^\star,\ n \in \mathbb{N}. \quad \text{Its form is } a_n = a_0\, r^n.$$

5.1.2 Summations

A summation of a sequence $(a_j)_j$ starting form the integer rang m and finishing at the rank n, $0 \le m \le n$, represents the sum

$$\sum_{j=m}^{n} a_j = a_m + a_{m+1} + a_{m+2} + \cdots + a_n.$$

The variable j is called the **index of summation**, running from its lower limit m to its upper limit n. Any other letter could be used to denote this index.
1) What does $\sum_{j=1}^{6} j$ stand for?

$$\sum_{j=1}^{6} j = 1 + 2 + 3 + 4 + 5 + 6 = 21.$$

It is tedious to calculate this manually.
2) How can we express the sum of the first 1000 terms of the sequence $(a_n)_n$ where $a_n = n^2$ for $n = 1, 2, 3, \ldots$? We write it as $\sum_{n=1}^{1000} n^2$. What is the value of this summation?

Example 5.1.3 (Arithmetic Sum)**.** How does:

$$s_n := 1 + 2 + 3 + \cdots + n = \sum_{j=1}^{n} j = \frac{n(n+1)}{2}, \tag{5.12}$$

work?
Observation:

$$\begin{aligned} s_n &= 1 + 2 + 3 + \cdots + n = n + (n-1) + (n-2) + \cdots + 1 \\ 2s_n &= (n+1) + (n+1) + \cdots + (n+1) \quad \text{(with } n \text{ terms)} \\ s_n &= \frac{n(n+1)}{2}. \end{aligned}$$

Example 5.1.4 (Geometric Sum)**.** How does: $S_n = 1 + a + a^2 + a^3 + \cdots + a^n = \sum_{j=0}^{n} a^j$ work?

If $a = 1$, then $S = n + 1$. If $a \neq 1$, then

$$\begin{aligned} aS_n &= a + a^2 + a^3 + \cdots + a^{n+1} \\ (aS_n - S_n) &= (a-1)S_n = a^{n+1} - 1 \\ S_n &= \frac{a^{n+1}-1}{a-1}. \end{aligned} \tag{5.13}$$

For instance $1 + 2 + 4 + 8 + \cdots + 1024 = 2047$.

Example 5.1.5 (Other useful Sums)**.** *We have the following computations*

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}, \tag{5.14}$$

$$\sum_{j=1}^{n} j^3 = \left(\frac{n(n+1)}{2}\right)^2. \tag{5.15}$$

5.1.3 Double Summations

Corresponding to nested loops in programming languages, there are double summations:

$$\sum_{j=1}^{n}\sum_{k=1}^{m} a_{j,k}.$$

5.2 Limits

5.2.1 Definition

We define x as the **limit** of the sequence (x_n), denoted as:

$$x_n \underset{n\to\infty}{\longrightarrow} x \quad \text{or} \quad \lim_{n\to\infty} x_n = x,$$

if the following condition is satisfied:

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N},\ \ s.t.\ \ \forall n \geq N,\ |x_n - x| < \varepsilon,$$

which reads: for every real number $\varepsilon > 0$, there exists a natural number N such that for all $n \geq N$, the inequality $|x_n - x| < \varepsilon$ holds. In simpler terms, for any given measure of closeness ε, the terms of the sequence eventually become that close to the limit x. The sequence (x_n) is said to **converge** to or **tend to** the limit x.

If a sequence (x_n) converges to a limit x, it is called **convergent**, and x is its unique limit. Otherwise, the sequence is said to be **divergent**. For instance $x_n \underset{n\to\infty}{\longrightarrow} \pm\infty$ is formulated as follows:

$$\forall R > 0,\ \exists N \in \mathbb{N},\ \ s.t.\ \ \forall n \geq N,\ |x_n| > R.$$

5.2.2 Examples

Example 5.2.1. The sequence (x_n) defined by $x_n = \frac{1}{n}$, $n \geq 1$, converges to the limit $x = 0$. This is an example of a non-null sequence tending to 0:

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

Example 5.2.2. The sequence (y_n) defined by $y_n = \frac{\sin(n)}{n}$ also converges to the limit $x = 0$, as the absolute value of the terms approaches zero: we use $|\lim_{n\to\infty} y_n| \leq \lim_{n\to\infty} |y_n|$, thus

$$\lim_{n\to\infty} y_n| = \left|\lim_{n\to\infty} \frac{\sin(n)}{n}\right| \leq \lim_{n\to\infty} \frac{1}{n} = 0 \Longrightarrow \lim_{n\to\infty} y_n = 0.$$

Example 5.2.3. The sequence (z_n) defined by $z_n = \cos(n\pi)$ has no limit as

$$z_{2n} = 1 \quad \text{and} \quad z_{2n+1} = -1.$$

Example 5.2.4. The sequence (u_n) defined by $u_n = e^{-n} \underset{n\to\infty}{\longrightarrow} 0$

5.3 Series

5.3.1 Definition

The n**-th partial sum**, S_n, is the sum of the first n terms of the sequence $(a_k)_k$ is given by:

$$S_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k.$$

A **series** is the sum of all the terms:

$$S = a_1 + a_2 + a_3 + \cdots = \sum_{n=1}^{\infty} a_n.$$

A series is said to be **convergent** (or that it **converges**) if the sequence $(S_n)_n$ converge in the sense of Section 5.2. series that does not converge is said to be **divergent** or to **diverge**.

Note that

$$\sum_{n=1}^{\infty} a_n \text{ converges} \Longrightarrow a_n \underset{n\to\infty}{\longrightarrow} 0.$$

Thus, if a_n does not tend to zero, then $\sum_{n=1}^{\infty} a_n$ diverge !

