CHAPTER 1

Basic Probability Theory

PART I: THEORY

It is assumed that the reader has had a course in elementary probability. In this chapter we discuss more advanced material, which is required for further developments.

1.1 OPERATIONS ON SETS

Let $\mathcal{S}$ denote a **sample space**. Let E_1, E_2 be subsets of $\mathcal{S}$. We denote the **union** by $E_1 \cup E_2$ and the intersection by $E_1 \cap E_2$. $\bar{E} = \mathcal{S} - E$ denotes the **complement** of E. By DeMorgan's laws $\overline{E_1 \cup E_2} = \bar{E}_1 \cap \bar{E}_2$ and $\overline{E_1 \cap E_2} = \bar{E}_1 \cup \bar{E}_2$.

Given a sequence of sets $\{E_n, n \geq 1\}$ (finite or infinite), we define

$$\sup_{n \geq 1} E_n = \bigcup_{n \geq 1} E_n, \quad \inf_{n \geq 1} E_n = \bigcap_{n \geq 1} E_n. \tag{1.1.1}$$

Furthermore, $\liminf_{n \to \infty}$ and $\limsup_{n \to \infty}$ are defined as

$$\liminf_{n \to \infty} E_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} E_k, \quad \limsup_{n \to \infty} E_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} E_k. \tag{1.1.2}$$

If a point of $\mathcal{S}$ belongs to $\limsup_{n \to \infty} E_n$, it belongs to infinitely many sets E_n. The sets $\liminf_{n \to \infty} E_n$ and $\limsup_{n \to \infty} E_n$ always exist and

$$\liminf_{n \to \infty} E_n \subset \limsup_{n \to \infty} E_n. \tag{1.1.3}$$

Examples and Problems in Mathematical Statistics, First Edition. Shelemyahu Zacks.

If $\liminf_{n\to\infty} E_n = \limsup_{n\to\infty} E_n$, we say that a limit of $\{E_n, n \geq 1\}$ exists. In this case,

$$\lim_{n\to\infty} E_n = \liminf_{n\to\infty} E_n = \limsup_{n\to\infty} E_n. \tag{1.1.4}$$

A sequence $\{E_n, n \geq 1\}$ is called **monotone increasing** if $E_n \subset E_{n+1}$ for all $n \geq 1$. In this case $\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$. The sequence is **monotone decreasing** if $E_n \supset E_{n+1}$, for all $n \geq 1$. In this case $\lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$. We conclude this section with the definition of a **partition** of the sample space. A collection of sets $\mathcal{D} = \{E_1, \ldots, E_k\}$ is called a finite **partition** of $\mathcal{S}$ if all elements of $\mathcal{D}$ are **pairwise disjoint** and their union is $\mathcal{S}$, i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j$; $E_i, E_j \in \mathcal{D}$; and $\bigcup_{i=1}^{k} E_i = \mathcal{S}$. If $\mathcal{D}$ contains a countable number of sets that are mutually exclusive and $\bigcup_{i=1}^{\infty} E_i = \mathcal{S}$, we say that $\mathcal{D}$ is a countable partition.

1.2 ALGEBRA AND σ-FIELDS

Let $\mathcal{S}$ be a sample space. An algebra $\mathcal{A}$ is a collection of subsets of $\mathcal{S}$ satisfying

$$\begin{aligned} &\textbf{(i)} \quad \mathcal{S} \in \mathcal{A}; \\ &\textbf{(ii)} \quad \text{if } E \in \mathcal{A} \text{ then } \bar{E} \in \mathcal{A}; \\ &\textbf{(iii)} \quad \text{if } E_1, E_2 \in \mathcal{A} \text{ then } E_1 \cup E_2 \in \mathcal{A}. \end{aligned} \tag{1.2.1}$$

We consider $\emptyset = \bar{\mathcal{S}}$. Thus, (i) and (ii) imply that $\emptyset \in \mathcal{A}$. Also, if $E_1, E_2 \in \mathcal{A}$ then $E_1 \cap E_2 \in \mathcal{A}$.

The **trivial algebra** is $\mathcal{A}_0 = \{\emptyset, \mathcal{S}\}$. An algebra $\mathcal{A}_1$ is a subalgebra of $\mathcal{A}_2$ if all sets of $\mathcal{A}_1$ are contained in $\mathcal{A}_2$. We denote this inclusion by $\mathcal{A}_1 \subset \mathcal{A}_2$. Thus, the trivial algebra $\mathcal{A}_0$ is a subalgebra of every algebra $\mathcal{A}$. We will denote by $\mathcal{A}(\mathcal{S})$, the algebra generated by all subsets of $\mathcal{S}$ (see Example 1.1).

If a sample space $\mathcal{S}$ has a finite number of points n, say $1 \leq n < \infty$, then the collection of all subsets of $\mathcal{S}$ is called the **discrete algebra** generated by the elementary events of $\mathcal{S}$. It contains 2^n events.

Let $\mathcal{D}$ be a partition of $\mathcal{S}$ having k, $2 \leq k$, disjoint sets. Then, the algebra generated by $\mathcal{D}$, $\mathcal{A}(\mathcal{D})$, is the algebra containing all the $2^k - 1$ unions of the elements of $\mathcal{D}$ and the empty set.

An algebra on $\mathcal{S}$ is called a **σ-field** if, in addition to being an algebra, the following holds.

(iv) If $E_n \in \mathcal{A}$, $n \geq 1$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}$.

We will denote a σ-field by $\mathcal{F}$. In a σ-field $\mathcal{F}$ the supremum, infinum, limsup, and liminf of any sequence of events belong to $\mathcal{F}$. If $\mathcal{S}$ is finite, the discrete algebra $\mathcal{A}(\mathcal{S})$ is a σ-field. In Example 1.3 we show an algebra that is not a σ-field.

The minimal σ-field containing the algebra generated by $\{(-\infty, x], -\infty < x < \infty\}$ is called the **Borel σ-field** on the real line $\mathbb{R}$.

A sample space $\mathcal{S}$, with a σ-field $\mathcal{F}$, $(\mathcal{S}, \mathcal{F})$ is called a **measurable space**.

The following lemmas establish the existence of smallest σ-field containing a given collection of sets.

Lemma 1.2.1. *Let $\mathcal{E}$ be a collection of subsets of a sample space $\mathcal{S}$. Then, there exists a smallest σ-field $\mathcal{F}(\mathcal{E})$, containing the elements of $\mathcal{E}$.*

Proof. The algebra of all subsets of $\mathcal{S}$, $\mathcal{A}(\mathcal{S})$ obviously contains all elements of $\mathcal{E}$. Similarly, the σ-field $\mathcal{F}$ containing all subsets of $\mathcal{S}$, contains all elements of $\mathcal{E}$. Define the σ-field $\mathcal{F}(\mathcal{E})$ to be the **intersection** of all σ-fields, which contain all elements of $\mathcal{E}$. Obviously, $\mathcal{F}(\mathcal{E})$ is an algebra. QED

A collection $\mathcal{M}$ of subsets of $\mathcal{S}$ is called a **monotonic class** if the limit of any monotone sequence in $\mathcal{M}$ belongs to $\mathcal{M}$.

If $\mathcal{E}$ is a collection of subsets of $\mathcal{S}$, let $\mathcal{M}^*(\mathcal{E})$ denote the smallest monotonic class containing $\mathcal{E}$.

Lemma 1.2.2. *A necessary and sufficient condition of an algebra $\mathcal{A}$ to be a σ-field is that it is a monotonic class.*

Proof. (**i**) Obviously, if $\mathcal{A}$ is a σ-field, it is a monotonic class.

(**ii**) Let $\mathcal{A}$ be a monotonic class.

Let $E_n \in \mathcal{A}, n \geq 1$. Define $B_n = \bigcup_{i=1}^{n} E_i$. Obviously $B_n \subset B_{n+1}$ for all $n \geq 1$. Hence $\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. But $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} E_n$. Thus, $\sup_{n\geq 1} E_n \in \mathcal{A}$. Similarly, $\bigcap_{n=1}^{\infty} E_n \in \mathcal{A}$. Thus, $\mathcal{A}$ is a σ-field. QED

Theorem 1.2.1. *Let $\mathcal{A}$ be an algebra. Then $\mathcal{M}^*(\mathcal{A}) = \mathcal{F}(\mathcal{A})$, where $\mathcal{F}(\mathcal{A})$ is the smallest σ-field containing $\mathcal{A}$.*

Proof. See Shiryayev (1984, p. 139).

The measurable space $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ is the real line and $\mathcal{B} = \mathcal{F}(\mathbb{R})$, called the **Borel measurable space**, plays a most important role in the theory of statistics. Another important measurable space is $(\mathbb{R}^n, \mathcal{B}^n)$, $n \geq 2$, where $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ is the Euclidean n-space, and $\mathcal{B}^n = \mathcal{B} \times \cdots \times \mathcal{B}$ is the smallest σ-field containing $\mathbb{R}^n$, $\emptyset$, and all n-dimensional rectangles $I = I_1 \times \cdots \times I_n$, where

$$I_i = (a_i, b_i], \quad i = 1, \ldots, n, \quad -\infty < a_i < b_i < \infty.$$

The measurable space $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ is used as a basis for probability models of experiments with infinitely many trials. $\mathbb{R}^\infty$ is the space of ordered sequences $\mathbf{x} = (x_1, x_2, \ldots)$, $-\infty < x_n < \infty$, $n = 1, 2, \ldots$. Consider the cylinder sets

$$\mathcal{C}(I_1 \times \cdots \times I_n) = \{\mathbf{x} : x_i \in I_i,\ i = 1, \ldots, n\}$$

and

$$\mathcal{C}(B_1 \times \cdots \times B_n) = \{\mathbf{x} : x_i \in B_i,\ i = 1, \ldots, n\}$$

where B_i are Borel sets, i.e., $B_i \in \mathcal{B}$. The smallest σ-field containing all these cylinder sets, $n \geq 1$, is $\mathcal{B}(\mathbb{R}^\infty)$. Examples of Borel sets in $\mathcal{B}(\mathbb{R}^\infty)$ are

(a) $\{\mathbf{x} : \mathbf{x} \in \mathbb{R}^\infty, \sup_{n \geq 1} x_n > a\}$

or

(b) $\{\mathbf{x} : \mathbf{x} \in \mathbb{R}^\infty, \limsup_{n \to \infty} x_n \leq a\}$.

1.3 PROBABILITY SPACES

Given a measurable space $(\mathcal{S}, \mathcal{F})$, a **probability model** ascribes a countably additive function P on $\mathcal{F}$, which assigns a probability $P\{A\}$ to all sets $A \in \mathcal{F}$. This function should satisfy the following properties.

(A.1) If $A \in \mathcal{F}$ then $0 \leq P\{A\} \leq 1$.

(A.2) $P\{\mathcal{S}\} = 1$. (1.3.1)

(A.3) If $\{E_n, n \geq 1\} \in \mathcal{F}$ is a sequence of **disjoint** sets in $\mathcal{F}$, then

$$P\left\{\bigcup_{n=1}^{\infty} E_n\right\} = \sum_{n=1}^{\infty} P\{E_n\}. \tag{1.3.2}$$

Recall that if $A \subset B$ then $P\{A\} \le P\{B\}$, and $P\{\bar{A}\} = 1 - P\{A\}$. Other properties will be given in the examples and problems. In the sequel we often write AB for $A \cap B$.

Theorem 1.3.1. *Let $(\mathcal{S}, \mathcal{F}, P)$ be a probability space, where $\mathcal{F}$ is a σ-field of subsets of $\mathcal{S}$ and P a probability function. Then*

(i) *if $B_n \subset B_{n+1}$, $n \ge 1$, $B_n \in \mathcal{F}$, then*

$$P\left\{\lim_{n\to\infty} B_n\right\} = \lim_{n\to\infty} P\{B_n\}. \tag{1.3.3}$$

(ii) *if $B_n \supset B_{n+1}$, $n \ge 1$, $B_n \in \mathcal{F}$, then*

$$P\left\{\lim_{n\to\infty} B_n\right\} = \lim_{n\to\infty} P\{B_n\}. \tag{1.3.4}$$

Proof. (**i**) Since $B_n \subset B_{n+1}$, $\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n$. Moreover,

$$P\left\{\bigcup_{n=1}^{\infty} B_n\right\} = P\{B_1\} + \sum_{n=2}^{\infty} P\{B_n - B_{n-1}\}. \tag{1.3.5}$$

Notice that for $n \ge 2$, since $\bar{B}_n B_{n-1} = \emptyset$,

$$\begin{aligned} P\{B_n - B_{n-1}\} &= P\{B_n \bar{B}_{n-1}\} \\ &= P\{B_n\} - P\{B_n B_{n-1}\} = P\{B_n\} - P\{B_{n-1}\}. \end{aligned} \tag{1.3.6}$$

Also, in (1.3.5)

$$\begin{aligned} P\{B_1\} + \sum_{n=2}^{\infty} P\{B_n - B_{n-1}\} &= \lim_{N\to\infty} \left(P\{B_1\} + \sum_{n=2}^{N} (P\{B_n\} - P\{B_{n-1}\}) \right) \\ &= \lim_{N\to\infty} P\{B_N\}. \end{aligned} \tag{1.3.7}$$

Thus, Equation (1.3.3) is proven.

(**ii**) Since $B_n \supset B_{n+1}$, $n \geq 1$, $\bar{B}_n \subset \bar{B}_{n+1}$, $n \geq 1$. $\lim_{n \to \infty} B_n = \bigcap_{n=1}^{\infty} B_n$. Hence,

$$\begin{aligned} P\left(\lim_{n\to\infty} B_n\right) &= 1 - P\left\{\overline{\bigcap_{n=1}^{\infty} B_n}\right\} \\ &= 1 - P\left\{\bigcup_{n=1}^{\infty} \bar{B}_n\right\} \\ &= 1 - \lim_{n\to\infty} P\{\bar{B}_n\} = \lim_{n\to\infty} P\{B_n\}. \end{aligned}$$

QED

Sets in a probability space are called events.

1.4 CONDITIONAL PROBABILITIES AND INDEPENDENCE

The conditional probability of an event $A \in \mathcal{F}$ given an event $B \in \mathcal{F}$ such that $P\{B\} > 0$, is defined as

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}}. \tag{1.4.1}$$

We see first that $P\{\cdot \mid B\}$ is a probability function on $\mathcal{F}$. Indeed, for every $A \in \mathcal{F}$, $0 \leq P\{A \mid B\} \leq 1$. Moreover, $P\{\mathcal{S} \mid B\} = 1$ and if A_1 and A_2 are disjoint events in $\mathcal{F}$, then

$$\begin{aligned} P\{A_1 \cup A_2 \mid B\} &= \frac{P\{(A_1 \cup A_2)B\}}{P\{B\}} \\ &= \frac{P\{A_1B\} + P\{A_2B\}}{P\{B\}} = P\{A_1 \mid B\} + P\{A_2 \mid B\}. \end{aligned} \tag{1.4.2}$$

If $P\{B\} > 0$ and $P\{A\} \neq P\{A \mid B\}$, we say that the events A and B are **dependent**. On the other hand, if $P\{A\} = P\{A \mid B\}$ we say that A and B are **independent** events. Notice that two events are independent if and only if

$$P\{AB\} = P\{A\}P\{B\}. \tag{1.4.3}$$

Given n events in $\mathcal{A}$, namely $A_1, \ldots, A_n$, we say that they are **pairwise** independent if $P\{A_iA_j\} = P\{A_i\}P\{A_j\}$ for any $i \neq j$. The events are said to be independent in triplets if

$$P\{A_iA_jA_k\} = P\{A_i\}P\{A_j\}P\{A_k\}$$

for any $i \neq j \neq k$. Example 1.4 shows that pairwise independence does not imply independence in triplets.

Given n events $A_1, \ldots, A_n$ of $\mathcal{F}$, we say that they are **independent** if, for any $2 \leq k \leq n$ and any k-tuple $(1 \leq i_1 < i_2 < \cdots < i_k \leq n)$,

$$P\left\{\bigcap_{j=1}^{k} A_{i_j}\right\} = \prod_{j=1}^{k} P\{A_{i_j}\}. \tag{1.4.4}$$

Events in an infinite sequence $\{A_1, A_2, \ldots\}$ are said to be **independent** if $\{A_1, \ldots, A_n\}$ are independent, for each $n \geq 2$. Given a sequence of events $A_1, A_2, \ldots$ of a σ-field $\mathcal{F}$, we have seen that

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k.$$

This event means that points w in $\limsup_{n\to\infty} A_n$ belong to infinitely many of the events $\{A_n\}$. Thus, the event $\limsup_{n\to\infty} A_n$ is denoted also as $\{A_n, \text{i.o.}\}$, where i.o. stands for "infinitely often."

The following important theorem, known as the **Borel–Cantelli Lemma**, gives conditions under which $P\{A_n, \text{i.o.}\}$ is either 0 or 1.

Theorem 1.4.1 (Borel–Cantelli). *Let $\{A_n\}$ be a sequence of sets in $\mathcal{F}$.*

(i) *If $\sum_{n=1}^{\infty} P\{A_n\} < \infty$, then $P\{A_n, i.o.\} = 0$.*

(ii) *If $\sum_{n=1}^{\infty} P\{A_n\} = \infty$ and $\{A_n\}$ are independent, then $P\{A_n, i.o.\} = 1$.*

Proof. (**i**) Notice that $B_n = \bigcup_{k=n}^{\infty} A_k$ is a decreasing sequence. Thus

$$P\{A_n, \text{i.o.}\} = P\left\{\bigcap_{n=1}^{\infty} B_n\right\} = \lim_{n\to\infty} P\{B_n\}.$$

But

$$P\{B_n\} = P\left\{\bigcup_{k=n}^{\infty} A_k\right\} \leq \sum_{k=n}^{\infty} P\{A_k\}.$$

The assumption that $\sum_{n=1}^{\infty} P\{A_n\} < \infty$ implies that $\lim_{n\to\infty}\sum_{k=n}^{\infty} P\{A_k\} = 0$.

(**ii**) Since $A_1, A_2, \ldots$ are independent, $\bar{A}_1, \bar{A}_2, \ldots$ are independent. This implies that

$$P\left\{\bigcap_{k=1}^{\infty} \bar{A}_k\right\} = \prod_{k=1}^{\infty} P\{\bar{A}_k\} = \prod_{k=1}^{\infty}(1 - P\{A_k\}).$$

If $0 < x \le 1$ then $\log(1 - x) \le -x$. Thus,

$$\begin{aligned}\log \prod_{k=1}^{\infty}(1 - P\{A_k\}) &= \sum_{k=1}^{\infty} \log(1 - P\{A_k\}) \\ &\le -\sum_{k=1}^{\infty} P\{A_k\} = -\infty\end{aligned}$$

since $\sum_{n=1}^{\infty} P\{A_n\} = \infty$. Thus $P\left\{\bigcap_{k=1}^{\infty} \bar{A}_k\right\} = 0$ for all $n \ge 1$. This implies that $P\{A_n, \text{i.o.}\} = 1$. QED

We conclude this section with the celebrated **Bayes Theorem**.

Let $\mathcal{D} = \{B_i, i \in J\}$ be a partition of $\mathcal{S}$, where J is an index set having a finite or countable number of elements. Let $B_j \in \mathcal{F}$ and $P\{B_j\} > 0$ for all $j \in J$. Let $A \in \mathcal{F}$, $P\{A\} > 0$. We are interested in the conditional probabilities $P\{B_j \mid A\}$, $j \in J$.