5.3.2 Examples of convergent and divergent series

(a) **Harmonic series (divergent):** The reciprocals of positive integers produce a divergent series $a_n = \frac{1}{n}$:

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \cdots \to \infty.$$

(b) **Alternating harmonic series (convergent):** Alternating the signs of the reciprocals of positive integers produces a convergent series $a_n = \frac{(-1)^n}{n}$:

$$\frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots = \ln(2).$$

(c) **Reciprocals of prime numbers (divergent):** The reciprocals of prime numbers produce a divergent series $a_n = \frac{1}{p_n}$, where p_n is the n-th prime number:

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \cdots \to \infty.$$

(d) **Reciprocals of triangular numbers (convergent):** The reciprocals of triangular numbers produce a convergent series with $a_n = \frac{2}{n(n+1)}$, see (5.12):

$$\frac{1}{1} + \frac{1}{3} + \frac{1}{6} + \frac{1}{10} + \frac{1}{15} + \frac{1}{21} + \cdots = 2.$$

(e) **Reciprocals of factorials (convergent):** The reciprocals of factorials produce a convergent series $a_n = \frac{1}{n!}$:

$$\frac{1}{1} + \frac{1}{1} + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \cdots = e. \tag{5.31}$$

(f) **Reciprocals of square numbers (convergent):** The reciprocals of square numbers produce a convergent series (the Basel problem) with $a_n = \frac{1}{n^2}$:

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \cdots = \frac{\pi^2}{6}.$$

(g) **Reciprocals of powers of $n > 1$ (convergent):** An application of (5.13) gives that the reciprocals of powers of any $x > 1$ produce a convergent series with $a_n = \frac{1}{x^n}$:

$$\frac{1}{1} + \frac{1}{x} + \frac{1}{x^2} + \frac{1}{x^3} + \frac{1}{x^4} + \cdots = \frac{x}{x-1}.$$

In particular the reciprocals of powers of 2 produce a convergent series is

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \cdots = 2.$$

(h) **Reciprocals of Fibonacci numbers (convergent):** The reciprocals of Fibonacci numbers produce a convergent series with $1/a_n$ = Fibonacci sequence:

$$\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{8}+\cdots=\psi\approx 3.359885.$$

This convergence can be analyzed using (5.11):

$$\frac{1}{a_n}=\frac{\sqrt{5}}{\varphi^n-(-\varphi)^{-n}}.$$

5.3.3 Methods for determining series convergence or divergence

1) **Absolute convergence test.** Let $a_n \in \mathbb{R}$, Then

$$\sum_{n=1}^{\infty}|a_n| \text{ converges} \Longrightarrow \sum_{n=1}^{\infty} a_n \text{ converges.}$$

2) **Comparison test.** Let $a_n, b_n > 0$ satisfying one of the following conditions:

$$a_n \le b_n, \quad \text{or} \quad \lim_{n\to\infty}\frac{a_n}{b_n}=0.$$

Then

$$\sum_{n=1}^{\infty} b_n \text{ converges} \Longrightarrow \sum_{n=1}^{\infty} a_n \text{ converges.}$$

Conversely,

$$\sum_{n=1}^{\infty} a_n \text{ diverge} \Longrightarrow \sum_{n=1}^{\infty} b_n \text{ diverge.}$$

3) **Equivalence test.** Let a_n, b_n satisfying one of the following conditions:

$$0 \le b_n \le a_n \le b_{n+1}, \quad \text{or} \quad \lim_{n\to\infty}\frac{a_n}{b_n}=c, \text{ for some } 0<c<\infty.$$

Then

$$\sum_{n=1}^{\infty} a_n \Longleftrightarrow \sum_{n=1}^{\infty} b_n \text{ converges.}$$

4) **Integral test.** If $a_n = f(n)$, where $f(x)$ is positive, continuous, and monotonically decreasing, then

$$\int_1^{\infty} f(x)\,dx \text{ converges} \Longleftrightarrow \sum_{n=1}^{\infty} a_n \text{ converges.}$$

The notion of convergence of $\int_1^{\infty} f(x)\,dx$ will be seen Chapter 7.

5) **Ratio Test.** Let

$$r=\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right|.$$

- If $r<1$, the series is absolutely convergent.

- If $r > 1$, the series diverges.
- If $r = 1$, the test is inconclusive.

6) **Root test (or n-th root test).** Let

$$r = \limsup_{n\to\infty} \sqrt[n]{|a_n|}.$$

- If $r < 1$, the series converges.
- If $r > 1$, the series diverges.
- If $r = 1$, the test is inconclusive.

7) **Alternating series test.** Let $a_n \geq 0$. Then

$$a_n \text{ is monotonically decreasing to } 0 \Longrightarrow \sum_{n=1}^{\infty} (-1)^n \, a_n \text{converge.}$$

The above criteria help to explain the convergence and divergence of the examples in Subsection 5.3.2.

Chapter 6

Limits, Continuity and Derivability of Functions

6.1 Limit of a function in a point

A real valued function $f$is said to have a limit l at a point a, if any sequence $(x_n)_{n\in\mathbb{N}}$ of points in the domain D that converges to a, the corresponding sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to l. In mathematical notation:

$$\forall (x_n)_{n\in\mathbb{N}} \subset D, \quad \lim_{n\to\infty} x_n = a \Longrightarrow \lim_{n\to\infty} f(x_n) = l.$$

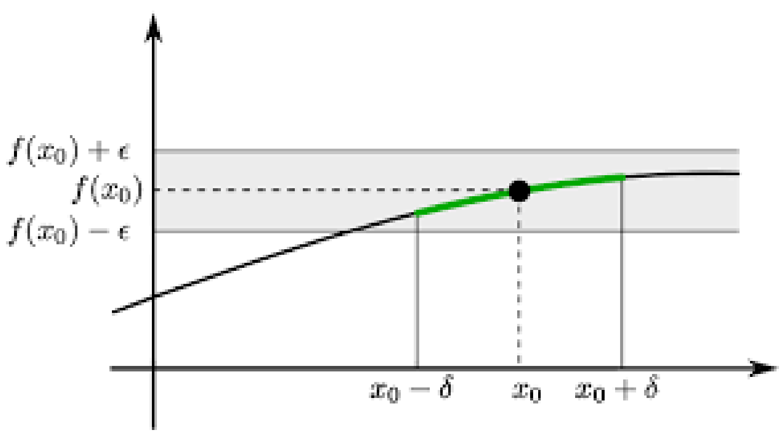
We denote l by $\lim_{x\to a} f(x)$.

6.1.1 Continuity a function in a point

Given a function $f : D \to \mathbb{R}$ and an element $x_0 \in D$, f is said to be continuous at the point x_0 if $\lim_{x\to x_0)} f(x) = f(x_0)$. In mathematical notation: for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in D$:

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

One can alternatively define continuity by requiring $\lim_{x\to a_-} f(x) = \lim_{x\to a, x<a} f(x)$ and

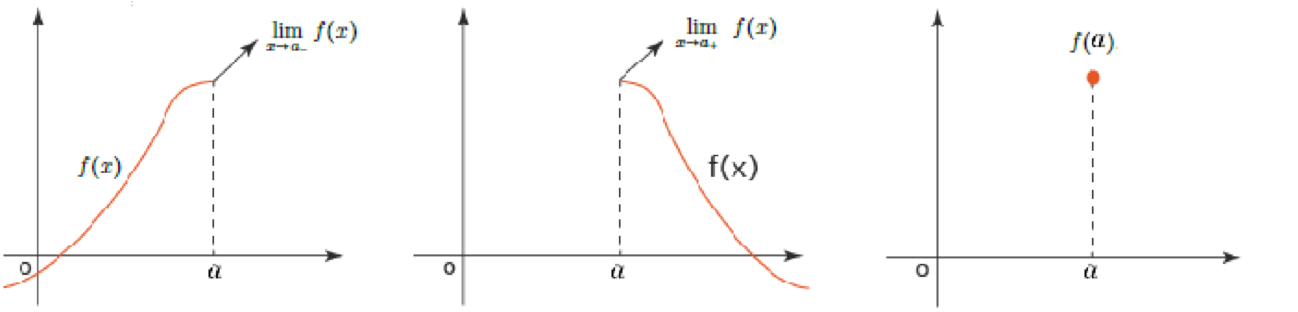


$\lim_{x\to a_+} f(x) = \lim_{x\to a, x>a} f(x)$ exist and are equal. We then say that f is continuous at x_0 and we have

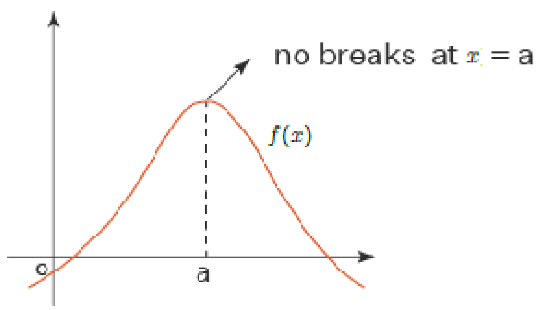
$$\lim_{x\to a} f(x) = \lim_{x\to a_-} f(x) = \lim_{x\to a_+} f(x) = f(x_0).$$

A function is said to be continuous on $D = [a, b]$ if it is continuous at each point of D.

Example 6.1.1. Given the graph of $f(x)$, shown below, determine if $f(x)$ is continuous at $x = -2$, $x = 0$, and $x = 3$. The graph represents a function with two distinct pieces. The first piece, for $x < -2$, starts at $(-4, -2)$ and increases to $(-2, 2)$. The second piece starts with an open dot at $(-2, -1)$, increases, then decreases to $(3, 0)$, and continues increasing. It has open dots at $(-2, -1)$ and $(3, 0)$, and closed dots at $(0, 1)$ and $(3, -1)$. The function is continuous except at two points where the pencil must be lifted.



These three together will make the function f continous at x = a



$\lim_{x\to a} f(x) = f(a)$ $\Longrightarrow$ f is continuous at x = a

Example 6.1.2. Determine where the function below is defined and continuous.

$$h(t) = \frac{8t^t - 16t^2 - 120t}{t^2 - 2t - 15}$$

Solution The numerator is of the form $t(4t-20)(2t+6)$, and denominator is of the form $(t+3)(t-5)$ function h has no problem of definition:

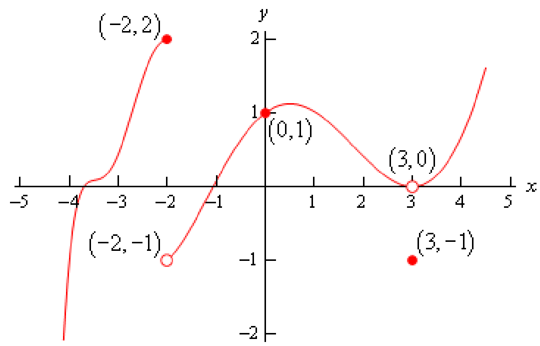
$$h(t) = \frac{t(4t-20)(2t+6)}{(t+3)(t-5)} = 8t.$$

Thus h is well defined and continuous on $\mathbb{R}$.