Theorem 1.4.2 (Bayes).

$$P\{B_j \mid A\} = \frac{P\{B_j\}P\{A \mid B_j\}}{\sum_{j' \in J} P\{B_{j'}\}P\{A \mid B_{j'}\}}. \tag{1.4.5}$$

Proof. Left as an exercise. QED

Bayes Theorem is widely used in scientific inference. Examples of the application of Bayes Theorem are given in many elementary books. Advanced examples of Bayesian inference will be given in later chapters.

1.5 RANDOM VARIABLES AND THEIR DISTRIBUTIONS

Random variables are finite real value functions on the sample space $\mathcal{S}$, such that measurable subsets of $\mathcal{F}$ are mapped into Borel sets on the real line and thus can be assigned probability measures. The situation is simple if $\mathcal{S}$ contains only a finite or countably infinite number of points.

In the general case, $\mathcal{S}$ might contain non-countable infinitely many points. Even if $\mathcal{S}$ is the space of all infinite binary sequences $w = (i_1, i_2, \ldots)$, the number of points in $\mathcal{S}$ is non-countable. To make our theory rich enough, we will require that the probability space will be $(\mathcal{S}, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-field. A random variable X is a finite real value function on $\mathcal{S}$. We wish to define the distribution function of X, on $\mathbb{R}$, as

$$F_X(x) = P\{w : X(w) \le x\}. \tag{1.5.1}$$

For this purpose, we must require that every Borel set on $\mathbb{R}$ has a measurable inverse image with respect to $\mathcal{F}$. More specifically, given $(\mathcal{S}, \mathcal{F}, P)$, let $(\mathbb{R}, \mathcal{B})$ be Borel measurable space where $\mathbb{R}$ is the real line and $\mathcal{B}$ the Borel σ-field of subsets of $\mathbb{R}$. A subset of $(\mathbb{R}, B)$ is called a Borel set if B belongs to $\mathcal{B}$. Let $X : \mathcal{S} \to \mathbb{R}$. The inverse image of a Borel set B with respect to X is

$$X^{-1}(B) = \{w : X(w) \in B\}. \tag{1.5.2}$$

A function $X : \mathcal{S} \to \mathbb{R}$ is called $\mathcal{F}$-measurable if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$. Thus, a **random variable with respect to** $(\mathcal{S}, \mathcal{F}, P)$ **is an** $\mathcal{F}$**-measurable function on** $\mathcal{S}$. The class $\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}\}$ is also a σ-field, generated by the random variable X. Notice that $\mathcal{F}_X \subset \mathcal{F}$.

By definition, every random variable X has a distribution function F_X. The **probability measure** $P_X\{\cdot\}$ induced by X on $(\mathbb{R}, B)$ is

$$P_X\{B\} = P\{X^{-1}(B)\}, \quad B \in \mathcal{B}. \tag{1.5.3}$$

A distribution function F_X is a real value function satisfying the properties

- **(i)** $\lim_{x \to -\infty} F_X(x) = 0$;
- **(ii)** $\lim_{x \to \infty} F_X(x) = 1$;
- **(iii)** If $x_1 < x_2$ then $F_X(x_1) \le F_X(x_2)$; and
- **(iv)** $\lim_{\epsilon \downarrow 0} F_X(x + \epsilon) = F_X(x)$, and $\lim_{\epsilon \uparrow 0} F(x - \epsilon) = F_X(x-)$, all $-\infty < x < \infty$.

Thus, a distribution function F is right-continuous.

Given a distribution function F_X, we obtain from (1.5.1), for every $-\infty < a < b < \infty$,

$$P\{w : a < X(w) \le b\} = F_X(b) - F_X(a) \tag{1.5.4}$$

and

$$P\{w : X(w) = x_0\} = F_X(x_0) - F_X(x_0-). \tag{1.5.5}$$

Thus, if F_X is continuous at a point x_0, then $P\{w : X(w) = x_0\} = 0$. If X is a random variable, then $Y = g(X)$ is a random variable only if g is $\mathcal{B}$-(Borel) measurable, i.e., for any $B \in \mathcal{B}$, $g^{-1}(B) \in \mathcal{B}$. Thus, if $Y = g(X)$, g is $\mathcal{B}$-measurable and X $\mathcal{F}$-measurable, then Y is also $\mathcal{F}$-measurable. The distribution function of Y is

$$F_Y(y) = P\{w : g(X(w)) \le y\}. \tag{1.5.6}$$

Any two random variables X, Y having the same distribution are **equivalent**. We denote this by $Y \sim X$.

A distribution function F may have a countable number of distinct points of discontinuity. If x_0 is a point of discontinuity, $F(x_0) - F(x_0-) > 0$. In between points of discontinuity, F is continuous. If F assumes a constant value between points of discontinuity (step function), it is called **discrete**. Formally, let $-\infty < x_1 < x_2 < \cdots < \infty$ be points of discontinuity of F. Let $I_A(x)$ denote the indicator function of a set A, i.e.,

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then a discrete F can be written as

$$\begin{aligned} F_{\mathrm{d}}(x) &= \sum_{i=1}^{\infty} I_{[x_i, x_{i+1})}(x) F(x_i) \\ &= \sum_{\{x_i \le x\}} (F(x_i) - F(x_i-)). \end{aligned} \tag{1.5.7}$$

Let μ_1 and μ_2 be measures on $(\mathbb{R}, \mathcal{B})$. We say that μ_1 is **absolutely continuous** with respect to μ_2, and write $\mu_1 \ll \mu_2$, if $B \in \mathcal{B}$ and $\mu_2(B) = 0$ then $\mu_1(B) = 0$. Let λ denote the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$. For every interval $(a, b]$, $-\infty < a < b < \infty$, $\lambda((a, b]) = b - a$. The celebrated **Radon–Nikodym** Theorem (see Shiryayev, 1984, p. 194) states that if $\mu_1 \ll \mu_2$ and μ_1, μ_2 are σ-finite measures on $(\mathbb{R}, \mathcal{B})$, there exists a $\mathcal{B}$-measurable nonnegative function $f(x)$ so that, for each $B \in \mathcal{B}$,

$$\mu_1(B) = \int_B f(x) d\mu_2(x) \tag{1.5.8}$$

where the **Lebesgue integral** in (1.5.8) will be discussed later. In particular, if P_c is absolutely continuous with respect to the Lebesgue measure λ, then there exists a function $f \geq 0$ so that

$$P_c\{B\} = \int_B f(x)\lambda(x), \quad B \in \mathcal{B}. \tag{1.5.9}$$

Moreover,

$$F_c(x) = \int_{-\infty}^{x} f(y)dy, \quad -\infty < x < \infty. \tag{1.5.10}$$

A distribution function F is called **absolutely continuous** if there exists a nonnegative function f such that

$$F(\xi) = \int_{-\infty}^{\xi} f(x)dx, \quad -\infty < \xi < \infty. \tag{1.5.11}$$

The function f, which can be represented for "almost all x" by the derivative of F, is called the **probability density function** (p.d.f.) corresponding to F.

If F is absolutely continuous, then $f(x) = \dfrac{d}{dx}F(x)$ "almost everywhere." The term "almost everywhere" or "almost all" x means for all x values, excluding maybe on a set N of Lebesgue measure zero. Moreover, the probability assigned to any interval $(\alpha, \beta]$, $\alpha \leq \beta$, is

$$P\{\alpha < X \leq \beta\} = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f(x)dx. \tag{1.5.12}$$

Due to the continuity of F we can also write

$$P\{\alpha < X \leq \beta\} = P\{\alpha \leq X \leq \beta\}.$$

Often the density functions f are Riemann integrable, and the above integrals are Riemann integrals. Otherwise, these are all Lebesgue integrals, which are defined in the next section.

There are continuous distribution functions that are not absolutely continuous. Such distributions are called **singular**. An example of a singular distribution is the **Cantor distribution** (see Shiryayev, 1984, p. 155).

Finally, every distribution function $F(x)$ is a **mixture** of the three types of distributions—discrete distribution $F_{\mathrm{d}}(\cdot)$, absolutely continuous distributions $F_{\mathrm{ac}}(\cdot)$, and singular distributions $F_{\mathrm{s}}(\cdot)$. That is, for some $0 \leq p_1, p_2, p_3 \leq 1$ such that $p_1 + p_2 + p_3 = 1$,

$$F(x) = p_1 F_{\mathrm{d}}(x) + p_2 F_{\mathrm{ac}}(x) + p_3 F_{\mathrm{s}}(x).$$

In this book we treat only mixtures of $F_{\mathrm{d}}(x)$ and $F_{\mathrm{ac}}(x)$.

1.6 THE LEBESGUE AND STIELTJES INTEGRALS

1.6.1 General Definition of Expected Value: The Lebesgue Integral

Let $(\mathcal{S}, \mathcal{F}, P)$ be a probability space. If X is a random variable, we wish to define the integral

$$E\{X\} = \int_{\mathcal{S}} X(w)P(dw). \tag{1.6.1}$$

We define first $E\{X\}$ for nonnegative random variables, i.e., $X(w) \geq 0$ for all $w \in \mathcal{S}$. Generally, $X = X^+ - X^-$, where $X^+(w) = \max(0, X(w))$ and $X^-(w) = -\min(0, X(w))$.

Given a nonnegative random variable X we construct for a given finite integer n the events

$$A_{k,n} = \left\{w : \frac{k-1}{2^n} \leq X(w) < \frac{k}{2^n}\right\}, \quad k = 1, 2, \ldots, n2^n$$

and

$$A_{n2^n+1,n} = \{w : X(w) \geq n\}.$$

These events form a partition of $\mathcal{S}$. Let X_n, $n \geq 1$, be the discrete random variable defined as

$$X_n(w) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{A_{k,n}}(w) + nI_{A_{n2^n+1,n}}(w). \tag{1.6.2}$$

Notice that for each w, $X_n(w) \leq X_{n+1}(w) \leq \ldots \leq X(w)$ for all n. Also, if $w \in A_{k,n}$, $k = 1, \ldots, n2^n$, then $|X(w) - X_n(w)| \leq \frac{1}{2^n}$. Moreover, $A_{n2^n+1,n} \supset A_{(n+1)2^{n+1},n+1}$, all $n \geq 1$. Thus

$$\begin{aligned}\lim_{n\to\infty} A_{n2^n+1,n} &= \bigcap_{n=1}^{\infty}\{w : X(w) \geq n\} \\ &= \emptyset.\end{aligned}$$

Thus for all $w \in \mathcal{S}$

$$\lim_{n\to\infty} X_n(w) = X(w). \tag{1.6.3}$$

Now, for each discrete random variable $X_n(w)$

$$E\{X_n\} = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} P\{A_{k,n}\} + nP\{w : X(w) > n\}. \tag{1.6.4}$$

Obviously $E\{X_n\} \le n$, and $E\{X_{n+1}\} \ge E\{X_n\}$. Thus, $\lim_{n\to\infty} E\{X_n\}$ exists (it might be $+\infty$). Accordingly, **the Lebesgue integral** is defined as

$$\begin{aligned} E\{X\} &= \int X(w)P\{dw\} \\ &= \lim_{n\to\infty} E\{X_n\}. \end{aligned} \tag{1.6.5}$$

The Lebesgue integral may exist when the Riemann integral does not. For example, consider the probability space $(\mathcal{I}, \mathcal{B}, P)$ where $\mathcal{I} = \{x : 0 \le x \le 1\}$, $\mathcal{B}$ the Borel σ-field on $\mathcal{I}$, and P the Lebesgue measure on $[\mathcal{B}]$. Define

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is irrational on } [0, 1] \\ 1, & \text{if } x \text{ is rational on } [0, 1]. \end{cases}$$

Let $B_0 = \{x : 0 \le x \le 1, \ f(x) = 0\}$, $B_1 = [0, 1] - B_0$. The Lebesgue integral of f is

$$\int_0^1 f(x)dx = 0 \cdot P\{B_0\} + 1 \cdot P\{B_1\} = 0,$$

since the Lebesgue measure of B_1 is zero. On the other hand, the Riemann integral of $f(x)$ does not exist. Notice that, contrary to the construction of the Riemann integral, the Lebesgue integral $\int f(x)P\{dx\}$ of a nonnegative function f is obtained by partitioning the **range** of the function f to 2^n subintervals $\mathcal{D}_n = \{B_j^{(n)}\}$ and constructing a discrete random variable $\hat{f}_n = \sum_{j=1}^{2^n} f_{n,j}^* I\{x \in B_j^{(n)}\}$, where $f_{n,j} = \inf\{f(x) : x \in B_j^{(n)}\}$. The expected value of $\hat{f}_n$ is $E\{\hat{f}_n\} = \sum_{j=1}^{2^n} f_{n,j}^* P(X \in B_j^{(n)})$. The sequence $\{E\{\hat{f}_n\}, n \ge 1\}$ is nondecreasing, and its limit exists (might be $+\infty$). Generally, we define

$$E\{X\} = E\{X^+\} - E\{X^-\} \tag{1.6.6}$$

if either $E\{X^+\} < \infty$ or $E\{X^-\} < \infty$.

If $E\{X^+\} = \infty$ and $E\{X^-\} = \infty$, we say that $E\{X\}$ does not exist. As a special case, if F is absolutely continuous with density f, then

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

provided $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. If F is discrete then

$$E\{X\} = \sum_{n=1}^{\infty} x_n P\{X = x_n\}$$

provided it is absolutely convergent.

From the definition (1.6.4), it is obvious that if $P\{X(w) \geq 0\} = 1$ then $E\{X\} \geq 0$. This immediately implies that if X and Y are two random variables such that $P\{w : X(w) \geq Y(w)\} = 1$, then $E\{X - Y\} \geq 0$. Also, if $E\{X\}$ exists then, for all $A \in \mathcal{F}$,

$$E\{|X| I_A(X)\} \leq E\{|X|\},$$

and $E\{X I_A(X)\}$ exists. If $E\{X\}$ is finite, $E\{X I_A(X)\}$ is also finite. From the definition of expectation we immediately obtain that for any finite constant c,

$$\begin{aligned} E\{cX\} &= cE\{X\}, \\ E\{X + Y\} &= E\{X\} + E\{Y\}. \end{aligned} \tag{1.6.7}$$

Equation (1.6.7) implies that the expected value is a **linear functional**, i.e., if $X_1, \ldots, X_n$ are random variables on $(\mathcal{S}, \mathcal{F}, P)$ and $\beta_0, \beta_1, \ldots, \beta_n$ are finite constants, then, if all expectations exist,

$$E\left\{\beta_0 + \sum_{i=1}^{n} \beta_i X_i\right\} = \beta_0 + \sum_{i=1}^{n} \beta_i E\{X_i\}. \tag{1.6.8}$$

We present now a few basic theorems on the convergence of the expectations of sequences of random variables.

Theorem 1.6.1 (Monotone Convergence). *Let $\{X_n\}$ be a monotone sequence of random variables and Y a random variable.*

(i) *Suppose that $X_n(w) \underset{n\to\infty}{\nearrow} X(w)$, $X_n(w) \geq Y(w)$ for all n and all $w \in \mathcal{S}$, and $E\{Y\} > -\infty$. Then*

$$\lim_{n\to\infty} E\{X_n\} = E\{X\}.$$

(ii) *If* $X_n(w) \underset{n\to\infty}{\searrow} X(w)$, $X_n(w) \le Y(w)$, *for all n and all* $w \in \mathcal{S}$, *and* $E\{Y\} < \infty$, *then*

$$\lim_{n\to\infty} E\{X_n\} = E\{X\}.$$

Proof. See Shiryayev (1984, p. 184). QED

Corollary 1.6.1. *If* $X_1, X_2, \ldots$ *are nonnegative random variables, then*

$$E\left\{\sum_{n=1}^{\infty} X_n\right\} = \sum_{n=1}^{\infty} E\{X_n\}. \tag{1.6.9}$$

Theorem 1.6.2 (Fatou). *Let* X_n, $n \ge 1$ *and* Y *be random variables.*

(i) *If* $X_n(w) \ge Y(w)$, $n \ge 1$, *for each* w *and* $E\{Y\} > -\infty$, *then*

$$E\left\{\varliminf_{n\to\infty} X_n\right\} \le \varliminf_{n\to\infty} E\{X_n\};$$

(ii) *if* $X_n(w) \le Y(w)$, $n \ge 1$, *for each* w *and* $E\{Y\} < \infty$, *then*

$$\varlimsup_{n\to\infty} E\{X_n\} \le E\left\{\varlimsup_{n\to\infty} X_n\right\};$$

(iii) *if* $|X_n(w)| \le Y(w)$ *for each* w, *and* $E\{Y\} < \infty$, *then*

$$E\left\{\varliminf_{n\to\infty} X_n\right\} \le \varliminf_{n\to\infty} E\{X_n\} \le \varlimsup_{n\to\infty} E\{X_n\} \le E\left\{\varlimsup_{n\to\infty} X_n\right\}. \tag{1.6.10}$$

Proof. (i)

$$\varliminf_{n\to\infty} X_n(w) = \lim_{n\to\infty} \inf_{m\ge n} X_m(w).$$

The sequence $Z_n(w) = \inf\limits_{m\ge n} X_m(w)$, $n \ge 1$ is monotonically increasing for each w, and $Z_n(w) \ge Y(w)$, $n \ge 1$. Hence, by Theorem 1.6.1,

$$\lim_{n\to\infty} E\{Z_n\} = E\left\{\lim_{n\to\infty} Z_n\right\}.$$

Or

$$E\left\{\varliminf_{n\to\infty} X_n\right\} = \lim_{n\to\infty} E\{Z_n\} = \varliminf_{n\to\infty} E\{Z_n\} \leq \varliminf_{n\to\infty} E\{X_n\}.$$

The proof of (ii) is obtained by defining $Z_n(w) = \sup_{m\geq n} X_m(w)$, and applying the previous theorem. Part (iii) is a result of (i) and (ii). QED

Theorem 1.6.3 (Lebesgue Dominated Convergence). *Let Y, X, X_n, $n \geq 1$, be random variables such that $|X_n(w)| \leq Y(w)$, $n \geq 1$ for almost all w, and $E\{Y\} < \infty$. Assume also that $P\left\{w : \lim_{n\to\infty} X_n(w) = X(w)\right\} = 1$. Then $E\{|X|\} < \infty$ and*

$$\lim_{n\to\infty} E\{X_n\} = E\left\{\lim_{n\to\infty} X_n\right\}, \tag{1.6.11}$$

and

$$\lim_{n\to\infty} E\{|X_n - X|\} = 0. \tag{1.6.12}$$

Proof. By Fatou's Theorem (Theorem 1.6.2)

$$E\left\{\varliminf_{n\to\infty} X_n\right\} \leq \varliminf_{n\to\infty} E\{X_n\} \leq \varlimsup_{n\to\infty} E\{X_n\} \leq E\left\{\varlimsup_{n\to\infty} X_n\right\}.$$

But since $\lim_{n\to\infty} X_n(w) = X(w)$, with probability 1,

$$E\{X\} = E\left\{\lim_{n\to\infty} X_n\right\} = \lim_{n\to\infty} E\{X_n\}.$$

Moreover, $|X(w)| < Y(w)$ for almost all w (with probability 1). Hence, $E\{|X|\} < \infty$. Finally, since $|X_n(w) - X(w)| \leq 2Y(w)$, with probability 1

$$\lim_{n\to\infty} E\{|X_n - X|\} = E\left\{\lim_{n\to\infty} |X_n - X|\right\} = 0.$$

QED

We conclude this section with a theorem on change of variables under Lebesgue integrals.