6.1.2 Fact

A consequence of continuity is the following fact. If $f(x)$ is continuous at $x = b$ and $\lim_{x\to a} g(x) = b$, then:

$$\lim_{x\to a} f(g(x)) = f\left(\lim_{x\to a} g(x)\right).$$



Example 6.1.3. Evaluate the following limit: $\lim_{x\to 0} e^{\sin x}$

6.2 Intermediate value theorem (nice consequence of continuity)

Suppose that $f(x)$ is continuous on $[a, b]$ and let M be any number between $f(a)$ and $f(b)$. Then there exists a number c such that:

$$a < c < b \quad \text{and} \quad f(c) = M.$$

The intermediate value theorem says is that a continuous function will take on all values between $f(a)$ and $f(b)$.

Example 6.2.1. Check that $p(x) = 2x^3 - 5x^2 - 10x + 5$ has a root somewhere in the interval $[-1, 2]$.

Example 6.2.2. If possible, determine if $f(x) = 20\sin(x+3)\cos\left(\frac{x^2}{2}\right)$ takes the following values in the interval $[0, 5]$.

(i) Does $f(x) = 10$?

(ii) Does $f(x) = -10$?

6.3 Differentiation of a function

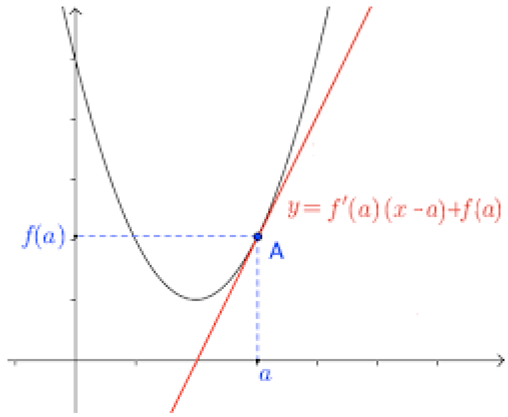
A function $f : I \to \mathbb{R}$ is differentiable at a point a if its domain I contains an open interval containing a, and the limit

$$L = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

exists. This means that, for every positive real number ε, there exists a positive real number δ such that, for every h such that $0 < |h| < \delta$, then $f(a+h)$ is defined, and

$$\left| L - \frac{f(a+h) - f(a)}{h} \right| < \varepsilon,$$

where the vertical bars denote the absolute value. The value L is denoted by $f'(x)$. The derivative tells us the slope of a function at any point. For example:



- The slope of a constant function (like 3) is always 0.
- The slope of a linear function like $2x$ is 2.

6.3.1 Common functions and their derivatives

Here are useful rules to help you work out the derivatives of many functions. Note: the little mark $'$ means derivative of, and f and g are functions.

Function	**Derivative**
Constant c	0
Line ax, $a \in \mathbb{R}$	a
Power function x^α	$\alpha x^{\alpha-1}$
Exponential a^x, $a > 0$	$\ln(a) \cdot a^x$
Exponential function e^x	e^x
Logarithm $\ln(x)$	$\frac{1}{x}$
Sine, $\sin(x)$	$\cos(x)$
Cosine, $\cos(x)$	$-\sin(x)$
Tangent, $\tan(x)$	$\sec^2(x)$
Inverse Sine, $\sin^{-1}(x)$	$\frac{1}{\sqrt{1-x^2}}$
Inverse Cosine, $\cos^{-1}(x)$	$-\frac{1}{\sqrt{1-x^2}}$
Inverse Tangent, $\tan^{-1}(x)$	$\frac{1}{1+x^2}$
Rule	**Derivative**
Multiplication by a constant: cf	cf'
Sum rule: $f + g$	$f' + g'$
Difference rule: $f - g$	$f' - g'$
Product rule: fg	$f \cdot g' + f' \cdot g$
Quotient rule: $\frac{f}{g}$	$\frac{f' \cdot g - g' \cdot f}{g^2}$
Power rule: f^α	$\alpha \, f' \, f^{\alpha-1}$
Composition rule $f(g(x))$	$f'(g(x)) \cdot g'(x)$

6.3.2 Examples of differentiation

Example 1: Power rule

What is the derivative of $\frac{d}{dx}x^3$? Using the Power rule where $n = 3$: $\frac{d}{dx}x^n = nx^{n-1}$. So,

$$\frac{d}{dx}x^3 = 3x^{3-1} = 3x^2.$$

Example 2: Derivative of $\frac{1}{x} = x^{-1}$

Using the Power rule where $n = -1$: $\frac{d}{dx}x^n = nx^{n-1}$. So,

$$\frac{d}{dx}x^{-1} = -1 \cdot x^{-1-1} = -x^{-2} = -\frac{1}{x^2}.$$

Example 3: Multiplication by a constant

What is $\frac{d}{dx}5x^3$? Using the Power rule:

$$\frac{d}{dx}x^3 = 3x^2.$$

So:

$$\frac{d}{dx}5x^3 = 5 \cdot \frac{d}{dx}x^3 = 5 \cdot 3x^2 = 15x^2.$$

Example 5: Sum rule

What is the derivative of $x^2 + x^3$? The Sum Rule says:

$$\frac{d}{dx}(f+g) = \frac{d}{dx}f + \frac{d}{dx}g.$$

Using the Power Rule $\frac{d}{dx}x^2 = 2x, \quad \frac{d}{dx}x^3 = 3x^2$. So,

$$\frac{d}{dx}(x^2 + x^3) = 2x + 3x^2.$$

Example 6: Difference rule

What is $\frac{d}{dv}(v^3 - v^4)$? The Difference Rule says $\frac{d}{dv}(f-g) = \frac{d}{dv}f - \frac{d}{dv}g$. So,

$$\frac{d}{dv}(v^3 - v^4) = 3v^2 - 4v^3.$$

6.4 Problems

Find the domain, the continuity and the derivative of the given function.