Theorem 1.6.4. *Let X be a random variable with respect to $(\mathcal{S}, \mathcal{F}, P)$. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then for each $B \in \mathcal{B}$,*

$$\int_B g(x) P_X\{dx\} = \int_{X^{-1}(B)} g(X(w)) P\{dw\}. \tag{1.6.13}$$

The proof of the theorem is based on the following steps.

1. If $A \in \mathcal{B}$ and $g(x) = I_A(x)$ then

$$\begin{aligned}\int_B g(x)P_X\{dx\} &= \int_B I_A(x)P_X\{dx\} = P_X\{A \cap B\}\\ &= P\{w : X^{-1}(A) \cap X^{-1}(B)\}\\ &= \int_{X^{-1}(B)} g(X(w))P\{dw\}.\end{aligned}$$

2. Show that Equation (1.6.13) holds for simple random variables.

3. Follow the steps of the definition of the Lebesgue integral.

1.6.2 The Stieltjes–Riemann Integral

Let g be a function of a real variable and F a distribution function. Let $(\alpha, \beta]$ be a half-closed interval. Let

$$\alpha = x_0 < x_1 < \cdots < x_{n-1} < x_n = \beta$$

be a partition of $(\alpha, \beta]$ to n subintervals $(x_{i-1}, x_i]$, $i = 1, \ldots, n$. In each subinterval choose x_i', $x_{i-1} < x_i' \le x_i$ and consider the sum

$$S_n = \sum_{i=1}^{n} g(x_i')[F(x_i) - F(x_{i-1})]. \tag{1.6.14}$$

If, as $n \to \infty$, $\max_{1 \le i \le n} |x_i - x_{i-1}| \to 0$ and if $\lim_{n\to\infty} S_n$ exists (finite) independently of the partitions, then the limit is called the **Stieltjes–Riemann integral** of g with respect to F. We denote this integral as

$$\int_\alpha^\beta g(x)dF(x).$$

This integral has the usual linear properties, i.e.,

(i) $\int_\alpha^\beta cg(x)dF(x) = c\int_\alpha^\beta g(x)dF(x)$;

(ii) $\int_\alpha^\beta (g_1(x) + g_2(x))dF(x) = \int_\alpha^\beta g_1(x)dF(x) + \int_\alpha^\beta g_2(x)dF(x)$; (1.6.15)

and

(iii) $$\int_{\alpha}^{\beta} g(x)d(\gamma F_1(x) + \delta F_2(x)) = \gamma \int_{\alpha}^{\beta} g(x)dF_1(x) + \delta \int_{\alpha}^{\beta} g(x)dF_2(x).$$

One can integrate by parts, if all expressions exist, according to the formula

$$\int_{\alpha}^{\beta} g(x)dF(x) = [g(\beta)F(\beta) - g(\alpha)F(\alpha)] - \int_{\alpha}^{\beta} g'(x)F(x)dx, \quad (1.6.16)$$

where $g'(x)$ is the derivative of $g(x)$. If F is strictly discrete, with jump points $-\infty < \xi_1 < \xi_2 < \cdots < \infty$,

$$\int_{\alpha}^{\beta} g(x)dF(x) = \sum_{j=1}^{\infty} I\{\alpha < \xi_j \leq \beta\} g(\xi_j) p_j, \quad (1.6.17)$$

where $p_j = F(\xi_j) - F(\xi_j -)$, $j = 1, 2, \ldots$. If F is absolutely continuous, then at almost all points,

$$F(x + dx) - F(x) = f(x)dx + o(dx),$$

as $dx \to 0$. Thus, in the absolutely continuous case

$$\int_{\alpha}^{\beta} g(x)dF(x) = \int_{\alpha}^{\beta} g(x)f(x)dx. \quad (1.6.18)$$

Finally, the **improper Stieltjes–Riemann integral**, if it exists, is

$$\int_{-\infty}^{\infty} g(x)dF(x) = \lim_{\substack{\beta \to \infty \\ \alpha \to -\infty}} \int_{\alpha}^{\beta} g(x)dF(x). \quad (1.6.19)$$

If B is a set obtained by union and complementation of a sequence of intervals, we can write, by setting $g(x) = I\{x \in B\}$,

$$\begin{aligned} P\{B\} &= \int_{-\infty}^{\infty} I\{x \in B\} dF(x) \\ &= \int_{B} dF(x), \end{aligned} \quad (1.6.20)$$

where F is either discrete or absolutely continuous.

1.6.3 Mixtures of Discrete and Absolutely Continuous Distributions

Let F_{d} be a discrete distribution and let F_{ac} be an absolutely continuous distribution function. Then for all α $0 \le \alpha \le 1$,

$$F(x) = \alpha F_{\mathrm{d}}(x) + (1-\alpha)F_{\mathrm{ac}}(x) \tag{1.6.21}$$

is also a distribution function, which is a mixture of the two types. Thus, for such mixtures, if $-\infty < \xi_1 < \xi_2 < \cdots < \infty$ are the jump points of F_{d}, then for every $-\infty < \gamma \le \delta < \infty$ and $B = (\gamma, \delta]$,

$$\begin{aligned} P\{B\} &= \int_{\gamma}^{\delta} dF(x) \\ &= \alpha \sum_{j=1}^{\infty} I\{\gamma < \xi_j \le \delta\} dF_{\mathrm{d}}(\xi_j) + (1-\alpha)\int_{\gamma}^{\delta} dF_{\mathrm{ac}}(x). \end{aligned} \tag{1.6.22}$$

Moreover, if $B^+ = [\gamma, \delta]$ then

$$P\{B^+\} = P\{B\} + dF_{\mathrm{d}}(\gamma).$$

The expected value of X, when $F(x) = pF_{\mathrm{d}}(x) + (1-p)F_{\mathrm{ac}}(x)$ is,

$$E\{X\} = p\sum_{\{j\}} \xi_j f_{\mathrm{d}}(\xi_j) + (1-p)\int_{-\infty}^{\infty} x f_{\mathrm{ac}}(x)dx, \tag{1.6.23}$$

where $\{\xi_j\}$ is the set of jump points of F_{d}; f_{d} and f_{ac} are the corresponding p.d.f.s. We assume here that the sum and the integral are absolutely convergent.

1.6.4 Quantiles of Distributions

The p-**quantiles** or **fractiles** of distribution functions are inverse points of the distributions. More specifically, the p-quantile of a distribution function F, designated by x_p or $F^{-1}(p)$, is the smallest value of x at which $F(x)$ is greater or equal to p, i.e.,

$$x_p = F^{-1}(p) = \inf\{x : F(x) \ge p\}. \tag{1.6.24}$$

The inverse function defined in this fashion is unique. The **median** of a distribution, $x_{.5}$, is an important parameter characterizing the **location** of the distribution. The **lower** and **upper quartiles** are the .25- and .75-quantiles. The difference between these quantiles, $R_Q = x_{.75} - x_{.25}$, is called the **interquartile range**. It serves as one of the measures of **dispersion** of distribution functions.

1.6.5 Transformations

From the distribution function $F(x) = \alpha F_{\mathrm{d}}(x) + (1-\alpha)F_{\mathrm{ac}}(x)$, $0 \le \alpha \le 1$, we can derive the distribution function of a transformed random variable $Y = g(X)$, which is

$$\begin{aligned} F_Y(y) &= P\{g(X) \le y\} \\ &= P\{X \in B_y\} \\ &= \alpha \sum_{j=1}^{\infty} I\{\xi_j \in B_y\} dF_{\mathrm{d}}(\xi_j) + (1-\alpha)\int_{B_y} dF_{\mathrm{ac}}(x) \end{aligned} \tag{1.6.25}$$

where

$$B_y = \{x : g(x) \le y\}.$$

In particular, if F is absolutely continuous and if g is a strictly increasing differentiable function, then the p.d.f. of Y, $h(y)$, is

$$f_Y(y) = f_X(g^{-1}(y))\left(\frac{d}{dy}g^{-1}(y)\right), \tag{1.6.26}$$

where $g^{-1}(y)$ is the inverse function. If $g'(x) < 0$ for all x, then

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy}g^{-1}(y)\right|. \tag{1.6.27}$$

Suppose that X is a continuous random variable with p.d.f. $f(x)$. Let $g(x)$ be a differentiable function that is not necessarily one-to-one, like $g(x) = x^2$. Excluding cases where $g(x)$ is a constant over an interval, like the indicator function, let $m(y)$ denote the number of roots of the equation $g(x) = y$. Let $\xi_j(y)$, $j = 1, \ldots, m(y)$ denote the roots of this equation. Then the p.d.f. of $Y = g(x)$ is

$$f_Y(y) = \sum_{j=1}^{m(y)} f_X(\xi_j(y)) \cdot \frac{1}{|g'(\xi_j(y))|} \tag{1.6.28}$$

if $m(y) > 0$ and zero otherwise.

1.7 JOINT DISTRIBUTIONS, CONDITIONAL DISTRIBUTIONS AND INDEPENDENCE

1.7.1 Joint Distributions

Let $(X_1, \dots, X_k)$ be a vector of k random variables defined on the same probability space. These random variables represent variables observed in the same experiment. The joint distribution function of these random variables is a real value function F of k real arguments $(\xi_1, \dots, \xi_k)$ such that

$$F(\xi_1, \dots, \xi_k) = P\{X_1 \le \xi_1, \dots, X_k \le \xi_k\}. \tag{1.7.1}$$

The joint distribution of two random variables is called a **bivariate distribution function**.

Every bivariate distribution function F has the following properties.

$$\begin{aligned}
&\textbf{(i)} \quad \lim_{\xi_1 \to -\infty} F(\xi_1, \xi_2) = \lim_{\xi_2 \to -\infty} F(\xi_1, \xi_2) = 0, \quad \text{for all } \xi_1, \xi_2;\\
&\textbf{(ii)} \quad \lim_{\xi_1 \to \infty} \lim_{\xi_2 \to \infty} F(\xi_1, \xi_2) = 1;\\
&\textbf{(iii)} \quad \lim_{\epsilon \downarrow 0} F(\xi_1 + \epsilon, \xi_2 + \epsilon) = F(\xi_1, \xi_2) \ \text{ for all } (\xi_1, \xi_2);\\
&\textbf{(iv)} \quad \text{for any } a < b,\ c < d,\ F(b, d) - F(a, d) - F(b, c) + F(a, c) \ge 0.
\end{aligned} \tag{1.7.2}$$

Property (iii) is the right continuity of $F(\xi_1, \xi_2)$. Property (iv) means that the probability of every rectangle is nonnegative. Moreover, the total increase of $F(\xi_1, \xi_2)$ is from 0 to 1. The similar properties are required in cases of a larger number of variables.

Given a bivariate distribution function F. The univariate distributions of X_1 and X_2 are F_1 and F_2 where

$$F_1(x) = \lim_{y \to \infty} F(x, y), \ \ F_2(y) = \lim_{x \to \infty} F(x, y). \tag{1.7.3}$$

F_1 and F_2 are called the **marginal distributions** of X_1 and X_2, respectively. In cases of joint distributions of three variables, we can distinguish between three marginal bivariate distributions and three marginal univariate distributions. As in the univariate case, multivariate distributions are either discrete, absolutely continuous, singular, or mixtures of the three main types. In the discrete case there are at most a countable number of points $\{(\xi_1^{(j)}, \dots, \xi_k^{(j)}), j = 1, 2, \dots\}$ on which the distribution concentrates. In this case the joint probability function is

$$p(x_1, \dots, x_k) = \begin{cases} P\{X_1 = \xi_1^{(j)}, \dots, X_k = \xi_k^{(j)}\}, & \text{if } (x_1, \dots, x_k) = (\xi_1^{(j)}, \dots, \xi_k^{(j)}) \\ & j = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases} \tag{1.7.4}$$

Such a discrete p.d.f. can be written as

$$p(x_1, \dots, x_k) = \sum_{j=1}^{\infty} I\{(x_1, \dots, x_k) = (\xi_1^{(j)}, \dots, \xi_k^{(j)})\} p_j,$$

where $p_j = P\{X_1 = \xi_1^{(j)}, \dots, X_k = \xi_k^{(j)}\}$.

In the absolutely continuous case there exists a nonnegative function $f(x_1, \dots, x_k)$ such that

$$F(\xi_1, \dots, \xi_k) = \int_{-\infty}^{\xi_1} \cdots \int_{-\infty}^{\xi_k} f(x_1, \dots, x_k) dx_1 \dots dx_k. \tag{1.7.5}$$

The function $f(x_1, \dots, x_k)$ is called the **joint density function**.

The marginal probability or density functions of single variables or of a subvector of variables can be obtained by summing (in the discrete case) or integrating, in the absolutely continuous case, the joint distribution functions (densities) with respect to the variables that are not under consideration, over their range of variation.

Although the presentation here is in terms of k discrete or k absolutely continuous random variables, the joint distributions can involve some discrete and some continuous variables, or mixtures.

If X_1 has an absolutely continuous marginal distribution and X_2 is discrete, we can introduce the function $N(B)$ on $\mathcal{B}$, which counts the number of jump points of X_2 that belong to B. $N(B)$ is a σ-finite measure. Let $\lambda(B)$ be the Lebesgue measure on $\mathcal{B}$. Consider the σ-finite measure on $\mathcal{B}^{(2)}$, $\mu(B_1 \times B_2) = \lambda(B_1)N(B_2)$. If X_1 is absolutely continuous and X_2 discrete, their joint probability measure $P_{\mathbf{X}}$ is absolutely continuous with respect to μ. There exists then a nonnegative function $f_{\mathbf{X}}$ such that

$$F_{\mathbf{X}}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{\mathbf{X}}(y_1, y_2) dy_1 dN(y_2).$$

The function $f_{\mathbf{X}}$ is a joint p.d.f. of X_1, X_2 with respect to μ. The joint p.d.f. $f_{\mathbf{X}}$ is positive only at jump point of X_2.

If $X_1, \ldots, X_k$ have a joint distribution with p.d.f. $f(x_1, \ldots, x_k)$, the **expected** value of a function $g(X_1, \ldots, X_k)$ is defined as

$$E\{g(X_1, \ldots, X_k)\} = \int g(x_1, \ldots, x_k) dF(x_1, \ldots, x_k). \tag{1.7.6}$$

We have used here the conventional notation for Stieltjes integrals.

Notice that if (X, Y) have a joint distribution function $F(x, y)$ and if X is discrete with jump points of $F_1(x)$ at $\xi_1, \xi_2, \ldots$, and Y is absolutely continuous, then, as in the previous example,

$$\int g(x, y) dF(x, y) = \sum_{j=1}^{\infty} \int g(\xi_j, y) f(\xi_j, y) dy$$

where $f(x, y)$ is the joint p.d.f. A similar formula holds for the case of X, absolutely continuous and Y, discrete.

1.7.2 Conditional Expectations: General Definition

Let $X(w) \geq 0$, for all $w \in \mathcal{S}$, be a random variable with respect to $(\mathcal{S}, \mathcal{F}, P)$. Consider a σ-field $\mathcal{G}$, $\mathcal{G} \subset \mathcal{F}$. The conditional expectation of X given $\mathcal{G}$ is defined as a $\mathcal{G}$-measurable random variable $E\{X \mid \mathcal{G}\}$ satisfying

$$\int_A X(w) P\{dw\} = \int_A E\{X \mid \mathcal{G}\}(w) P\{dw\} \tag{1.7.7}$$

for all $A \in \mathcal{G}$. Generally, $E\{X \mid \mathcal{G}\}$ is defined if $\min\{E\{X^+ \mid \mathcal{G}\}, E\{X^- \mid \mathcal{G}\}\} < \infty$ and $E\{X \mid \mathcal{G}\} = E\{X^+ \mid \mathcal{G}\} - E\{X^- \mid \mathcal{G}\}$. To see that such conditional expectations exist, where $X(w) \geq 0$ for all w, consider the σ-finite measure on $\mathcal{G}$,

$$Q(A) = \int_A X(w) P\{dw\}, \quad A \in \mathcal{G}. \tag{1.7.8}$$

Obviously $Q \ll P$ and by Radon–Nikodym Theorem, there exists a nonnegative, $\mathcal{G}$-measurable random variable $E\{X \mid \mathcal{G}\}$ such that

$$Q(A) = \int_A E\{X \mid \mathcal{G}\}(w) P\{dw\}. \tag{1.7.9}$$

According to the Radon–Nikodym Theorem, $E\{X \mid \mathcal{G}\}$ is determined only up to a set of P-measure zero.

If $B \in \mathcal{F}$ and $X(w) = I_B(w)$, then $E\{X \mid \mathcal{G}\} = P\{B \mid \mathcal{G}\}$ and according to (1.6.13),

$$\begin{aligned} P\{A \cap B\} &= \int_A I_B(w) P\{dw\} \\ &= \int_A P\{B \mid \mathcal{G}\} P\{dw\}. \end{aligned} \tag{1.7.10}$$

Notice also that if X is $\mathcal{G}$-measurable then $X = E\{X \mid \mathcal{G}\}$ with probability 1.

On the other hand, if $\mathcal{G} = \{\emptyset, \mathcal{S}\}$ is the trivial algebra, then $E\{X \mid \mathcal{G}\} = E\{X\}$ with probability 1.

From the definition (1.7.7), since $\mathcal{S} \in \mathcal{G}$,

$$\begin{aligned} E\{X\} &= \int_{\mathcal{S}} X(w) P\{dw\} \\ &= \int_{\mathcal{S}} E\{X \mid \mathcal{G}\} P\{dw\}. \end{aligned}$$

This is the law of iterated expectation; namely, for all $\mathcal{G} \subset \mathcal{F}$,

$$E\{X\} = E\{E\{X \mid \mathcal{G}\}\}. \tag{1.7.11}$$

Furthermore, if X and Y are two random variables on $(\mathcal{S}, \mathcal{F}, P)$, the collection of all sets $\{Y^{-1}(B), B \in \mathcal{B}\}$, is a σ-field generated by Y. Let $\mathcal{F}_Y$ denote this σ-field. Since Y is a random variable, $\mathcal{F}_Y \subset \mathcal{F}$. We define

$$E\{X \mid Y\} = E\{X \mid \mathcal{F}_Y\}. \tag{1.7.12}$$

Let y_0 be such that $f_Y(y_0) > 0$.

Consider the $\mathcal{F}_Y$-measurable set $A_\delta = \{w : y_0 < Y(w) \le y_0 + \delta\}$. According to (1.7.7)

$$\begin{aligned} \int_{A\delta} X(w) P(dw) &= \int_{-\infty}^{\infty} \int_{y_0}^{y_0+\delta} x f_{XY}(x, y) dx dy \\ &= \int_{y_0}^{y_0+\delta} E\{X \mid Y = y\} f_Y(y) dy. \end{aligned} \tag{1.7.13}$$

The left-hand side of (1.7.13) is, if $E\{|X|\} < \infty$,

$$\begin{aligned} \int_{-\infty}^{\infty} x \int_{y_0}^{y_0+\delta} f_{XY}(x, y) dy dx &= \int_{y_0}^{y_0+\delta} f_Y(y_0) \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y)}{f_Y(y_0)} dx dy \\ &= f_Y(y_0)\delta \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y_0)}{f_Y(y_0)} dx + o(\delta), \quad \text{as } \delta \to 0 \end{aligned}$$

where $\lim_{\delta \to 0} \frac{o(\delta)}{\delta} = 0$. The right-hand side of (1.7.13) is

$$\int_{y_0}^{y_0+\delta} E\{X \mid Y = y\} f_Y(y) dy = E\{X \mid Y = y_0\} f_Y(y_0)\delta + o(\delta), \quad \text{as } \delta \to 0.$$

Dividing both sides of (1.7.13) by $f_Y(y_0)\delta$, we obtain that

$$\begin{aligned} E\{X \mid Y = y_0\} &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y_0) dx \\ &= \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y_0)}{f_X(y_0)} dx. \end{aligned}$$

We therefore define for $f_Y(y_0) > 0$

$$f_{X|Y}(x \mid y_0) = \frac{f_{XY}(x, y_0)}{f_Y(y_0)}. \tag{1.7.14}$$

More generally, for $k > 2$ let $f(x_1, \ldots, x_k)$ denote the joint p.d.f. of $(X_1, \ldots, X_k)$. Let $1 \leq r < k$ and $g(x_1, \ldots, x_r)$ denote the marginal joint p.d.f. of $(X_1, \ldots, X_r)$. Suppose that $(\xi_1, \ldots, \xi_r)$ is a point at which $g(\xi_1, \ldots, \xi_r) > 0$. The **conditional** p.d.f. of $X_{r+1}, \ldots, X_k$ given $\{X_1 = \xi_1, \ldots, X_r = \xi_r\}$ is defined as

$$h(x_{r+1}, \ldots, x_k \mid \xi_1, \ldots, \xi_r) = \frac{f(\xi_1, \ldots, \xi_r, x_{r+1}, \ldots, x_k)}{g(\xi_1, \ldots, \xi_r)}. \tag{1.7.15}$$

We remark that conditional distribution functions are not defined on points $(\xi_1, \ldots, \xi_r)$ such that $g(\xi_1, \ldots, \xi_r) = 0$. However, it is easy to verify that the probability associated with this set of points is zero. Thus, the definition presented here is sufficiently general for statistical purposes. Notice that $f(x_{r+1}, \ldots, x_k \mid \xi_1, \ldots, \xi_r)$ is, for a fixed point $(\xi_1, \ldots, \xi_r)$ at which it is well defined, a nonnegative function of $(x_{r+1}, \ldots, x_k)$ and that

$$\int dF(x_{r+1}, \ldots, x_k \mid \xi_1, \ldots, \xi_r) = 1.$$

Thus, $f(x_{r+1}, \ldots, x_k \mid \xi_1, \ldots, \xi_r)$ is indeed a joint p.d.f. of $(X_{r+1}, \ldots, X_k)$. The point $(\xi_1, \ldots, \xi_r)$ can be considered a parameter of the conditional distribution.

If $\psi(X_{r+1}, \ldots, X_k)$ is an (integrable) function of $(X_{r+1}, \ldots, X_k)$, the **conditional expectation** of $\psi(X_{r+1}, \ldots, X_k)$ given $\{X_1 = \xi_1, \ldots, X_r = \xi_r\}$ is

$$E\{\psi(X_{r+1}, \ldots, X_k) \mid \xi_1, \ldots, \xi_r\} = \int \psi(x_{r+1}, \ldots, x_k) dF(x_{r+1}, \ldots, x_k \mid \xi_1, \ldots, \xi_r). \tag{1.7.16}$$

This conditional expectation exists if the integral is absolutely convergent.

1.7.3 Independence

Random variables $X_1, \ldots, X_n$, on the same probability space, are called **mutually independent** if, for any Borel sets $B_1, \ldots, B_n$,

$$P\{w : X_1(w) \in B_1, \ldots, X_n(w) \in B_n\} = \prod_{j=1}^{n} P\{w : X_j \in B_j\}. \quad (1.7.17)$$

Accordingly, the joint distribution function of any k-tuple $(X_{i_1}, \ldots, X_{i_k})$ is a product of their marginal distributions. In particular,

$$F_{X_1 \cdots X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i). \quad (1.7.18)$$

Equation (1.7.18) implies that if $X_1, \ldots, X_n$ have a joint p.d.f. $f_{\mathbf{X}}(x_1, \ldots, x_n)$ and if they are independent, then

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j). \quad (1.7.19)$$

Moreover, if $g(X_1, \ldots, X_n) = \prod_{j=1}^{n} g_j(X_j)$, where $g(x_1, \ldots, x_n)$ is $\mathcal{B}^{(n)}$-measurable and $g_j(x)$ are $\mathcal{B}$-measurable, then under independence

$$E\{g(X_1, \ldots, X_n)\} = \prod_{j=1}^{n} E\{g_j(X_j)\}. \quad (1.7.20)$$

Probability models with independence structure play an important role in statistical theory. From (1.7.12) and (1.7.21), we imply that if $\mathbf{X}^{(r)} = (X_1, \ldots, X_r)$ and $\mathbf{Y}^{(r)} = (X_{r+1}, \ldots, X_n)$ are independent subvectors, then the conditional distribution of $\mathbf{X}^{(r)}$ given $\mathbf{Y}^{(r)}$ is independent of $\mathbf{Y}^{(r)}$, i.e.,

$$f(x_1, \ldots, x_r \mid x_{r+1}, \ldots, x_n) = f(x_1, \ldots, x_r) \quad (1.7.21)$$

with probability one.

1.8 MOMENTS AND RELATED FUNCTIONALS

A **moment** of **order** r, $r = 1, 2, \ldots,$ of a distribution $F(x)$ is

$$\mu_r = E\{X^r\}. \quad (1.8.1)$$

The moments of $Y = X - \mu_1$ are called **central moments** and those of $|X|$ are called **absolute** moments. It is simple to prove that the existence of an absolute moment of order r, $r > 0$, implies the existence of all moments of order s, $0 < s \leq r$, (see Section 1.13.3).

Let $\mu_r^* = E\{(X - \mu_1)^r\}$, $r = 1, 2, \ldots$ denote the rth central moment of a distribution. From the binomial expansion and the linear properties of the expectation operator we obtain the relationship between moments (about the origin) μ_r and center moments m_r

$$\mu_r^* = \sum_{j=0}^{r} (-1)^j \binom{r}{j} \mu_{r-j} \mu_i^j, \quad r = 1, 2, \ldots \tag{1.8.2}$$

where $\mu_0 \equiv 1$.

A distribution function F is called **symmetric about a point** ξ_0 **if its** p.d.f. **is symmetric about** ξ_0, i.e.,

$$f(\xi_0 + h) = f(\xi_0 - h), \quad \text{all } 0 \leq h < \infty.$$

From this definition we immediately obtain the following results.

(i) If F is symmetric about ξ_0 and $E\{|X|\} < \infty$, then $\xi_0 = E\{X\}$.

(ii) If F is symmetric, then all **central** moments of **odd order** are zero, i.e., $E\{(X - E\{X\})^{2m+1}\} = 0$, $m = 0, 1, \ldots$, provided $E|X|^{2m+1} < \infty$.

The central moment of the second order occupies a central role in the theory of statistics and is called the **variance** of X. The variance is denoted by $V\{X\}$. The square-root of the variance, called the **standard deviation**, is a measure of dispersion around the expected value. We denote the standard deviation by σ. The variance of X is equal to

$$V\{X\} = E\{X^2\} - (E\{X\})^2. \tag{1.8.3}$$

The variance is always nonnegative, and hence for every distribution having a finite second moment $E\{X^2\} \geq (E\{X\})^2$. One can easily verify from the definition that if X is a random variable and a and b are constants, then $V\{a + bX\} = b^2 V\{X\}$.

The variance is equal to zero if and only if the distribution function is concentrated at one point (a degenerate distribution).

A famous inequality, called the **Chebychev inequality**, relates the probability of X concentrating around its mean, and the standard deviation σ.

Theorem 1.8.1 (Chebychev). *If F_X has a finite standard deviation σ, then, for every $a > 0$,*

$$P\{w : |X(w) - \mu| \leq a\} \geq 1 - \frac{\sigma^2}{a^2}, \tag{1.8.4}$$

where $\mu = E\{X\}$.

Proof.

$$
\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 dF_X(x) \\
&= \int_{\{|x-\mu|\le a\}} (x-\mu)^2 dF_X(x) + \int_{\{|x-\mu|>a\}} (x-\mu)^2 dF_X(x) \qquad (1.8.5) \\
&\ge a^2 P\{w : |X(w)-\mu| > a\}.
\end{aligned}
$$

Hence,

$$P\{w : |X(w)-\mu| \le a\} = 1 - P\{w : |X(w)-\mu| > a\} \ge 1 - \frac{\sigma^2}{a^2}.$$

QED

Notice that in the proof of the theorem, we used the Riemann–Stieltjes integral. The theorem is true for **any** type of distribution for which $0 \le \sigma < \infty$. The Chebychev inequality is a crude inequality. Various types of better inequalities are available, under additional assumptions (see Zelen and Severv, 1968; Rohatgi, 1976, p. 102).

The **moment generating function** (m.g.f.) of a random variable X, denoted by M, is defined as

$$M(t) = E\{\exp(tX)\}, \qquad (1.8.6)$$

where t is such that $M(t) < \infty$. Obviously, at $t = 0$, $M(0) = 1$. However, $M(t)$ may not exist when $t \ne 0$. Assume that $M(t)$ exists for all t in some interval (a, b), $a < 0 < b$. There is a one-to-one correspondence between the distribution function F and the moment generating function M. M is analytic on (a, b), and can be differentiated under the expectation integral. Thus

$$\frac{d^r}{dt^r} M(t) = E\{X^r \exp\{tX\}\}, \quad r = 1, 2, \ldots. \qquad (1.8.7)$$

Under this assumption the rth derivative of $M(t)$ evaluated at $t = 0$ yields the moment of order r.

To overcome the problem of M being undefined in certain cases, it is useful to use the **characteristic function**

$$\phi(t) = E\{e^{itX}\}, \qquad (1.8.8)$$

where $i = \sqrt{-1}$. The characteristic function exists for all t since

$$|\phi(t)| = \left| \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \le \int_{-\infty}^{\infty} |e^{itx}| dF(x) = 1. \qquad (1.8.9)$$

Indeed, $|e^{itx}| = 1$ for all x and all t.

If X assumes nonnegative integer values, it is often useful to use the **probability generating function** (p.g.f.)

$$G(t) = \sum_{j=0}^{\infty} t^j p_j, \tag{1.8.10}$$

which is convergent if $|t| < 1$. Moreover, given a p.g.f. of a nonnegative integer value random variable X, its p.d.f. can be obtained by the formula

$$P\{w : X(w) = k\} = \frac{1}{k!} \frac{d^k}{dt^k} G(t)\Big|_{t=0}. \tag{1.8.11}$$

The logarithm of the moment generating function is called **cumulants generating function**. We denote this generating function by K. K exists for all t for which M is finite. Both M and K are analytic functions in the interior of their domains of convergence. Thus we can write for t close to zero

$$K(t) = \log M(t) = \sum_{j=0}^{\infty} \frac{\kappa_j}{j!} t^j \tag{1.8.12}$$

The coefficients $\{\kappa_j\}$ are called **cumulants**. Notice that $\kappa_0 = 0$, and κ_j, $j \geq 1$, can be obtained by differentiating $K(t)$ j times, and setting $t = 0$. Generally, the relationships between the cumulants and the moments of a distribution are, for $j = 1, \ldots, 4$

$$\begin{aligned}
\kappa_1 &= \mu_1 \\
\kappa_2 &= \mu_2 - \mu_1^2 = \mu_2^* \\
\kappa_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 = \mu_3^* \\
\kappa_4 &= \mu_4^* - 3(\mu_2^*)^2.
\end{aligned} \tag{1.8.13}$$

The following two indices

$$\beta_1 = \frac{\mu_3^*}{\sigma^3} \tag{1.8.14}$$

and

$$\beta_2 = \frac{\mu_4^*}{\sigma^4}, \tag{1.8.15}$$

where $\sigma^2 = \mu_2^*$ is the variance, are called coefficients of **skewness** (asymmetry) and **kurtosis** (steepness), respectively. If the distribution is symmetric, then $\beta_1 = 0$. If $\beta_1 > 0$ we say that the distribution is positively skewed; if $\beta_1 < 0$, it is negatively skewed. If $\beta_2 > 3$ we say that the distribution is **steep**, and if $\beta_2 < 3$ we say that the distribution is **flat**.

The following equation is called the law of total variance.

If $E\{X^2\} < \infty$ then

$$V\{X\} = E\{V\{X \mid Y\}\} + V\{E\{X \mid Y\}\}, \tag{1.8.16}$$

where $V\{X \mid Y\}$ denotes the conditional variance of X given Y.

It is often the case that it is easier to find the conditional mean and variance, $E\{X \mid Y\}$ and $V\{X \mid Y\}$, than to find $E\{X\}$ and $V\{X\}$ directly. In such cases, formula (1.8.16) becomes very handy.

The product central moment of two variables (X, Y) is called the **covariance** and denoted by $\text{cov}(X, Y)$. More specifically

$$\begin{aligned} \text{cov}(X, Y) &= E\{[X - E\{X\}][Y - E\{Y\}]\} \\ &= E\{XY\} - E\{X\}E\{Y\}. \end{aligned} \tag{1.8.17}$$

Notice that $\text{cov}(X, Y) = \text{cov}(Y, X)$, and $\text{cov}(X, X) = V\{X\}$. Notice that if X is a random variable having a finite first moment and a is any finite **constant**, then $\text{cov}(a, X) = 0$. Furthermore, whenever the second moments of X and Y exist the covariance exists. This follows from **the Schwarz inequality** (see Section 1.13.3), i.e., if F is the joint distribution of (X, Y) and F_X, F_Y are the marginal distributions of X and Y, respectively, then

$$\left(\int g(x)h(y)dF(x, y)\right)^2 \le \left(\int g^2(x)dF_X(x)\right)\left(\int h^2(y)dF_Y(y)\right) \tag{1.8.18}$$

whenever $E\{g^2(X)\}$ and $E\{h^2(Y)\}$ are finite. In particular, for any two random variables having second moments

$$\text{cov}^2(X, Y) \le V\{X\}V\{Y\}.$$

The ratio

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{V\{X\}V\{Y\}}} \tag{1.8.19}$$

is called the **coefficient of correlation** (Pearson's product moment correlation). From (1.8.18) we deduce that $-1 \le \rho \le 1$. The sign of ρ is that of $\text{cov}(X, Y)$.

The **m.g.f.** of a multivariate distribution is a function of k variables

$$M(t_1, \ldots, t_k) = E\left\{\exp\left\{\sum_{i=1}^{k} t_i X_i\right\}\right\}. \tag{1.8.20}$$

Let $X_1, \ldots, X_k$ be random variables having a joint distribution. Consider the linear transformation $Y = \sum_{j=1}^{k} \beta_j X_j$, where $\beta_1, \ldots, \beta_k$ are constants. Some formulae for the moments and covariances of such linear functions are developed here. Assume that all the moments under consideration exist. Starting with the expected value of Y we prove:

$$E\left\{\sum_{i=1}^{k} \beta_i X_i\right\} = \sum_{i=1}^{k} \beta_i E\{X_i\}. \tag{1.8.21}$$

This result is a direct implication of the definition of the integral as a linear operator.

Let $\mathbf{X}$ denote a random vector in a column form and $\mathbf{X}'$ its transpose. The expected value of a random vector $\mathbf{X}' = (X_1, \ldots, X_k)$ is defined as the corresponding vector of expected values, i.e.,

$$E\{\mathbf{X}'\} = (E\{X_1\}, \ldots, E\{X_k\}). \tag{1.8.22}$$

Furthermore, let Σ denote a $k \times k$ matrix with elements that are the variances and covariances of the components of $\mathbf{X}$. In symbols

$$\Sigma = (\sigma_{ij}; i, j = 1, \ldots, k), \tag{1.8.23}$$

where $\sigma_{ij} = \text{cov}(X_i, X_j)$, $\sigma_{ii} = V\{X_i\}$. If $Y = \boldsymbol{\beta}'\mathbf{X}$ where $\boldsymbol{\beta}$ is a vector of constants, then

$$\begin{aligned} V\{Y\} &= \boldsymbol{\beta}' \Sigma \boldsymbol{\beta} \\ &= \sum_i \sum_j \beta_i \beta_j \sigma_{ij} \\ &= \sum_{i=1}^{k} \beta_i^2 \sigma_{ii} + \sum_{i \neq j}\sum \beta_i \beta_j \sigma_{ij}. \end{aligned} \tag{1.8.24}$$

The result given by (1.8.24) can be generalized in the following manner. Let $Y_1 = \boldsymbol{\beta}'\mathbf{X}$ and $Y_2 = \boldsymbol{\alpha}'\mathbf{X}$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are arbitrary constant vectors. Then

$$\text{cov}(Y_1, Y_2) = \boldsymbol{\alpha}' \Sigma \boldsymbol{\beta}. \tag{1.8.25}$$

Finally, if $\mathbf{X}$ is a k-dimensional random vector with covariance matrix Σ and $\mathbf{Y}$ is an m-dimensional vector $\mathbf{Y} = A\mathbf{X}$, where A is an $m \times k$ matrix of constants, then the covariance matrix of $\mathbf{Y}$ is

$$V[\mathbf{Y}] = A \Sigma A'. \tag{1.8.26}$$

In addition, if the covariance matrix of $\mathbf{X}$ is $\boldsymbol{\Sigma}$, then the covariance matrix of $\mathbf{Y} = \boldsymbol{\xi} + A\mathbf{X}$ is $\mathbf{V}$, where $\boldsymbol{\xi}$ is a vector of constants, and A is a matrix of constants. Finally, if $\mathbf{Y} = A\mathbf{X}$ and $\mathbf{Z} = B\mathbf{X}$, where A and B are matrices of constants with compatible dimensions, then the covariance matrix of $\mathbf{Y}$ and $\mathbf{Z}$ is

$$C[\mathbf{Y}, \mathbf{Z}] = A\boldsymbol{\Sigma} B'. \tag{1.8.27}$$

We conclude this section with an important theorem concerning a characteristic function. Recall that ϕ is generally a complex valued function on $\mathbb{R}$, i.e.,

$$\phi(t) = \int_{-\infty}^{\infty} \cos(tx)dF(x) + i \int_{-\infty}^{\infty} \sin(tx)dF(x).$$

Theorem 1.8.2. *A characteristic function ϕ, of a distribution function F, has the following properties.*

(i) $|\phi(t)| \le \phi(0) = 1$*;*

(ii) *$\phi(t)$ is a uniformly continuous function of t, on $\mathbb{R}$;*

(iii) *$\phi(t) = \overline{\phi(-t)}$, where $\bar{z}$ denotes the complex conjugate of z;*

(iv) *$\phi(t)$ is real valued if and only if F is symmetric around $x_0 = 0$;*

(v) *if $E\{|X|^n\} < \infty$ for some $n \ge 1$, then the rth order derivative $\phi^{(r)}(t)$ exists for every $1 \le r \le n$, and*

$$\phi^{(r)}(t) = \int_{-\infty}^{\infty} (ix)^r e^{itx} dF(x), \tag{1.8.28}$$

$$\mu_r = \frac{1}{i^r}\phi^{(r)}(0), \tag{1.8.29}$$

and

$$\phi(t) = \sum_{j=1}^{n} \frac{(it)^j}{j!}\mu_j + \frac{(it)^n}{n!} R_n(t), \tag{1.8.30}$$

where $|R_n(t)| \le 3E\{|X|^n\}$, $R_n(t) \to 0$ as $t \to 0$;

(vi) *if $\phi^{(2n)}(0)$ exists and is finite, then $\mu_{2n} < \infty$;*

(vii) *if $E\{|X|^n\} < \infty$ for all $n \ge 1$ and*

$$\overline{\lim_{n\to\infty}} \left(\frac{E\{|X|^n\}}{n!}\right)^{1/n} = \frac{1}{R} < \infty, \tag{1.8.31}$$

then

$$\phi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mu_n, \quad |t| < R. \tag{1.8.32}$$

Proof. The proof of (**i**) and (**ii**) is based on the fact that $|e^{itx}| = 1$ for all t and all x. Now, $\int e^{-itx} dF(x) = \phi(-t) = \overline{\phi(t)}$. Hence (**iii**) is proven.