1. $f(x) = 6x^3 - 9x + 4$

Solution:

2. $y = 2t^4 - 10t^2 + 13t$

Solution:

3. $g(z) = 4z^7 - 3z - 7 + 9z$

Solution:

4. $h(y) = y^{-4} - 9y^{-3} + 8y^{-2} + 12$
Solution:

5. $y = \sqrt{x} + 8\sqrt[3]{x} - 2\sqrt[4]{x}$
Solution:

6. $f(x) = 10\sqrt[5]{x^3} - \sqrt{x^7} + 6\sqrt[3]{x^8} - 3$
Solution:

7. $f(t) = 4t^{-1} - 6t^3 + 8t^5$
Solution:

8. $R(z) = 6\sqrt[3]{z} + \frac{1}{8}z^4 - \frac{1}{3}z^{10}$
Solution:

9. $z = x(3x^2 - 9)$
Solution:

10. $g(y) = (y - 4)(2y + y^2)$
Solution:

11. $h(x) = 4x^3 - 7x + 8x$
Solution:

12. $f(y) = y^5 - 5y^3 + 2yy^3$
Solution:

Determine where the function is not changing.

1. $f(x) = x^3 + 9x^2 - 48x + 2$

Solution:

2. $y = 2z^4 - z^3 - 3z^2$

Solution:

Find the tangent line to the given function.

1. $g(x) = 16x - 4\sqrt{x}$ at $x = 4$

Solution:

2. $f(x) = 7x^4 + 8x - 6 + 2x$ at $x = -1$

Solution:

Determine where the function is increasing and decreasing.

1. $h(z) = 6 + 40z^3 - 5z^4 - 4z^5$

Solution:

2. $R(x) = (x + 1)(x - 2)^2$

Solution:

Tangent line parallel to a given line

Determine where, if anywhere, the tangent line to $f(x) = x^3 - 5x^2 + x$ is parallel to the line $y = 4x + 23$.

Solution:

Solve each of the following inequalities

1. $u^2 + 4u \geq 21$

Solution:

2. $x^2 + 8x + 12 < 0$

Solution:

3. $4t^2 \leq 15 - 17t$

Solution:

4. $z^2 + 34 > 12z$

Solution:

5. $y^2 - 2y + 1 \leq 0$

Solution:

6. $t^4 + t^3 - 12t^2 < 0$

Solution:

Simplify each expression

1. $-\frac{36x^3}{42x^2}$

Solution:

2. $\frac{16r^2}{16r^3}$

Solution:

3. $\frac{16p^2}{28p}$

Solution:

4. $\frac{32n^2}{24n}$
Solution:

5. $-\frac{70n^2}{28n}$
Solution:

6. $\frac{15n}{30n^3}$
Solution:

7. $\frac{2r-4}{r-2}$
Solution:

8. $\frac{45}{10a-10}$
Solution:

9. $\frac{x-4}{3x^2-12x}$
Solution:

10. $\frac{15a-3}{24}$
Solution:

11. $\frac{v-5}{v^2-10v+25}$
Solution:

12. $\frac{x+6}{x^2+5x-6}$
Solution:

Chapter 7

Simple and Generalized Integrals

7.1 Definite Integrals

7.1.1 Riemann Integrals

The Riemann integral is a foundational concept in calculus, providing a method for calculating the area under a curve. The approach involves approximating this area by dividing it into a series of rectangles and summing their areas.

Let $[a, b]$ be a closed interval. A subdivision (or partition) of $[a, b]$ of size n is a finite sequence of points $\{x_0, x_1, x_2, \ldots, x_n\}$ such that

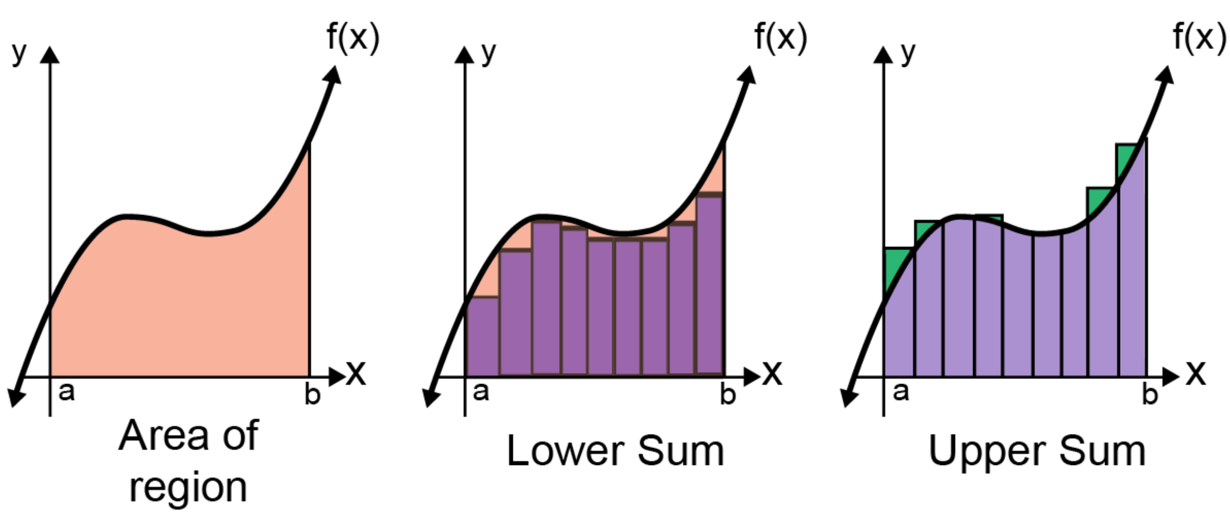
$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