(**iv**) Suppose $F(x)$ is symmetric around $x_0 = 0$. Then $dF(x) = dF(-x)$ for all x. Therefore, since $\sin(-tx) = -\sin(tx)$ for all x, $\int_{-\infty}^{\infty} \sin(tx) dF(x) = 0$, and $\phi(t)$ is real. If $\phi(t)$ is real, $\phi(t) = \overline{\phi(t)}$. Hence $\phi_X(t) = \phi_{-X}(t)$. Thus, by the one-to-one correspondence between ϕ and F, for any Borel set B, $P\{X \in B\} = P\{-X \in B\} = P\{X \in -B\}$. This implies that F is symmetric about the origin.

(**v**) If $E\{|X|^n\} < \infty$, then $E\{|X|^r\} < \infty$ for all $1 \le r \le n$. Consider

$$\frac{\phi(t+h) - \phi(t)}{h} = E\left\{e^{itX}\left(\frac{e^{ihX} - 1}{h}\right)\right\}.$$

Since $\left|\frac{e^{ihx} - 1}{h}\right| \le |x|$, and $E\{|X|\} < \infty$, we obtain from the Dominated Convergence Theorem that

$$\begin{aligned}
\phi^{(1)}(t) &= \lim_{h \to 0}\left(\frac{\phi(t+h) - \phi(t)}{h}\right) \\
&= E\left\{e^{itX} \lim_{h \to 0} \frac{e^{ihX} - 1}{h}\right\} \\
&= iE\{Xe^{itX}\}.
\end{aligned}$$

Hence $\mu_1 = \frac{1}{i}\phi^{(1)}(0)$.

Equations (1.8.28)–(1.8.29) follow by induction. Taylor expansion of e^{iy} yields

$$e^{iy} = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}(\cos(\theta_1 y) + i \sin(\theta_2 y)),$$

where $|\theta_1| \le 1$ and $|\theta_2| \le 1$. Hence

$$\begin{aligned}
\phi(t) &= E\{e^{itX}\} \\
&= \sum_{k=0}^{n-1} \frac{(it)^k}{k!}\mu_k + \frac{(it)^n}{n!}(\mu_n + R_n(t)),
\end{aligned}$$

where

$$R_n(t) = E\{X^n(\cos(\theta_1 tX) + i\,\sin(\theta_2 tX) - 1)\}.$$

Since $|\cos(ty)| \le 1$, $|\sin(ty)| \le 1$, evidently $R_n(t) \le 3E\{|X|^n\}$. Also, by dominated convergence, $\lim_{t\to 0} R_n(t) = 0$.

(**vi**) By induction on n. Suppose $\phi^{(2)}(0)$ exists. By L'Hospital's rule,

$$\begin{aligned}
\phi^{(2)}(0) &= \lim_{h\to 0} \frac{1}{2}\left[\frac{\phi'(2h) - \phi'(0)}{2h} + \frac{\phi'(0) - \phi'(-2h)}{2h}\right] \\
&= \lim_{h\to 0} \frac{1}{4h^2}[\phi(2h) - 2\phi(0) + \phi(-2h)] \\
&= \lim_{h\to 0} \int \left(\frac{e^{ihx} - e^{-ihx}}{2h}\right)^2 dF(x) \\
&= -\lim_{h\to 0} \int \left(\frac{\sin(hx)}{hx}\right)^2 x^2 dF(x).
\end{aligned}$$

By Fatou's Lemma,

$$\begin{aligned}
\phi^{(2)}(0) &\le -\int \lim_{h\to 0}\left(\frac{\sin(hx)}{hx}\right)^2 x^2 dF(x) \\
&= -\int x^2 dF(x) = -\mu_2.
\end{aligned}$$

Thus, $\mu_2 \le -\phi^{(2)}(0) < \infty$. Assume that $0 < \mu_{2k} < \infty$. Then, by (v),

$$\begin{aligned}
\phi^{(2k)}(t) &= \int (ix)^{2k} e^{itx} dF(x) \\
&= (-1)^k \int e^{itx} dG(x),
\end{aligned}$$

where $dG(x) = x^{2k} dF(x)$, or

$$G(x) = \int_{-\infty}^{x} u^{2k} dF(u).$$

Notice that $G(\infty) = \mu_{2k} < \infty$. Thus, $\dfrac{(-1)^k \phi^{(2k)}(t)}{G(\infty)}$ is the characteristic function of the distribution $G(x)/G(\infty)$. Since $\dfrac{1}{G(\infty)} > 0$, $\int x^{2h+2} dF(x) = \int x^2 dG(x) < \infty$. This proves that $\mu_{2k} < \infty$ for all $k = 1, \ldots, n$.

(**vii**) Assuming (1.8.31), if $0 < t_0 < R$, $\overline{\lim_{n\to\infty}} \frac{(E\{|X|^n\})^{1/n}}{n!} < \frac{1}{t_0}$. Therefore,

$$\overline{\lim_{n\to\infty}} \frac{(E\{|X|^n\}t_0^n)^{1/n}}{n!} < 1.$$

By Stirling's approximation, $\lim_{n\to\infty}(n!)^{1/n} = 1$. Thus, for $0 < t_0 < R$,

$$\overline{\lim_{n\to\infty}} \left(\frac{E\{|X|^n\}t_0^n}{n!}\right)^{1/n} < 1.$$

Accordingly, by Cauchy's test, $\sum_{n=1}^{\infty} \frac{E\{|X|^n\}t_0^n}{n!} < \infty$. By (iv), for any $n \geq 1$, for any t, $|t| \leq t_0$

$$\phi(t) = \sum_{k=0}^{n} \frac{(it)^n}{n!}\mu_n + R_n^*(t),$$

where $|R_n^*(t)| \leq 3\frac{|t|^n}{n!}E\{|X|^n\}$. Thus, for every t, $|t| < R$, $\overline{\lim_{n\to\infty}}|R_n^*(t)| = 0$, which implies that

$$\phi(t) = \sum_{k=1}^{\infty} \frac{(it)^k}{k!}\mu_k, \quad \text{for all } |t| < R.$$

QED

1.9 MODES OF CONVERGENCE

In this section we formulate many definitions and results in terms of random vectors $\mathbf{X} = (X_1, X_2, \cdots, X_k)'$, $1 \leq k < \infty$. The notation $||\mathbf{X}||$ is used for the Euclidean norm, i.e., $||\mathbf{x}||^2 = \sum_{i=1}^{k} x_i^2$.

We discuss here four modes of convergence of sequences of random vectors to a random vector.

(i) Convergence in distribution, $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$;
(ii) Convergence in probability, $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$;
(iii) Convergence in rth mean, $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$; and
(iv) Convergence almost surely, $\mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{X}$.

A sequence $\mathbf{X}_n$ *is said to converge in distribution to* $\mathbf{X}$, $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ *if the corresponding distribution functions* F_n *and* F *satisfy*

$$\lim_{n\to\infty} \int g(\mathbf{x}) dF_n(\mathbf{x}) = \int g(\mathbf{x}) dF(\mathbf{x}) \tag{1.9.1}$$

for every continuous bounded function g *on* $\mathbb{R}^k$.

One can show that this definition is equivalent to the following statement.

A sequence $\{\mathbf{X}_n\}$ *converges in distribution to* $\mathbf{X}$, $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ *if* $\lim_{n\to\infty} F_n(\mathbf{x}) = F(\mathbf{x})$ *at all continuity points* $\mathbf{x}$ *of* F.

If $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ we say that F_n converges to F weakly. The notation is $F_n \xrightarrow{w} F$ or $F_n \Rightarrow F$.

We define now convergence in probability.

A sequence $\{\mathbf{X}_n\}$ *converges in probability to* $\mathbf{X}$, $\mathbf{X}_n \xrightarrow[n\to\infty]{p} X$ *if, for each* $\epsilon > 0$,

$$\lim_{n\to\infty} P\{||\mathbf{X}_n - \mathbf{X}|| > \epsilon\} = 0. \tag{1.9.2}$$

We define now convergence in rth mean.

A sequence of random vectors $\{\mathbf{X}_n\}$ *converges in* r*th mean,* $r > 0$, *to* $\mathbf{X}$, $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$ *if* $E\{||\mathbf{X}_n - \mathbf{X}||^r\} \to 0$ *as* $n \to \infty$.

A fourth mode of convergence is

A sequence of random vectors $\{\mathbf{X}_n\}$ *converges almost-surely to* $\mathbf{X}$, $\mathbf{X}_n \xrightarrow{\text{a.s.}} X$, *as* $n \to \infty$ *if*

$$P\{\lim_{n\to\infty} \mathbf{X}_n = \mathbf{X}\} = 1. \tag{1.9.3}$$

The following is an equivalent definition.

$\mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{X}$ *as* $n \to \infty$ *if and only if, for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\{||\mathbf{X}_m - \mathbf{X}|| < \epsilon,\ \forall m \geq n\} = 1. \tag{1.9.4}$$

Equation (1.9.4) is equivalent to

$$P\{\overline{\lim_{n\to\infty}} ||\mathbf{X}_n - \mathbf{X}|| < \epsilon\} = 1.$$

But,

$$\begin{aligned} P\{\overline{\lim_{n\to\infty}} ||\mathbf{X}_n - \mathbf{X}|| < \epsilon\} &= 1 - P\{\overline{\lim_{n\to\infty}} ||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon\} \\ &= 1 - P\{||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon,\ \ \text{i.o.}\}. \end{aligned}$$

By the Borel–Cantelli Lemma (Theorem 1.4.1), a sufficient condition for $\mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{X}$ is

$$\sum_{n=1}^{\infty} P\{||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon\} < \infty \tag{1.9.5}$$

for all $\epsilon > 0$.

Theorem 1.9.1. *Let* $\{\mathbf{X}_n\}$ *be a sequence of random vectors. Then*

(a) $\mathbf{X}_n \xrightarrow{a.s.} \mathbf{X}$ *implies* $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.
(b) $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$, $r > 0$, *implies* $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.
(c) $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ *implies* $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$.

Proof. (a) Since $\mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{X}$, for any $\epsilon > 0$,

$$\begin{aligned} 0 &= P\{\overline{\lim_{n\to\infty}} ||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon\} \\ &= \lim_{n\to\infty} P\left\{\bigcup_{m\geq n} ||\mathbf{X}_m - \mathbf{X}|| \geq \epsilon\right\} \\ &\geq \lim_{n\to\infty} P\{||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon\}. \end{aligned} \tag{1.9.6}$$

The inequality (1.9.6) implies that $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.

(b) It can be immediately shown that, for any $\epsilon > 0$,

$$E\{||\mathbf{X}_n - \mathbf{X}||^r\} \geq \epsilon^r P\{||\mathbf{X}_n - \mathbf{X}|| \geq \epsilon\}.$$

Thus, $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$ implies $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.

(c) Let $\epsilon > 0$. If $\mathbf{X}_n \leq \mathbf{x}_0$ then either $\mathbf{X} \leq \mathbf{x}_0 + \epsilon\mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)'$, or $||\mathbf{X}_n - \mathbf{X}|| > \epsilon$. Thus, for all n,

$$F_n(\mathbf{x}_0) \leq F(\mathbf{x}_0 + \epsilon\mathbf{1}) + P\{||\mathbf{X}_n - \mathbf{X}|| > \epsilon\}.$$

Similarly,

$$F(\mathbf{x}_0 - \epsilon\mathbf{1}) \leq F_n(\mathbf{x}_0) + P\{||\mathbf{X}_n - \mathbf{X}|| > \epsilon\}.$$

Finally, since $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$,

$$F(\mathbf{x}_0 - \epsilon\mathbf{1}) \leq \underline{\lim_{n\to\infty}} F_n(\mathbf{x}_0) \leq \overline{\lim_{n\to\infty}} F_n(\mathbf{x}_0) \leq F(\mathbf{x}_0 + \epsilon\mathbf{1}).$$

Thus, if $\mathbf{x}_0$ is a continuity point of F, by letting $\epsilon \to 0$, we obtain

$$\lim_{n\to\infty} F_n(\mathbf{x}_0) = F(\mathbf{x}_0).$$

QED

Theorem 1.9.2. *Let $\{\mathbf{X}_n\}$ be a sequence of random vectors. Then*

(a) *if $\mathbf{c} \in \mathbb{R}^k$, then $\mathbf{X}_n \xrightarrow{d} \mathbf{c}$ implies $\mathbf{X}_n \xrightarrow{p} \mathbf{c}$;*
(b) *if $\mathbf{X}_n \xrightarrow{a.s.} \mathbf{X}$ and $||\mathbf{X}_n||^r \le Z$, for some $r > 0$ and some (positive) random variable Z, with $E\{Z\} < \infty$, then $\mathbf{X}_n \xrightarrow{r} \mathbf{X}$.*

For proof, see Ferguson (1996, p. 9). Part (b) is implied also from Theorem 1.13.3.

Theorem 1.9.3. *Let $\{X_n\}$ be a sequence of nonnegative random variables such that $X_n \xrightarrow{a.s.} X$ and $E\{X_n\} \to E\{X\}$, $E\{X\} < \infty$. Then*

$$E\{|X_n - X|\} \to 0, \quad as \ \ n \to \infty. \tag{1.9.7}$$

Proof. Since $E\{X_n\} \to E\{X\} < \infty$, for sufficiently large n, $E\{X_n\} < \infty$. For such n,

$$\begin{aligned} E\{|X_n - X|\} &= E\{(X - X_n)I\{X \ge X_n\}\} + E\{(X_n - X)I\{X_n > X\}\} \\ &= 2E\{(X - X_n)I\{X \ge X_n\}\} + E\{X - X_n\}. \end{aligned}$$

But,

$$0 \le (X - X_n)I\{X \ge X_n\} < X.$$

Therefore, by the Lebesgue Dominated Convergence Theorem,

$$\lim_{n\to\infty} E\{(X - X_n)I\{X \ge X_n\}\} = 0.$$

This implies (1.9.7). QED

1.10 WEAK CONVERGENCE

The following theorem plays a major role in weak convergence.

Theorem 1.10.1. *The following conditions are equivalent.*

(a) $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$;

(b) $E\{g(\mathbf{X}_n)\} \to E\{g(\mathbf{X})\}$, *for all continuous functions, g, that vanish outside a compact set;*

(c) $E\{g(\mathbf{X}_n)\} \to E\{g(\mathbf{X})\}$, *for all continuous bounded functions g;*

(d) $E\{g(\mathbf{X}_n)\} \to E\{g(\mathbf{X})\}$, *for all measurable functions g such that* $P\{\mathbf{X} \in C(g)\} = 1$, *where* $C(g)$ *is the set of all points at which g is continuous.*

For proof, see Ferguson (1996, pp. 14–16).

Theorem 1.10.2. *Let* $\{\mathbf{X}_n\}$ *be a sequence of random vectors in* $\mathbb{R}^k$, *and* $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$. *Then*

(i) $\mathbf{f}(\mathbf{X}_n) \xrightarrow{d} \mathbf{f}(\mathbf{X})$;

(ii) *if* $\{\mathbf{Y}_n\}$ *is a sequence such that* $\mathbf{X}_n - \mathbf{Y}_n \xrightarrow{p} \mathbf{0}$, *then* $\mathbf{Y}_n \xrightarrow{d} \mathbf{X}$;

(iii) *if* $\mathbf{X}_n \in \mathbb{R}^k$ *and* $\mathbf{Y}_n \in \mathbb{R}^l$ *and* $\mathbf{Y}_n \xrightarrow{d} \mathbf{c}$, *then*

$$\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{X} \\ \mathbf{c} \end{pmatrix}.$$

Proof. (**i**) Let $g : \mathbb{R}^l \to \mathbb{R}$ be bounded and continuous. Let $h(\mathbf{x}) = g(\mathbf{f}(\mathbf{x}))$. If $\mathbf{x}$ is a continuity point of $\mathbf{f}$, then $\mathbf{x}$ is a continuity point of h, i.e., $C(\mathbf{f}) \subset C(h)$. Hence $P\{X \in C(h)\} = 1$. By Theorem 1.10.1 (c), it is sufficient to show that $E\{g(\mathbf{f}(\mathbf{X}_n))\} \to E\{g(\mathbf{f}(\mathbf{X}))\}$. Theorem 1.10.1 (d) implies, since $P\{X \in C(h)\} = 1$ and $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, that $E\{h(\mathbf{X}_n)\} \to E\{h(\mathbf{X})\}$.

(**ii**) According to Theorem 1.10.1 (b), let g be a continuous function on $\mathbb{R}^k$ vanishing outside a compact set. Thus g is uniformly continuous and bounded. Let $\epsilon > 0$, find $\delta > 0$ such that, if $||\mathbf{x} - \mathbf{y}|| < \delta$ then $|g(\mathbf{x}) - g(\mathbf{y})| < \epsilon$. Also, g is bounded, say $|g(x)| \leq B < \infty$. Thus,

$$\begin{aligned} |E\{g(\mathbf{Y}_n)\} - E\{g(\mathbf{X})\}| &\leq |E\{g(\mathbf{Y}_n)\} - E\{g(\mathbf{X}_n)\}| + |E\{g(\mathbf{X}_n)\} - E\{g(\mathbf{X})\}| \\ &\leq E\{|g(\mathbf{Y}_n) - g(\mathbf{X}_n)|I\{||\mathbf{X}_n - \mathbf{Y}_n|| \leq \delta\}\} \\ &+ E\{|g(\mathbf{Y}_n) - g(\mathbf{X}_n)|I\{||\mathbf{X}_n - \mathbf{Y}_n|| > \delta\} \\ &+ |E\{g(\mathbf{X}_n)\} - E\{g(\mathbf{X})\}| \\ &\leq \epsilon + 2BP\{||\mathbf{X}_n - \mathbf{Y}_n|| > \delta\} \\ &+ |E\{g(\mathbf{X}_n)\} - E\{g(\mathbf{X})\}| \underset{n\to\infty}{\longrightarrow} \epsilon. \end{aligned}$$

Hence $\mathbf{Y}_n \xrightarrow{d} \mathbf{X}$.