Each $[x_{i-1}, x_i]$, for $i = 1, 2, \ldots, n$, is called a subinterval of the partition, it has equal or varying widths width $\Delta x_i = x_{i+1} - x_i$. Note that Δx_i depends on n and that $\Delta x_i \underset{n\to\infty}{\longrightarrow} 0$.

To approximate the integral of a continuous function $f : [a, b] \to \mathbb{R}$, rectangles are constructed with basis the subintervals $[x_{i-1}, x_i]$, the height of each is determined by $f(x_{i-1})$ the value of the function at the left endpoint (one may also choose the right endpoint, or midpoint ...). Then, the sum of the areas of these rectangles equals

$$S_n = \sum_{i=1}^{n} f(x_i)\, \Delta x_i.$$

As $n \to \infty$, this sum converges to a value L called the *Riemann integral*:



$$L = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\, \Delta x_i.$$

This value is denoted by $L = \int_a^b f(x)\, dx$. We will see that this method is widely used to visualize and compute the integrals, defined as in the following sections, as it connects the geometric idea of area with the analytical process of summation.

Example 7.1.1 (Approximation of $I = \int_0^1 x^2\, dx$)**.** With the subdivision

$$x_i = \frac{i}{n}, \quad 0 \le i \le n,$$

the Riemann sum corresponding to I is

$$S_n = \sum_{i=1}^{n} f(x_i)\Delta x_i = \sum_{i=1}^{n} \left(\frac{i}{n}\right)^2 \frac{i+1-i}{n} = \frac{1}{n^3} \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6\, n^3} = \frac{(n+1)(2n+1)}{6\, n^2},$$

see Formula 5.14

The approximation of the integral $\int_0^1 x^2\, dx$ with $n = 20$ is:

$$S_{20} = \frac{21 \times 41}{6 \times 20^2} \approx 0.35875.$$

For Comparison, the exact value of the integral is:

$$\int_0^1 x^2\, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3} \approx 0.33.$$

7.1.2 Definite integrals with antiderivatives

Definite integrals often rely on the use of antiderivatives (or primitives) for evaluation. Antiderivatives provide a direct way to evaluate definite integrals through the **Fundamental Theorem of Calculus**:

$$\int_a^b f(x)dx = F(b) - F(a), \quad \text{where } F'(x) = f(x). \tag{7.11}$$

7.1.3 Change of variables method for integrals

The *change of variables method* is an essential technique for simplifying integrals, particularly when dealing with complex or complicated expressions. This method involves substituting a new variable into the integral to make the computation more manageable. Below is the procedure for performing a change of variables on definite integrals.

Steps to perform change of variables

Let $g : [a,b] \to [c,d]$ be continuous bijection with inverse $g^{-1} : [c,d] \to [a,b]$. Note that $[c,d] = [g(a), g(b)]$ or $[c,d] = [g(b), g(a)]$, depending on the monotonicity of g (we know that a continuous injective function is strictly monotonic). Then, for a continuous function $f : [a,b] \to \mathbb{R}$, the integral

$$I = \int_c^d f\big(g^{-1}(x)\big)\, dx$$

can be computed by these steps:

(a) **Choose the substitution**: Define a new variable $x = g(u)$;

(b) **Calculate the derivative**: Compute the derivative $g'(u)$; If the original integral is from $x = a$ to $x = b$, then the new limits will be $g(a)$ and $g(b)$.

(c) **Substitute into the integral**: Replace x by $g(u)$, and the integral becomes

$$I = \int_c^d f\big(g^{-1}(x)\big)dx = \int_a^b f(u)\, |g'(u)| du.$$

(d) **Simplify the integral**: Perform any simplification to make the integral easier to solve.

(e) **Perform the integration**: Evaluate the integral in the new variable u.

(f) **Revert to the original variable (if needed)**: If you need the result in terms of the original variable, substitute back to get the final answer.

Example 7.1.2. Let's look at an example of a definite integral using the change of variables method:

$$I = \int_0^1 x\sqrt{1+x^2}\, dx.$$

(a) **Choose the substitution**: We select $u = 1 + x^2,\ x \in [0,1] \iff x = g(u) = \sqrt{u-1},\ u \in [1,2]$. This substitution simplifies the square root term.

(b) **Calculate the derivative**: $g'(u) = \frac{1}{2\sqrt{u-1}}$.

(c) **Substitute into the integral**:The original integral becomes:

$$I = \int_1^2 \sqrt{u-1}\,\sqrt{u}\,\frac{1}{2\sqrt{u-1}}\, du.$$

(d) **Simplify the integral**: We can now simplify the integral to:

$$I = \frac{1}{2}\int_1^2 \sqrt{u}\, du.$$

(e) **Perform the integration**: Integrating $u^{1/2}$ gives

$$I = \frac{1}{2}\left[\frac{2}{3}u^{3/2}\right]_1^2 = \frac{1}{3}[2^{3/2} - 1].$$

7.2 Integration by Parts

The *Integration by parts* is a method based on the product rule of differentiation. It is useful for integrating the product of two functions. The formula is derived from the following identity:

$$\frac{d}{dx}\big(u(x)v(x)\big) = u'(x)v(x) + u(x)v'(x)$$

By rearranging this, we get:

$$u(x)v(x) = \int u'(x)v(x)\,dx + \int u(x)v'(x)\,dx$$

Thus, the integration by parts formula is:

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int v(x)u'(x)\,dx$$

Steps to apply integration by parts

(a) Choose $u(x)$ and $v'(x)$ from the integrand. The function $u(x)$ should be chosen to simplify when differentiated, and $v'(x)$ should be easy to integrate.

(b) Compute $u'(x)$ (the derivative of $u(x)$) and $v(x)$ (the antiderivative of $v'(x)$).

(c) Substitute into the integration by parts formula.

(d) Simplify and integrate the remaining terms.

Example 7.2.1. Consider the integral:

$$\int_1^5 xe^x\,dx.$$

(a) Let $u = x$ (so $u'(x) = 1$) and $v'(x) = e^x$ (so $v(x) = e^x$).

(b) Using the formula:

$$\int_1^5 xe^x\,dx = \Big[xe^x\Big]_1^5 - \int_1^5 e^x\,dx$$

(c) The remaining integral is $\int_1^5 e^x\,dx = e^5 - e^1$.

(d) Therefore, the result is:

$$\int_1^5 xe^x\,dx = 5e^5 - e^1 - [e^5 - e^1] = 5e^5 - e^5 = 4e^5.$$

This shows how integration by parts can simplify an integral involving a product of functions.

7.3 Taylor formula with integral remainder

In calculus, *Taylor's theorem* provides an approximation of a function that is k-times differentiable around a given point using a polynomial of degree k, known as the k-th-order

Taylor polynomial. More precisely, let I be an open interval and $f : I \to \mathbb{R}$ be a function that has $n+1$ continuous derivatives in some neighborhood of $a \in I$. Taylor's theorem asserts that for any x in this neighborhood, $f(x)$ is represented by

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x),$$

where the remainder term is the integral

$$R_n(x) := \frac{1}{n!}\int_a^x f^{(n+1)}(t)\,(x-t)^n\,dt.$$

Example 7.3.1. Consider $f(x) = e^x$, $x \in \mathbb{R}$. Taylor's expansion of f around 0 gives that for any $x \in \mathbb{R}$

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + R_n(x),$$

with

$$R_n(x) = \frac{1}{n!}\int_0^x e^t\,(x-t)^n dt = \frac{x^{n+1}}{n!}\int_0^1 e^{xt}\,(1-t)^n\,dt.$$

Observing that $e^{xt} \le e^{|x|}$ for any $t \in [0,1]$, we deduce

$$\begin{aligned} |R_n(x)| &= \frac{|x|^{n+1}}{n!}\int_0^1 e^{xt}\,(1-t)^n\,dt \\ &\le \frac{|x|^{n+1}e^{|x|}}{n!}\int_0^1 (1-t)^n\,dt = \frac{|x|^{n+1}e^{|x|}}{(n+1)!} \underset{n\to\infty}{\longrightarrow} 0, \end{aligned}$$

we deduce that

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} \underset{n\to\infty}{\longrightarrow} e^x.$$

We have then recovered formula (5.31).

7.4 Generalized Integrals

This principle of this section is to handle infinite intervals or points of discontinuity by introducing limits to define integrals. More precisely, *Generalized (or improper) integrals* extend the concept of definite integrals to cases where the limits of integration or the integrand do not meet the standard requirements of continuity or boundedness. For instance, how to handle the integral

$$\int_1^\infty \frac{1}{x^2}dx$$

Why are generalized integrals needed? Many real-world problems (e.g., in physics, probability) require integrating functions over infinite intervals or functions with singularities.

Definition 7.4.1. *If f is a real-valued function defined on some open interval I, then the generalized integral of f on I is of two types*

1) If $I =]a,b]$ is a finite interval, and if f has a singularity in a, then the generalised integral $\int_a^b f(x)dx$ is defined if

$$\lim_{\epsilon \to 0} \int_{a+\epsilon}^{b} f(x)dx \text{ is finite.} \tag{7.41}$$

2) If $I = [a, \infty[$ is an infinite interval, then the generalised integral $\int_a^\infty f(x)dx$ is defined if

$$\int_a^\infty f(x)dx = \lim_{b\to\infty} \int_a^b f(x)dx \text{ is finite.}$$

If f is non negative, we denote $\int_I f(x)dx < \infty$ if the generalised integral converges.

Example 7.4.2. *a) f is not defined in a. For $]a,b] =]0,1]$ and $f(x) = 1/\sqrt{x}$, we have*

$$\int_0^1 \frac{1}{\sqrt{x}}dx = \lim_{\epsilon\to 0^+} \int_\epsilon^1 \frac{1}{\sqrt{x}}dx = \lim_{\epsilon\to 0^+} \left[2\sqrt{x}\right]_\epsilon^1 = 2.$$

b) f is defined on an infinite interval. For $f(x) = 1/x^2$, we have

$$\int_1^\infty \frac{1}{x^2}dx = \lim_{b\to\infty} \int_1^b \frac{1}{x^2}dx = \lim_{b\to\infty} \left[-\frac{1}{x}\right]_1^b = 1.$$

For $f(x) = e^{-x}$, we have

$$\int_1^\infty e^{-x}dx = \lim_{b\to\infty} \int_1^b e^{-x}dx = \lim_{b\to\infty} \left[-e^{-x}\right]_1^b = e^{-1}.$$

7.4.1 Removing the singularities

When the function has a discontinuity at a point in the integration range:

$$\int_a^b f(x)dx, \quad f(x) \text{ undefined at } c \in [a,b] \tag{7.42}$$

Split into two parts:

$$\int_a^b f(x)dx = \lim_{t\to c^-} \int_a^t f(x)dx + \lim_{t\to c^+} \int_t^b f(x)dx \tag{7.43}$$

7.4.2 Some applications of generalized integrals

A. Physics: Electric Fields The electric field generated by a point charge can involve improper integrals. For example:

$$\int_0^\infty \frac{1}{(x^2 + a^2)^{3/2}}dx. \tag{7.44}$$

Using substitution $x = a\tan\theta$, the integral simplifies, and the antiderivative leads to a finite value.