(iii)

$$P\left\{\left|\left|\begin{pmatrix}\mathbf{X}_n\\ \mathbf{Y}_n\end{pmatrix} - \begin{pmatrix}\mathbf{X}_n\\ \mathbf{c}\end{pmatrix}\right|\right| > \epsilon\right\} = P\{||\mathbf{Y}_n - \mathbf{c}|| > \epsilon\} \to 0, \quad \text{as} \quad n \to \infty.$$

Hence, from part (ii), $\begin{pmatrix}\mathbf{X}_n\\ \mathbf{Y}_n\end{pmatrix} \xrightarrow{d} \begin{pmatrix}\mathbf{X}\\ \mathbf{c}\end{pmatrix}$. QED

As a special case of the above theorem we get

Theorem 1.10.3 (Slutsky's Theorem). *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$. Then*

$$\begin{aligned} &\textit{(i)} \quad X_n + Y_n \xrightarrow{d} X + c; \\ &\textit{(ii)} \quad X_n Y_n \xrightarrow{d} cX; \\ &\textit{(iii)} \quad \text{if } c \neq 0 \text{ then } \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}. \end{aligned} \tag{1.10.1}$$

A sequence of distribution functions may not converge to a distribution function. For example, let X_n be random variables with

$$F_n(x) = \begin{cases} 0, & x < -n \\ \frac{1}{2}, & -n \leq x < n \\ 1, & n \leq x. \end{cases}$$

Then, $\lim_{n\to\infty} F_n(x) = \frac{1}{2}$ for all x. $F(x) = \frac{1}{2}$ for all x is not a distribution function. In this example, half of the probability mass escapes to $-\infty$ and half the mass escapes to $+\infty$. In order to avoid such situations, we require from collections (families) of probability distributions to be **tight**.

Let $\mathcal{F} = \{F_u, u \in \mathcal{U}\}$ be a family of distribution functions on $\mathbb{R}^k$. $\mathcal{F}$ is **tight** if, for any $\epsilon > 0$, there exists a **compact** set $C \subset \mathbb{R}^k$ such that

$$\sup_{u\in\mathcal{U}} \int I\{\mathbf{x} \in \mathbb{R}^k - C\} dF_u(\mathbf{x}) < \epsilon.$$

In the above, the sequence $F_n(x)$ is not tight.

If $\mathcal{F}$ is **tight**, then every sequence of distributions of $\mathcal{F}$ contains a subsequence converging weakly to a distribution function. (see Shiryayev, 1984, p. 315).

Theorem 1.10.4. *Let $\{F_n\}$ be a tight family of distribution functions on $\mathbb{R}$. A necessary and sufficient condition for $F_n \Rightarrow F$ is that, for each $t \in \mathbb{R}$, $\lim_{n\to\infty} \phi_n(t)$ exists, where $\phi_n(t) = \int e^{itx} dF_n(x)$ is the characteristic function corresponding to F_n.*

For proof, see Shiryayev (1984, p. 321).

Theorem 1.10.5 (Continuity Theorem). *Let $\{F_n\}$ be a sequence of distribution functions and $\{\phi_n\}$ the corresponding sequence of characteristic functions. Let F be a distribution function, with characteristic function ϕ. Then $F_n \Rightarrow F$ if and only if $\phi_n(\mathbf{t}) \to \phi(\mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^k$. (Shiryayev, 1984, p. 322).*

1.11 LAWS OF LARGE NUMBERS

1.11.1 The Weak Law of Large Numbers (WLLN)

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of identically distributed uncorrelated random vectors. Let $\boldsymbol{\mu} = E\{\mathbf{X}_1\}$ and let $\boldsymbol{\Sigma} = E\{(\mathbf{X}_1 - \mu)(\mathbf{X}_1 - \mu)'\}$ be finite. Then the means $\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$ converge in probability to $\boldsymbol{\mu}$, i.e.,

$$\bar{\mathbf{X}}_n \xrightarrow{p} \mu \quad \text{as } n \to \infty. \, \bar{\mathbf{X}}_n \xrightarrow{p} \boldsymbol{\mu} \quad \text{as } n \to \infty. \tag{1.11.1}$$

The proof is simple. Since $\text{cov}(\mathbf{X}_n, \mathbf{X}_{n'}) = \mathbf{0}$ for all $n \neq n'$, the covariance matrix of $\bar{\mathbf{X}}_n$ is $\frac{1}{n}\boldsymbol{\Sigma}$. Moreover, since $E\{\bar{\mathbf{X}}_n\} = \mu$,

$$E\{||\bar{\mathbf{X}}_n - \boldsymbol{\mu}||^2\} = \frac{1}{n} tr.\{\boldsymbol{\Sigma}\} \to 0 \quad \text{as } n \to \infty.$$

Hence $\bar{\mathbf{X}}_n \xrightarrow{2} \mu$, which implies that $\bar{\mathbf{X}}_n \xrightarrow{p} \mu$. Here $tr.\{\boldsymbol{\Sigma}\}$ denotes the trace of $\boldsymbol{\Sigma}$.

If $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent, and identically distributed, with $E\{\mathbf{X}_1\} = \boldsymbol{\mu}$, then the characteristic function of $\bar{\mathbf{X}}_n$ is

$$\phi_{\bar{\mathbf{X}}_n}(\mathbf{t}) = \left(\phi\left(\frac{\mathbf{t}}{n}\right)\right)^n, \tag{1.11.2}$$

where $\phi(\mathbf{t})$ is the characteristic function of $\mathbf{X}_1$. Fix $\mathbf{t}$. Then for large values of n,

$$\phi\left(\frac{\mathbf{t}}{n}\right) = 1 + \frac{i}{n}\mathbf{t}'\boldsymbol{\mu} + o\left(\frac{1}{n}\right), \text{ as } n \to \infty.$$

Therefore,

$$\phi_{\bar{\mathbf{X}}_n}(\mathbf{t}) = \left(1 + \frac{i}{n}\mathbf{t}'\boldsymbol{\mu} + o\left(\frac{1}{n}\right)\right)^n \to e^{i\mathbf{t}'\boldsymbol{\mu}}. \tag{1.11.3}$$

$\phi(\mathbf{t}) = e^{i\mathbf{t}'\mu}$ is the characteristic function of $\mathbf{X}$, where $P\{\mathbf{X} = \boldsymbol{\mu}\} = 1$. Thus, since $e^{i\mathbf{t}'\boldsymbol{\mu}}$ is continuous at $\mathbf{t} = \mathbf{0}$, $\bar{\mathbf{X}}_n \xrightarrow{d} \boldsymbol{\mu}$. This implies that $\bar{\mathbf{X}}_n \xrightarrow{p} \mu$ (left as an exercise).

1.11.2 The Strong Law of Large Numbers (SLLN)

Strong laws of large numbers, for independent random variables having finite expected values are of the form

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty,$$

where $\mu_i = E\{X_i\}$.

Theorem 1.11.1 (Cantelli). *Let $\{X_n\}$ be a sequence of independent random variables having uniformly bounded fourth-central moments, i.e.,*

$$0 \le E(X_n - \mu_n)^4 \le C < \infty \tag{1.11.4}$$

for all $n \ge 1$. Then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow[n\to\infty]{a.s.} 0. \tag{1.11.5}$$

Proof. Without loss of generality, we can assume that $\mu_n = E\{X_n\} = 0$ for all $n \ge 1$.

$$\begin{aligned} E\left\{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^4\right\} &= \frac{1}{n^4}\Bigg\{\sum_{i=1}^{n} E\{X_i^4\} \\ &\quad + 4\sum_{i\neq j}\sum E\{X_i^3 X_j\} + 3\sum_{i\neq j}\sum E\{X_i^2 X_j^2\} \\ &\quad + 6\sum\sum_{i\neq j\neq k}\sum E\{X_i^2 X_j X_k\} + \sum\sum_{i\neq j\neq k\neq l}\sum\sum E\{X_i X_j X_k X_l\}\Bigg\} \\ &= \frac{1}{n^4}\sum_{i=1}^{n}\mu_{4,i} + \frac{3}{n^4}\sum_{i\neq j}\sum \sigma_i^2\sigma_j^2, \end{aligned}$$

where $\mu_{4,i} = E\{X_i^4\}$ and $\sigma_i^2 = E\{X_i^2\}$. By the Schwarz inequality, $\sigma_i^2 \sigma_j^2 \le (\mu_{4,i} \cdot \mu_{4j})^{1/2}$ for all $i \neq j$. Hence,

$$E\{\bar{X}_n^4\} \le \frac{C}{n^3} + \frac{3n(n-1)C}{n^4} = O\left(\frac{1}{n^2}\right).$$

By Chebychev's inequality,

$$\begin{aligned} P\{|\bar{X}_n| \ge \epsilon\} &= P\{\bar{X}_n^4 \ge \epsilon^4\} \\ &\le \frac{E\{\bar{X}_n^4\}}{\epsilon^4}. \end{aligned}$$

Hence, for any $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P\{|\bar{X}_n| \ge \epsilon\} \le C^* \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

where C^* is some positive finite constant. Finally, by the Borel–Cantelli Lemma (Theorem 1.4.1),

$$P\{|\bar{X}_n| \ge \epsilon, \text{i.o.}\} = 0.$$

Thus, $P\{|\bar{X}_n| < \epsilon, \text{i.o.}\} = 1$. QED

Cantelli's Theorem is quite stringent, in the sense, that it requires the existence of the fourth moments of the independent random variables. Kolmogorov had relaxed this condition and proved that, if the random variables have finite variances, $0 < \sigma_n^2 < \infty$ and

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty, \tag{1.11.6}$$

then $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

If the random variables are **independent** and **identically distributed** (i.i.d.), then Kolmogorov showed that $E\{|X_1|\} < \infty$ is sufficient for the strong law of large numbers. To prove Kolmogorov's strong law of large numbers one has to develop more theoretical results. We refer the reader to more advanced probability books (see Shiryayev, 1984).

1.12 CENTRAL LIMIT THEOREM

The Central Limit Theorem (CLT) states that, under general valid conditions, the distributions of properly normalized sample means converge weakly to the standard normal distribution.

A continuous random variable Z is said to have a standard normal distribution, and we denote it $Z \sim N(0, 1)$ if its distribution function is absolutely continuous, having a p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty. \tag{1.12.1}$$

The c.d.f. of $N(0, 1)$, called the **standard normal integral** is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy. \tag{1.12.2}$$

The general family of normal distributions is studied in Chapter 2. Here we just mention that if $Z \sim N(0, 1)$, the moments of Z are

$$\mu_r = \begin{cases} \dfrac{(2k)!}{2^k k!}, & \text{if } r = 2k \\ 0, & \text{if } r = 2k + 1. \end{cases} \tag{1.12.3}$$

The characteristic function of $N(0, 1)$ is

$$\begin{aligned} \phi(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2 + itx} dx \\ &= e^{-\frac{1}{2}t^2}, \quad -\infty < t < \infty. \end{aligned} \tag{1.12.4}$$

A random vector $\bar{Z} = (Z_1, \ldots, Z_k)'$ is said to have a multivariate normal distribution with mean $\boldsymbol{\mu} = E\{\mathbf{Z}\} = \mathbf{0}$ and covariance matrix V (see Chapter 2), $\mathbf{Z} \sim N(\mathbf{0}, V)$ if the p.d.f. of $\mathbf{Z}$ is

$$f(\mathbf{Z}; V) = \frac{1}{(2\pi)^{k/2}|V|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{Z}' V^{-1} \mathbf{Z}\right\}.$$

The corresponding characteristic function is

$$\phi_{\mathbf{Z}}(\mathbf{t}) = \exp\left\{-\frac{1}{2}\mathbf{t}' V \mathbf{t}\right\}, \tag{1.12.5}$$

$\mathbf{t} \in \mathbb{R}^k$.

Using the method of characteristic functions, with the continuity theorem we prove the following simple two versions of the CLT. A proof of the Central Limit Theorem, which is not based on the continuity theorem of characteristic functions, can be obtained by the method of Stein (1986) for approximating expected values or probabilities.

Theorem 1.12.1 (CLT). *Let $\{X_n\}$ be a sequence of i.i.d. random variables having a finite positive variance, i.e., $\mu = E\{X_1\}$, $V\{X_1\} = \sigma^2$, $0 < \sigma^2 < \infty$. Then*

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1), \quad \textit{as } n \to \infty. \tag{1.12.6}$$

Proof. Notice that $\sqrt{n}\,\dfrac{\bar{X}_n - \mu}{\sigma} = \dfrac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i$, where $Z_i = \dfrac{X_i - \mu}{\sigma}$, $i \geq 1$. Moreover, $E\{Z_i\} = 0$ and $V\{Z_i\} = 1$, $i \geq 1$. Let $\phi_Z(t)$ be the characteristic function of Z_1. Then, since $E\{Z\} = 0$, $V\{Z\} = 1$, (1.8.33) implies that

$$\phi_Z(t) = 1 - \frac{t^2}{2} + o(t), \quad \text{as } t \to 0.$$

Accordingly, since $\{Z_n\}$ are i.i.d.,

$$\begin{aligned}
\phi_{\sqrt{n}\ \bar{Z}_n}(t) &= \phi_Z^n\left(\frac{t}{\sqrt{n}}\right) \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \\
&\to e^{-t^2/2} \quad \text{as } n \to \infty.
\end{aligned}$$

Hence, $\sqrt{n}\ \bar{Z}_n \xrightarrow{d} N(0, 1)$. QED

Theorem 1.12.1 can be generalized to random vector. Let $\bar{\mathbf{X}}_n = \dfrac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$, $n \geq 1$. The generalized CLT is the following theorem.

Theorem 1.12.2. *Let $\{\mathbf{X}_n\}$ be a sequence of i.i.d. random vectors with $E\{\mathbf{X}_n\} = 0$, and covariance matrix $E\{\mathbf{X}_n\mathbf{X}'_n\} = \mathbf{V}$, $n \geq 1$, where $\mathbf{V}$ is positive definite with finite eigenvalues. Then*

$$\sqrt{n}\,\bar{\mathbf{X}}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{V}). \tag{1.12.7}$$

Proof. Let $\phi_{\mathbf{X}}(\mathbf{t})$ be the characteristic function of $\mathbf{X}_1$. Then, since $E\{\mathbf{X}_1\} = \mathbf{0}$,

$$\begin{aligned}\phi_{\sqrt{n}\ \bar{X}_n}(\mathbf{t}) &= \phi^n_{\mathbf{X}}\left(\frac{\mathbf{t}}{\sqrt{n}}\right) \\ &= \left(1 - \frac{1}{2n}\mathbf{t}'V\mathbf{t} + o\left(\frac{\mathbf{t}}{\sqrt{n}}\right)\right)^n\end{aligned}$$

as $n \to \infty$. Hence

$$\lim_{n\to\infty} \phi_{\sqrt{n}\ \bar{X}_n}(\mathbf{t}) = \exp\left\{-\frac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}\right\}, \quad \mathbf{t} \in \mathbb{R}^k.$$

QED

When the random variables are independent but not identically distributed, we need a stronger version of the CLT. The following celebrated CLT is sufficient for most purposes.

Theorem 1.12.3 (Lindeberg–Feller). *Consider a triangular array of random variables* $\{X_{n,k}\}$, $k = 1, \ldots, n$, $n \geq 1$ *such that, for each* $n \geq 1$, $\{X_{n,k}, k = 1, \ldots, n\}$ *are independent, with* $E\{X_{n,k}\} = 0$ *and* $V\{X_{n,k}\} = \sigma^2_{n,k}$. *Let* $S_n = \sum_{k=1}^{n} X_{n,k}$ *and* $B^2_n = \sum_{k=1}^{n} \sigma^2_{n,k}$. *Assume that* $B_n > 0$ *for each* $n \geq 1$, *and* $B_n \nearrow \infty$, *as* $n \to \infty$. *If, for every* $\epsilon > 0$,

$$\frac{1}{B^2_n}\sum_{k=1}^{n} E\{X^2_{n,k} I\{|X_{n,k}| > \epsilon B_n\} \to 0 \tag{1.12.8}$$

as $n \to \infty$, *then* $S_n/B_n \xrightarrow{d} N(0, 1)$ *as* $n \to \infty$. *Conversely, if* $\max_{1\leq k\leq n} \frac{\sigma^2_{n,k}}{B^2_n} \to 0$ *as* $n \to \infty$ *and* $S_n/B_n \xrightarrow{d} N(0, 1)$, *then (1.12.8) holds.*

For a proof, see Shiryayev (1984, p. 326). The following theorem, known as Lyapunov's Theorem, is weaker than the Lindeberg–Feller Theorem, but is often sufficient to establish the CLT.

Theorem 1.12.4 (Lyapunov). *Let $\{X_n\}$ be a sequence of independent random variables. Assume that $E\{X_n\} = 0$, $V\{X_n\} > 0$ and $E\{|X_n|^3\} < \infty$, for all $n \geq 1$. Moreover, assume that $B_n^2 = \sum_{j=1}^{n} V\{X_j\} \nearrow \infty$. Under the condition*

$$\frac{1}{B_n^3} \sum_{j=1}^{n} E\{|X_j|^3\} \to 0 \ \ as \ \ n \to \infty, \tag{1.12.9}$$

the CLT holds, i.e., $S_n/B_n \xrightarrow{d} N(0, 1)$ as $n \to \infty$.

Proof. It is sufficient to prove that (1.12.9) implies the Lindberg–Feller condition (1.12.8). Indeed,

$$\begin{aligned} E\{|X_j|^3\} &= \int_{-\infty}^{\infty} |x|^3 dF_j(x) \\ &\geq \int_{\{x:|x|>\epsilon B_n\}} |x|^3 dF_j(x) \\ &\geq \epsilon B_n \int_{\{x:|x|>B_n \epsilon\}} x^2 dF_j(x). \end{aligned}$$

Thus,

$$\frac{1}{B_n^2} \sum_{j=1}^{n} \int_{\{x:|x|>\epsilon B_n\}} x^2 dF_j(x) \leq \frac{1}{\epsilon} \cdot \frac{1}{B_n^3} \sum_{j=1}^{n} E\{|X_j|^3\} \to 0.$$

QED

Stein (1986, p. 97) proved, using a novel approximation to expectation, that if $X_1, X_2, \ldots$ are independent and identically distributed, with $EX_1 = 0$, $EX_1^2 = 1$ and $\gamma = E\{|X_1|^3\} < \infty$, then, for all $-\infty < x < \infty$ and all $n = 1, 2, \ldots,$

$$\left| P \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \leq x \right\} - \Phi(x) \right| \leq \frac{6\gamma}{\sqrt{n}},$$

where $\Phi(x)$ is the c.d.f. of $N(0, 1)$. This immediately implies the CLT and shows that the convergence is uniform in x.

1.13 MISCELLANEOUS RESULTS

In this section we review additional results.

1.13.1 Law of the Iterated Logarithm

We denote by $\log_2(x)$ the function $\log(\log(x))$, $x > e$.

Theorem 1.13.1. *Let $\{X_n\}$ be a sequence of i.i.d. random variables, such that $E\{X_1\} = 0$ and $V\{X_1\} = \sigma^2$, $0 < \sigma < \infty$. Let $S_n = \sum_{i=1}^{n} X_i$. Then*

$$P\left\{\overline{\lim_{n\to\infty}} \frac{|S_n|}{\psi(n)} = 1\right\} = 1, \tag{1.13.1}$$

where $\psi(n) = (2\sigma^2 n \log_2(n))^{1/2}$, $n \geq 3$.

For proof, in the normal case, see Shiryayev (1984, p. 372).

The theorem means the sequence $|S_n|$ will cross the boundary $\psi(n)$, $n \geq 3$, only a finite number of times, with probability 1, as $n \to \infty$. Notice that although $E\{S_n\} = 0$, $n \geq 1$, the variance of S_n is $V\{S_n\} = n\sigma^2$ and $P\{|S_n| \nearrow \infty\} = 1$. However, if we consider $\frac{S_n}{n}$ then by the SLLN, $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$. If we divide only by $\sqrt{n}$ then, by the CLT, $\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1)$. The law of the iterated logarithm says that, for every $\epsilon > 0$, $P\left\{\frac{|S_n|}{\sigma\sqrt{n}} > (1+\epsilon)\sqrt{2\log_2(n)}, i.o.\right\} = 0$. This means, that the fluctuations of S_n are not too wild. In Example 1.19 we see that if $\{X_n\}$ are i.i.d. with $P\{X_1 = 1\} = P\{X_1 = -1\} = \frac{1}{2}$, then $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. But n goes to infinity faster than $\sqrt{n \log_2(n)}$. Thus, by (1.13.1), if we consider the sequence $W_n = \frac{S_n}{\sqrt{2n\log_2(n)}}$ then $P\{|W_n| < 1 + \epsilon, i.o.\} = 1$. $\{W_n\}$ fluctuates between -1 and 1 almost always.

1.13.2 Uniform Integrability

A sequence of random variables $\{X_n\}$ is **uniformly integrable** if

$$\lim_{c\to\infty} \sup_{n\geq 1} E\{|X_n| I\{|X_n| > c\}\} = 0. \tag{1.13.2}$$

Clearly, if $|X_n| \leq Y$ for all $n \geq 1$ and $E\{Y\} < \infty$, then $\{X_n\}$ is a uniformly integrable sequence. Indeed, $|X_n| I\{|X_n| > c\} \leq |Y| I\{|Y| > c\}$ for all $n \geq 1$. Hence,

$$\sup_{n\geq 1} E\{|X_n| I\{|X_n| > c\}\} \leq E\{|Y| I\{|Y| > c\}\} \to 0$$

as $c \to \infty$ since $E\{Y\} < \infty$.

Theorem 1.13.2. *Let $\{X_n\}$ be uniformly integrable. Then,*

(i)
$$E\{\varliminf_{n\to\infty} X_n\} \le \varliminf_{n\to\infty} E\{X_n\} \le \varlimsup_{n\to\infty} E\{X_n\} \le E\{\varlimsup_{n\to\infty} X_n\}; \tag{1.13.3}$$

(ii) *if in addition $X_n \xrightarrow{a.s.} X$, as $n \to \infty$, then X is integrable and*

$$\lim_{n\to\infty} E\{X_n\} = E\{X\}, \tag{1.13.4}$$

$$\lim_{n\to\infty} E\{|X_n - X|\} = 0. \tag{1.13.5}$$

Proof. (**i**) For every $c > 0$

$$E\{X_n\} = E\{X_n I\{X_n < -c\}\} + E\{X_n I\{X_n \ge -c\}\}. \tag{1.13.6}$$

By uniform integrability, for every $\epsilon > 0$, take c sufficiently large so that

$$\sup_{n\ge 1} |E\{X_n I\{X_n < -c\}\}| < \epsilon.$$

By Fatou's Lemma (Theorem 1.6.2),

$$\varliminf_{n\to\infty} E\{X_n I\{X_n \ge -c\}\} \ge E\left\{\varliminf_{n\to\infty} X_n I\{X_n \ge -c\}\right\}. \tag{1.13.7}$$

But $X_n I\{X_n \ge -c\}\} \ge X_n$. Therefore,

$$\varliminf_{n\to\infty} E\{X_n I\{X_n \ge -c\}\} \ge E\{\varliminf_{n\to\infty} X_n\}. \tag{1.13.8}$$

From (1.13.6)–(1.13.8), we obtain

$$\varliminf_{n\to\infty} E\{X_n\} \ge E\left\{\varliminf_{n\to\infty} X_n\right\} - \epsilon. \tag{1.13.9}$$

In a similar way, we show that

$$\varlimsup_{n\to\infty} E\{X_n\} \le E\{\varlimsup X_n\} + \epsilon. \tag{1.13.10}$$

Since ϵ is arbitrary we obtain (1.13.3). Part (ii) is obtained from (i) as in the Dominated Convergence Theorem (Theorem 1.6.3). QED

Theorem 1.13.3. *Let $X_n \ge 0$, $n \ge 1$, and $X_n \xrightarrow{a.s.} X$, $E\{X_n\} < \infty$. Then $E\{X_n\} \to E\{X\}$ if and only if $\{X_n\}$ is uniformly integrable.*

Proof. The sufficiency follows from part (ii) of the previous theorem.

To prove necessity, let

$$A = \{a : F_X(a) - F_X(a-) > 0\}.$$

Then, for each $c \notin A$

$$X_n I\{X_n < c\} \xrightarrow{\text{a.s.}} XI\{X < c\}.$$

The family $\{X_n I\{X_n < c\}\}$ is uniformly integrable. Hence, by sufficiency,

$$\lim_{n\to\infty} E\{X_n I\{X_n < c\}\} = E\{XI\{X < c\}\}$$

for $c \notin A$, $n \to \infty$. A has a countable number of jump points. Since $E\{X\} < \infty$, we can choose $c_0 \notin A$ sufficiently large so that, for a given $\epsilon > 0$, $E\{XI\{X \geq c_0\}\} < \frac{\epsilon}{2}$. Choose $N_0(\epsilon)$ sufficiently large so that, for $n \geq N_0(\epsilon)$,

$$E\{X_n I\{X_n \geq c_0\}\} \leq E\{XI\{X \geq c_0\}\} + \frac{\epsilon}{2}.$$

Choose $c_1 > c_0$ sufficiently large so that $E\{X_n I\{X_n \geq c_1\}\} \leq \epsilon$, $n \leq N_0$. Then $\sup_n E\{X_n I\{X_n \geq c_1\}\} \leq \epsilon$. QED

Lemma 1.13.1. *If $\{X_n\}$ is a sequence of uniformly integrable random variables, then*

$$\sup_{n\geq 1} E\{|X_n|\} < \infty. \tag{1.13.11}$$

Proof.

$$\begin{aligned}
\sup_{n\geq 1} E\{|X|_n\} &= \sup_{n\geq 1}(E\{|X_n|I\{|X_n| > c\}\} + E\{|X_n|I\{|X_n| \leq c\}\}) \\
&\leq \sup_{n\geq 1} E\{|X_n|I\{|X_n| > c\}\} + \sup_{n\geq 1} E\{|X_n|I\{|X_n| \leq c\}\} \\
&\leq \epsilon + c,
\end{aligned}$$

for $0 < c < \infty$ sufficiently large. QED

Theorem 1.13.4. *A necessary and sufficient condition for a sequence $\{X_n\}$ to be uniformly integrable is that*

$$\sup_{n\geq 1} E\{|X_n|\} \leq B < \infty \tag{1.13.12}$$

and

$$\sup_{n\geq 1} E\{|X_n|I_A\} \to 0 \ \textit{ when } \ P\{A\} \to 0. \tag{1.13.13}$$

Proof. (**i**) **Necessity**: Condition (1.13.12) was proven in the previous lemma. Furthermore, for any $0 < c < \infty$,

$$\begin{aligned} E\{|X_n|I_A\} &= E\{|X_n|I\{A \cap \{|X_n| \geq c\}\}\} \\ &\quad + E\{|X_n|I\{A \cap \{|X_n| < c\}\}\} \\ &\leq E\{|X_n|I\{|X_n| \geq c\}\} + cP(A). \end{aligned} \tag{1.13.14}$$

Choose c sufficiently large, so that $E\{|X_n|I\{|X_n| \geq c\}\} < \dfrac{\epsilon}{2}$ and A so that $P\{A\} < \dfrac{\epsilon}{2c}$, then $E\{|X_n|I_A\} < \epsilon$. This proves the necessity of (1.13.13).

(**ii**) **Sufficiency**: Let $\epsilon > 0$ be given. Choose $\delta(\epsilon)$ so that $P\{A\} < \delta(\epsilon)$, and $\sup_{n\geq 1} E\{|X_n|I_A\} \leq \epsilon$.

By Chebychev's inequality, for every $c > 0$,

$$P\{|X_n| \geq c\} \leq \frac{E\{|X_n|\}}{c}, \quad n \geq 1.$$

Hence,

$$\sup_{n\geq 1} P\{|X_n| \geq c\} \leq \frac{1}{c} \sup_{n\geq 1} E\{|X_n|\} \leq \frac{B}{c}. \tag{1.13.15}$$

The right-hand side of (1.13.15) goes to zero, when $c \to \infty$. Choose c sufficiently large so that $P\{|X_n| \geq c\} < \epsilon$. Such a value of c exists, independently of n, due to (1.13.15). Let $A = \left\{\bigcup_{n=1}^{\infty} |X_n| \geq c\right\}$. For sufficiently large c, $P\{A\} < \epsilon$ and, therefore,

$$\sup_{n\geq 1} E\{|X_n|I\{|X_n| \geq c\}\} \leq E\{|X_n|I_A\} \to 0$$

as $c \to \infty$. This establishes the uniform integrability of $\{X_n\}$. QED

Notice that according to Theorem 1.13.3, if $E|X_n|^r < \infty$, $r \geq 1$ and $X_n \xrightarrow{\text{a.s.}} X$, $\lim_{n\to\infty} E\{X_n^r\} = E\{X^r\}$ if and only if $\{X_n\}$ is a uniformly integrable sequence.

1.13.3 Inequalities

In previous sections we established several inequalities. The Chebychev inequality, the Kolmogorov inequality. In this section we establish some useful additional inequalities.

1. **The Schwarz Inequality**

Let (X, Y) be random variables with joint distribution function F_{XY} and marginal distribution functions F_X and F_Y, respectively. Then, for every Borel measurable and integrable functions g and h, such that $E\{g^2(X)\} < \infty$ and $E\{h^2(Y)\} < \infty$,

$$\left| \int g(x)h(y)dF_{XY}(x, y) \right| \leq \left(\int g^2(x)dF_X(x) \right)^{1/2} \left(\int h^2(y)dF_Y(y) \right)^{1/2}. \quad (1.13.16)$$

To prove (1.13.16), consider the random variable $Q(t) = (g(X) + th(Y))^2$, $-\infty < t < \infty$. Obviously, $Q(t) \geq 0$, for all t, $-\infty < t < \infty$. Moreover,

$$E\{Q(t)\} = E\{g^2(X)\} + 2tE\{g(X)h(Y)\} + t^2 E\{h^2(Y)\} \geq 0$$

for all t. But, $E\{Q(t)\} \geq 0$ for all t if and only if

$$(E\{g(X)h(Y)\})^2 \leq E\{g^2(X)\}E\{h^2(Y)\}.$$

This establishes (1.13.16).

2. **Jensen's Inequality**

A function $g : \mathbb{R} \to \mathbb{R}$ is called **convex** if, for any $-\infty < x < y < \infty$ and $0 \leq \alpha \leq 1$,

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

Suppose X is a random variable and $E\{|X|\} < \infty$. Then, if g is convex,

$$g(E\{X\}) \leq E\{g(X)\}. \quad (1.13.17)$$

To prove (1.13.17), notice that since g is convex, for every x_0, $-\infty < x_0 < \infty$, $g(x) \geq g(x_0) + (x - x_0)g^*(x_0)$ for all x, $-\infty < x < \infty$, where $g^*(x_0)$ is finite. Substitute $x_0 = E\{X\}$. Then

$$g(X) \geq g(E\{X\}) + (X - E\{X\})g^*(E\{X\})$$

with probability one. Since $E\{X - E\{X\}\} = 0$, we obtain (1.13.17).

3. **Lyapunov's Inequality**
   If $0 < s < r$ and $E\{|X|^r\} < \infty$, then

$$(E\{|X|^s\})^{1/s} \leq (E\{|X|^r\})^{1/r}. \tag{1.13.18}$$

To establish this inequality, let $t = r/s$. Notice that $g(x) = |x|^t$ is convex, since $t > 1$. Let $\xi = E\{|X|^s\}$, and $(|X|^s)^t = |X|^r$. Thus, by Jensen's inequality,

$$\begin{aligned} g(\xi) = (E|X|^s)^{r/s} &\leq E\{g(|X|^s)\} \\ &= E\{|X|^r\}. \end{aligned}$$

Hence, $E\{|X|^s\}^{1/s} \leq (E\{|X|^r\})^{1/r}$. As a result of Lyapunov's inequality we have the following chain of inequalities among absolute moments.

$$E\{|X|\} \leq (E\{X^2\})^{1/2} \leq (E\{|X|^3\})^{1/3} \leq \cdots . \tag{1.13.19}$$

4. **Hölder's Inequality**
   Let $1 < p < \infty$ and $1 < q < \infty$, such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$. $E\{|X|^p\} < \infty$ and $E\{|Y|^q\} < \infty$. Then

$$E\{|XY|\} \leq (E\{|X|^p\})^{1/p}(E\{|Y|^q\})^{1/q}. \tag{1.13.20}$$

Notice that the Schwarz inequality is a special case of Holder's inequality for $p = q = 2$.

For proof, see Shiryayev (1984, p. 191).

5. **Minkowsky's Inequality**
   If $E\{|X|^p\} < \infty$ and $E\{|Y|^p\} < \infty$ for some $1 \leq p < \infty$, then $E\{|X + Y|^p\} < \infty$ and

$$(E\{|X + Y|^p\})^{1/p} \leq (E|X|^p)^{1/p} + (E\{|Y|^p\})^{1/p}. \tag{1.13.21}$$

For proof, see Shiryayev (1984, p. 192).

1.13.4 The Delta Method

The delta method is designed to yield large sample approximations to nonlinear functions g of the sample mean $\bar{X}_n$ and its variance. More specifically, let $\{X_n\}$ be a sequence of i.i.d. random variables. Assume that $0 < V\{X\} < \infty$. By the SLLN, $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$, as $n \to \infty$, where $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$, and by the CLT, $\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{d} N(0, 1)$. Let $g : \mathbb{R} \to \mathbb{R}$ having third order continuous derivative. By the Taylor expansion of $g(\bar{X}_n)$ around μ,

$$g(\bar{X}_n) = g(\mu) + (\bar{X}_n - \mu)g^{(1)}(\mu) + \frac{1}{2}(\bar{X}_n - \mu)^2 g^{(2)}(\mu) + R_n, \quad (1.13.22)$$

where $R_n = \frac{1}{6}(\bar{X}_n - \mu)^3 g^{(3)}(\mu_n^*)$, where μ_n^* is a point between $\bar{X}_n$ and μ, i.e., $|\bar{X}_n - \mu_n^*| < |\bar{X}_n - \mu|$. Since we assumed that $g^{(3)}(x)$ is continuous, it is bounded on the closed interval $[\mu - \Delta, \mu + \Delta]$. Moreover, $g^{(3)}(\mu_n^*) \xrightarrow{\text{a.s.}} g^{(3)}(\mu)$, as $n \to \infty$. Thus $R_n \xrightarrow{p} 0$, as $n \to \infty$. The distribution of $g(\mu) + g^{(1)}(\mu)(\bar{X}_n - \mu)$ is asymptotically $N(g(\mu), (g^{(1)}(\mu))^2\sigma^2/n)$. $(\bar{X}_n - \mu)^2 \xrightarrow{2} 0$, as $n \to \infty$. Thus, $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} N(0, \sigma^2(g^{(1)}(\mu))^2)$. Thus, if $\bar{X}_n$ satisfies the CLT, an approximation to the expected value of $g(\bar{X}_n)$ is

$$E\{g(\bar{X}_n)\} \cong g(\mu) + \frac{\sigma^2}{2n} g^{(2)}(\mu). \quad (1.13.23)$$

An approximation to the variance of $g(\bar{X}_n)$ is

$$V\{g(\bar{X}_n)\} \cong \frac{\sigma^2}{n}(g^{(1)}(\mu))^2. \quad (1.13.24)$$

Furthermore, from (1.13.22)

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) = \sqrt{n}(\bar{X}_n - \mu)g^{(1)}(\mu) + D_n, \quad (1.13.25)$$

where

$$D_n = \frac{(\bar{X}_n - \mu)^2}{2} g^{(2)}(\mu_n^{**}), \quad (1.13.26)$$

and $|\mu_n^{**} - \bar{X}_n| \le |\mu - \bar{X}_n|$ with probability one. Thus, since $\bar{X}_n - \mu \to 0$ a.s., as $n \to \infty$, and since $|g^{(2)}(\mu_n^{**})|$ is bounded, $D_n \xrightarrow{p} 0$, as $n \to \infty$, then

$$\sqrt{n}\,\frac{g(\bar{X}_n) - g(\mu)}{\sigma|g^{(1)}(\mu)|} \xrightarrow{d} N(0, 1). \quad (1.13.27)$$

1.13.5 The Symbols o_p and O_p

Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables, $Y_n > 0$ a.s. for all $n \geq 1$. We say that $X_n = o_p(Y_n)$, i.e., X_n is of a smaller order of magnitude than Y_n in probability if

$$\frac{X_n}{Y_n} \xrightarrow{p} 0 \quad \text{as } n \to \infty. \tag{1.13.28}$$

We say that $X_n = O_p(Y_n)$, i.e., X_n has the same order of magnitude in probability as Y_n if, for all $\epsilon > 0$, there exists K_ϵ such that $\sup_n P\left\{\left|\frac{X_n}{Y_n}\right| > K_\epsilon\right\} < \epsilon$.

One can verify the following relations.

$$\begin{aligned}
&\textbf{(i)} \quad o_p(1) + O_p(1) = O_p(1),\\
&\textbf{(ii)} \quad O_p(1) + O_p(1) = O_p(1),\\
&\textbf{(iii)} \quad o_p(1) + o_p(1) = o_p(1),\\
&\textbf{(iv)} \quad O_p(1) \cdot O_p(1) = O_p(1),\\
&\textbf{(v)} \quad o_p(1) \cdot O_p(1) = o_p(1).
\end{aligned} \tag{1.13.29}$$

1.13.6 The Empirical Distribution and Sample Quantiles

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having a distribution F. The function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leq x\} \tag{1.13.30}$$

is called the **empirical distribution function** (EDF).

Notice that $E\{I\{X_i \leq x\}\} = F(x)$. Thus, the SLLN implies that at each x, $F_n(x) \xrightarrow{\text{a.s.}} F(x)$ as $n \to \infty$. The question is whether this convergence is uniform in x. The answer is given by

Theorem 1.13.5 (Glivenko–Cantelli). *Let $X_1, X_2, X_3, \ldots$ be i.i.d. random variables. Then*

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0, \quad \textit{as } n \to \infty. \tag{1.13.31}$$

For proof, see Sen and Singer (1993, p. 185).