B. Probability: Gaussian Distribution The Gaussian probability density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{7.45}$$

The total probability over $(-\infty, \infty)$ is:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{7.46}$$

The antiderivative involves advanced techniques base on the error function

$$\text{erf}(x) = \int_{-\infty}^{x} f(u)du, \ x \in \mathbb{R}.$$

7.4.3 Convergence and divergence of Riemann integrals

Riemann integrals I_a and J_a are defined by the values

$$I_a := \int_0^1 \frac{1}{x^a} dx \ \text{ and } \ J_a := \int_1^{\infty} \frac{1}{x^a} dx.$$

Do not mix the label of the latter Riemann integrals with the concept of the inteegrals obtrained by the Riemann sums in Subsection 7.1.1. It is easy to prove

$$I_a \text{ converges} \iff a < 1 \quad \text{and} \quad J_a \text{ converges} \iff a > 1.$$

7.4.4 Tests for convergence

Let $f, g : I \to \mathbb{R}_+$. The singularity is denoted by l, where

$$l = a \text{ is } I =]a, b], -\infty < a < b < \infty, \quad \text{or} \quad l = \infty \text{ if } I = [a, \infty[.$$

(i) **Comparison test:** If $0 \le f \le g$, or if $\lim_{x \to l} \frac{f(x)}{g(x)} = 0$, then

$$\int_I g(x)dx < \infty \implies \int_I f(x)dx < \infty.$$

(ii) **Equivalence test:** If $0 \le f \le g$ or if $\lim_{x \to l} \frac{f(x)}{g(x)} = c \in]0, \infty[$, then

$$\int_I g(x)dx < \infty \iff \int_I f(x)dx < \infty.$$

7.4.5 Problems

1) Apply the comparison test to

$$\int_1^{\infty} \frac{\ln x}{x^2} dx \quad \text{compared to} \quad \int_1^{\infty} \frac{1}{x^{3/2}} dx.$$

Solution: Both function $f(x) = \ln(x)/x^2$ and $g(x) = x^{-3/2}$ are nonnegative on $[1, \infty[$. Since

$$\lim_{x\to\infty} f(x)/g(x) = 0$$

and since the Riemann integral $J_{3/2} = \int_1^{\infty} \frac{1}{x^{3/2}} dx$ converges, we deduce that $\int_1^{\infty} \frac{\ln x}{x^2} dx < \infty$ also converge

2) Evaluate $I = \int_2^{\infty} \frac{1}{x(x+1)} dx$.
Solution: Using $\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$, we have

$$\begin{aligned} I &= \lim_{a\to+\infty} \Big[\int_2^a \Big(\frac{1}{x} - \frac{1}{x+1} \Big) dx \Big] = \lim_{a\to+\infty} \Big[\int_2^a \frac{1}{x} dx - \int_2^a \frac{1}{x+1} dx \Big] \\ &= \lim_{a\to+\infty} \Big[\int_2^a \frac{1}{x} dx - \int_3^{a+1} \frac{1}{x} dx \Big] = \lim_{a\to+\infty} \Big[\ln(2/3) - \ln((a+1)/a) \Big] \\ &= \ln(2/3) \end{aligned}$$

3) Show that $\int_0^1 \frac{\ln x}{x} dx$ diverges.

Solution: With $F(x) = \ln(x)$, note that

$$\frac{\ln(x)}{x} = F'(x)F(x) = G'(x), \text{ where } G(x) = \frac{1}{2} F^2(x),$$

which gives

$$\int_0^1 \frac{\ln x}{x} dx = \int_0^1 G'(x) dx = \lim_{a\to 0,\, a>0} \Big[G(x) \Big]_a^1 = G(1) - \lim_{a\to 0,\, a>0} G(a) = -\infty.$$

4) Prove that $I := \int_0^{\infty} \frac{1}{(1+x^2)^2} dx$ converges and find its value.

Solution: The integral I is obviously convergent since

$$0 < I \le \int_0^{\infty} \frac{1}{1+x^2} dx = [\arctan(x)]_0^{\infty} = \frac{\pi}{2} < \infty.$$

Using the substitution $x = \tan(\theta)$, we have that if $x \to 0$, then $\theta \to 0$ and if $x \to +\infty$, then $\theta \to \frac{\pi}{2}$. Moreover,

$$dx = \tan'(\theta) d\theta = \frac{1}{\cos(\theta)^2} d\theta, \quad (1+x^2)^2 = (1+\tan(\theta)^2)^2 = \frac{1}{\cos(\theta)^4}.$$

We deduce

$$I = \int_0^{\frac{\pi}{2}} \cos(\theta)^4 \, \frac{1}{\cos(\theta)^2} d\theta = \int_0^{\frac{\pi}{2}} \cos(\theta)^2 d\theta = \int_0^{\frac{\pi}{2}} \frac{1 + \cos(2\theta)}{2} d\theta = \frac{\pi}{4}.$$