The pth sample quantile $x_{n,p}$ is defined as

$$\begin{aligned} x_{n,p} &= F_n^{-1}(p) \\ &= \inf\{x : F_n(x) \geq p\} \end{aligned} \tag{1.13.32}$$

for $0 < p < 1$, where $F_n(x)$ is the EDF. When $F(x)$ is continuous then, the points of increase of $F_n(x)$ are the order statistics $X_{(1:n)} < \cdots < X_{(n:n)}$ with probability one. Also, $F_n(X_{(i:n)}) = \dfrac{i}{n}$, $i = 1, \ldots, n$. Thus,

$$\begin{aligned} &x_{n,p} = X_{(i(p):n)}, \quad \text{where} \\ &i(p) = \text{smallest integer } i \ \text{ such that } \ i \geq pn. \end{aligned} \tag{1.13.33}$$

Theorem 1.13.6. *Let F be a continuous distribution function, and $\xi_p = F^{-1}(p)$, and suppose that $F(\xi_p) = p$ and for any $\epsilon > 0$, $F(\xi_p - \epsilon) < p < F(\xi_p + \epsilon)$. Let $X_1, \ldots, X_n$ be i.i.d. random variables from this distribution. Then*

$$x_{n,p} \xrightarrow{a.s.} \xi_p \ \text{ as } \ n \to \infty.$$

For proof, see Sen and Singer (1993, p. 167).

The following theorem establishes the asymptotic normality of $x_{n,p}$.

Theorem 1.13.7. *Let $F(x)$ be an absolutely continuous distribution, with continuous p.d.f. $f(x)$. Let p, $0 < p < 1$, $\xi_p = F^{-1}(p)$ and $f(\xi_p) > 0$. Then*

$$\sqrt{n}(x_{n,p} - \xi_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{f^2(\xi_p)}\right). \tag{1.13.34}$$

For proof, see Sen and Singer (1993, p. 168).

The results of Theorems 1.13.6–1.13.7 will be used in Chapter 7 to establish the asymptotic relative efficiency of the sample median, relative to the sample mean.

PART II: EXAMPLES

Example 1.1. We illustrate here two algebras.

The sample space is finite

$$\mathcal{S} = \{1, 2, \ldots, 10\}.$$

Let $E_1 = \{1, 2\}$, $E_2 = \{9, 10\}$. The algebra generated by E_1 and E_2, $\mathcal{A}_1$, contains the events

$$\mathcal{A}_1 = \{\mathcal{S}, \emptyset, E_1, \bar{E}_1, E_2, \bar{E}_2, E_1 \cup E_2, \overline{E_1 \cup E_2}\}.$$

The algebra generated by the partition $\mathcal{D} = \{E_1, E_2, E_3, E_4\}$, where $E_1 = \{1, 2\}$, $E_2 = \{9, 10\}$, $E_3 = \{3, 4, 5\}$, $E_4 = \{6, 7, 8\}$ contains the $2^4 = 16$ events

$$\mathcal{A}_2 = \{\mathcal{S}, \emptyset, E_1, E_2, E_3, E_4, E_1 \cup E_2, E_1 \cup E_3, E_1 \cup E_4, E_2 \cup E_3, E_2 \cup E_4,$$
$$E_3 \cup E_4, E_1 \cup E_2 \cup E_3, E_1 \cup E_2 \cup E_4, E_1 \cup E_3 \cup E_4, E_2 \cup E_3 \cup E_4\}.$$

Notice that the complement of each set in $\mathcal{A}_2$ is in $\mathcal{A}_2$. $\mathcal{A}_1 \subset \mathcal{A}_2$. Also, $\mathcal{A}_2 \subset \mathcal{A}(\mathcal{S})$. ■

Example 1.2. In this example we consider a **random walk** on the integers. Consider an experiment in which a particle is initially at the origin, 0. In the first trial the particle moves to +1 or to −1. In the second trial it moves either one integer to the right or one integer to the left. The experiment consists of $2n$ such trials ($1 \le n < \infty$). The sample space $\mathcal{S}$ is finite and there are 2^{2n} points in $\mathcal{S}$, i.e., $\mathcal{S} = \{(i_1, \ldots, i_{2n}) : i_j = \pm 1, j = 1, \ldots, 2n\}$. Let $E_j = \left\{(i_1, \ldots, i_{2n}) : \sum_{k=1}^{2n} i_k = j\right\}$, $j = 0, \pm 2, \pm, \cdots, \pm 2n$. E_j is the event that, at the end of the experiment, the particle is at the integer j. Obviously, $-2n \le j \le 2n$. It is simple to show that j must be an even integer $j = \pm 2k$, $k = 0, 1, \ldots, n$. Thus, $\mathcal{D} = \{E_{2k}, k = 0, \pm 1, \ldots, \pm n\}$ is a partition of $\mathcal{S}$. The event E_{2k} consists of all elementary events in which there are $(n + k)$ +1s and $(n - k)$ −1s. Thus, E_{2k} is the union of $\binom{2n}{n+k}$ points of $\mathcal{S}$, $k = 0, \pm 1, \ldots, \pm n$.

The algebra generated by $\mathcal{D}$, $\mathcal{A}(\mathcal{D})$, consists of $\emptyset$ and $2^{2n+1} - 1$ unions of the elements of $\mathcal{D}$. ■

Example 1.3. Let $\mathcal{S}$ be the real line, i.e., $\mathcal{S} = \{x : -\infty < x < \infty\}$. We construct an algebra $\mathcal{A}$ generated by half-closed intervals: $E_x = (-\infty, x]$, $-\infty < x < \infty$. Notice that, for $x < y$, $E_x \cup E_y = (-\infty, y]$. The complement of E_x is $\bar{E}_x = (x, \infty)$. We will adopt the convention that $(x, \infty) \equiv (x, \infty]$.

Consider the sequence of intervals $E_n = \left(-\infty, 1 - \frac{1}{n}\right]$, $n \ge 1$. All $E_n \in \mathcal{A}$. However, $\bigcup_{n=1}^{\infty} E_n = (-\infty, 1)$. Thus $\lim_{n\to\infty} E_n$ **does not** belong to $\mathcal{A}$. $\mathcal{A}$ is **not** a σ-field. In order to make $\mathcal{A}$ into a σ-field we have to add to it all limit sets of sequences of events in $\mathcal{A}$. ■

Example 1.4. We illustrate here three events that are only pairwise independent.

Let $\mathcal{S} = \{1, 2, 3, 4\}$, with $P(w) = \frac{1}{4}$, for all $w \in \mathcal{S}$. Define the three events

$$A_1 = \{1, 2\}, \quad A_2 = \{1, 3\}, \quad A_3 = \{1, 4\}.$$

$P\{A_i\} = \frac{1}{2}, i = 1, 2, 3.$

$$A_1 \cap A_2 = \{1\}.$$
$$A_1 \cap A_3 = \{1\}.$$
$$A_2 \cap A_3 = \{1\}.$$

Thus

$$P\{A_1 \cap A_2\} = \frac{1}{4} = P\{A_1\}P\{A_2\}.$$
$$P\{A_1 \cap A_3\} = \frac{1}{4} = P\{A_1\}P\{A_3\}.$$
$$P\{A_2 \cap A_3\} = \frac{1}{4} = P\{A_2\}P\{A_3\}.$$

Thus, A_1, A_2, A_3 are pairwise independent. On the other hand,

$$A_1 \cap A_2 \cap A_3 = \{1\}$$

and

$$P\{A_1 \cap A_2 \cap A_3\} = \frac{1}{4} \neq P\{A_1\}P\{A_2\}P\{A_3\} = \frac{1}{8}.$$

Thus, the triplet (A_1, A_2, A_3) is not independent. ■

Example 1.5. An infinite sequence of trials, in which each trial results in either "success" S or "failure" F is called **Bernoulli trials** if all trials are independent and the probability of success in each trial is the same. More specifically, consider the sample space of countable sequences of Ss and Fs, i.e.,

$$\mathcal{S} = \{(i_1, i_2, \ldots) : i_j = S, F, \ \ j = 1, 2, \ldots\}.$$

Let

$$E_j = \{(i_1, i_2, \ldots) : i_j = S\}, \ \ j = 1, 2, \ldots.$$

We assume that $\{E_1, E_2, \ldots, E_n\}$ are mutually independent for all $n \geq 2$ and $P\{E_j\} = p$ for all $j = 1, 2, \ldots, 0 < p < 1$.

The points of $\mathcal{S}$ represent an infinite sequence of **Bernoulli trials**. Consider the events

$$A_j = \{(i_1, i_2, \ldots) : i_j = S, i_{j+1} = F, i_{j+2} = S\}$$
$$= E_j \cap \bar{E}_{j+1} \cap E_{j+2}$$

$j = 1, 2, \ldots$. $\{A_j\}$ are not independent.

Let $B_j = \{A_{3j+1}\}$, $j \geq 0$. The sequence $\{B_j, j \geq 1\}$ consists of mutually independent events. Moreover, $P(B_j) = p^2(1-p)$ for all $j = 1, 2, \ldots$. Thus, $\sum_{j=1}^{\infty} P(B_j) = \infty$ and the Borel–Cantelli Lemma implies that $P\{B_n, \text{i.o.}\} = 1$. That is, the pattern SFS will occur infinitely many times in a sequence of Bernoulli trials, with probability one. ■

Example 1.6. Let $\mathcal{S}$ be the sample space of $N = 2^n$ binary sequences of size n, $n < \infty$, i.e.,

$$\mathcal{S} = \{(i_1, \ldots, i_n) : i_j = 0, 1, \;\; j = 1, \ldots, n\}.$$

We assign the points $w = (i_1, \ldots, i_n)$ of $\mathcal{S}$, equal probabilities, i.e., $P\{(i_1, \ldots, i_n)\} = 2^{-n}$. Consider the partition $\mathcal{D} = \{B_0, B_1, \ldots, B_n\}$ to $k = n+1$ disjoint events, such that

$$B_j = \{(i_1, \ldots, i_n) : \sum_{l=1}^{n} i_l = j\}, \quad j = 0, \ldots, n.$$

B_j is the set of all points having exactly j ones and $(n-j)$ zeros. We define the discrete random variable corresponding to $\mathcal{D}$ as

$$X(w) = \sum_{j=0}^{n} j I_{B_j}(w).$$

The jump points of $X(w)$ are $\{0, 1, \ldots, n\}$. The probability distribution function of $X(w)$ is

$$f_X(x) = \sum_{j=0}^{n} I_{\{j\}}(x) P\{B_j\}.$$

It is easy to verify that

$$P\{B_j\} = \binom{n}{j} 2^{-n}, \quad j = 0, 1, \ldots, n$$

where

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}, \quad j = 0, 1, \dots, n.$$

Thus,

$$f_X(x) = \sum_{j=0}^{n} I_{\{j\}}(x) \binom{n}{x} 2^{-n}.$$

The distribution function (c.d.f.) is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{j=0}^{[x]} \binom{n}{j} 2^{-n}, & \end{cases}$$

where $[x]$ is the maximal integer value smaller or equal to x. The distribution function illustrated here is called a **binomial distribution** (see Section 2.2.1). ■

Example 1.7. Consider the random variable of Example 1.6. In that example $X(w) \in \{0, 1, \dots, n\}$ and $f_X(j) = \binom{n}{j} 2^{-n}$, $j = 0, \dots, n$. Accordingly,

$$E\{X\} = \sum_{j=0}^{n} j \binom{n}{j} 2^{-n} = \frac{n}{2} \sum_{j=0}^{n-1} \binom{n-1}{j} 2^{-(n-1)} = \frac{n}{2}.$$

■

Example 1.8. Let $(\mathcal{S}, \mathcal{F}, P)$ be a probability space where $\mathcal{S} = \{0, 1, 2, \dots\}$. $\mathcal{F}$ is the σ-field of all subsets of $\mathcal{S}$. Consider $X(w) = w$, with probability function

$$\begin{aligned} p_j &= P\{w : X(w) = j\} \\ &= e^{-\lambda} \frac{\lambda^j}{j!}, \quad j = 0, 1, 2, \dots \end{aligned}$$

for some λ, $0 < \lambda < \infty$. $0 < p_j < \infty$ for all j, and since $\sum_{j=0}^{\infty} \frac{\lambda^j}{j} = e^{\lambda}$, $\sum_{j=0}^{\infty} p_j = 1$.

Consider the partition $\mathcal{D} = \{A_1, A_2, A_3\}$ where $A_1 = \{w : 0 \leq w \leq 10\}$, $A_2 = \{w : 10 < w \leq 20\}$ and $A_3 = \{w : w \geq 21\}$. The probabilities of these sets are

$$q_1 = P\{A_1\} = e^{-\lambda} \sum_{j=0}^{10} \frac{\lambda^j}{j!},$$

$$q_2 = P\{A_2\} = e^{-\lambda} \sum_{j=11}^{20} \frac{\lambda^j}{j!}, \quad \text{and}$$

$$q_3 = P\{A_3\} = e^{-\lambda} \sum_{j=21}^{\infty} \frac{\lambda^j}{j!}.$$

The conditional distributions of X given A_i $i = 1, 2, 3$ are

$$f_{X|A_i}(x) = \frac{\dfrac{\lambda^x}{x!} I_{A_i}(x)}{\displaystyle\sum_{j=b_{i-1}}^{b_i-1} \frac{\lambda^j}{j!}}, \quad i = 1, 2, 3$$

where $b_0 = 0, b_1 = 11, b_2 = 21, b_3 = \infty$.

The conditional expectations are

$$E\{X \mid A_i\} = \lambda \frac{\displaystyle\sum_{j=(b_{i-1}-1)^+}^{b_i-2} \frac{\lambda^j}{j!}}{\displaystyle\sum_{j=b_{i-1}}^{b_i-1} \frac{\lambda^j}{j!}}, \quad i = 1, 2, 3$$

where $a+ = \max(a, 0)$. $E\{X \mid \mathcal{D}\}$ is a random variable, which obtains the values $E\{X \mid A_1\}$ with probability q_1, $E\{X \mid A_2\}$ with probability q_2, and $E\{X \mid A_3\}$ with probability q_3. ■

Example 1.9. Consider two discrete random variables X, Y on $(\mathcal{S}, \mathcal{F}, P)$ such that the jump points of X and Y are the nonnegative integers $\{0, 1, 2, \ldots\}$. The joint probability function of (X, Y) is

$$f_{XY}(x, y) = \begin{cases} e^{-\lambda} \dfrac{\lambda^y}{(y+1)!}, & x = 0, 1, \ldots, y;\ y = 0, 1, 2, \ldots \\ 0, & \text{otherwise,} \end{cases}$$

where λ, $0 < \lambda < \infty$, is a specified parameter.

First, we have to check that

$$\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} f_{XY}(x, y) = 1.$$

Indeed,

$$\begin{aligned} f_Y(y) &= \sum_{x=0}^{y} f_{XY}(x, y) \\ &= e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, \ldots \end{aligned}$$

and

$$\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

The conditional p.d.f. of X given $\{Y = y\}$, $y = 0, 1, \ldots$ is

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{1}{1+y}, & x = 0, 1, \ldots, y \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned} E\{X \mid Y = y\} &= \frac{1}{1+y} \sum_{x=0}^{y} x \\ &= \frac{y}{2}, \quad y = 0, 1, \ldots \end{aligned}$$

and, as a random variable,

$$E\{X \mid Y\} = \frac{Y}{2}.$$

Finally,

$$E\{E\{X \mid Y\}\} = \sum_{y=0}^{\infty} \frac{y}{2} e^{-\lambda} \frac{\lambda^y}{y!} = \frac{\lambda}{2}.$$

■

Example 1.10. In this example we show an absolutely continuous distribution for which $E\{X\}$ **does not** exist.

Let $F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)$. This is called the **Cauchy distribution**. The density function (p.d.f.) is

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

It is a symmetric density around $x = 0$, in the sense that $f(x) = f(-x)$ for all x. The expected value of X having this distribution does not exist. Indeed,

$$\begin{aligned}\int_{-\infty}^{\infty} |x| f(x) dx &= \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2} dx \\ &= \frac{1}{\pi} \lim_{T\to\infty} \log(1+T^2) = \infty.\end{aligned}$$

■

Example 1.11. We show here a mixture of discrete and absolutely continuous distributions.

Let

$$F_{\text{ac}}(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - \exp\{-\lambda x\}, & \text{if } x \geq 0 \end{cases}$$

$$F_{\text{d}}(x) = \begin{cases} 0, & \text{if } x < 0 \\ e^{-\mu}\sum_{j=0}^{[x]} \frac{\mu^j}{j!}, & \text{if } x \geq 0 \end{cases}$$

where $[x]$ designates the maximal integer not exceeding x; λ and μ are real positive numbers. The mixed distribution is, for $0 \leq \alpha \leq 1$,

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ \alpha e^{-\mu}\sum_{j=0}^{[x]} \frac{\mu^j}{j!} + (1-\alpha)[1 - \exp(-\lambda x)], & \text{if } x \geq 0. \end{cases}$$

This distribution function can be applied with appropriate values of α, λ, and μ for modeling the length of telephone conversations. It has discontinuities at the nonnegative integers and is continuous elsewhere. ■

Example 1.12. Densities derived after transformations.

Let X be a random variable having an absolutely continuous distribution with p.d.f. f_X.

A. If $Y = X^2$, the number of roots are

$$m(y) = \begin{cases} 0, & \text{if } y < 0 \\ 1, & \text{if } y = 0 \\ 2, & \text{if } y > 0. \end{cases}$$

Thus, the density of Y is

$$f_Y(y) = \begin{cases} 0, & y \leq 0 \\ \dfrac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], & y > 0. \end{cases}$$

B. If $Y = \cos X$

$$m(y) = \begin{cases} 0, & \text{if } |y| > 1 \\ \infty, & \text{if } |y| \leq 1. \end{cases}$$

For every y, such that $|y| < 1$, let $\xi(y)$ be the value of $\cos^{-1}(y)$ in the interval $(0, \pi)$. Then, if $f(x)$ is the p.d.f. of X, the p.d.f. of $Y = \cos X$ is, for $|y| < 1$,

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}} \sum_{j=0}^{\infty} \{f_X(\xi(y) + 2\pi j) + \\ + f_X(\xi(y) - 2\pi j) + f_X(-\xi(y) + 2\pi j) + f_X(-\xi(y) - 2\pi j)\}.$$

The density does not exist for $|y| \geq 1$. ■

Example 1.13. Three cases of joint p.d.f.

A. Both X_1, X_2 are discrete, with jump points on $\{0, 1, 2, \ldots\}$. Their joint p.d.f. for $0 < \lambda < \infty$ is,

$$f_{X_1X_2}(x_1, x_2) = \binom{x_2}{x_1} 2^{-x_2} e^{-\lambda} \frac{\lambda^{x_2}}{x_2!},$$

for $x_1 = 0, \ldots, x_2$, $x_2 = 0, 1, \ldots$. The marginal p.d.f. are

$$f_{X_1}(x_1) = e^{-\lambda/2} \frac{(\lambda/2)^{x_1}}{x_1!}, \quad x_1 = 0, 1, \ldots \text{ and}$$

$$f_{X_2}(x_2) = e^{-\lambda} \frac{\lambda^{x_2}}{x_2!}, \quad x_2 = 0, 1, \ldots.$$

B. Both X_1 and X_2 are absolutely continuous, with joint p.d.f.

$$f_{X_1X_2}(x, y) = 2I_{(0,1)}(x)I_{(0,x)}(y).$$

The marginal distributions of X_1 and X_2 are

$$f_{X_1}(x) = 2x I_{(0,1)}(x) \text{ and}$$
$$f_{X_2}(y) = 2(1-y)I_{(0,1)}(y).$$

C. X_1 is discrete with jump points $\{0, 1, 2, \ldots\}$ and X_2 absolutely continuous. The joint p.d.f., with respect to the σ-finite measure $dN(x_1)dy$ is, for $0 < \lambda < \infty$,

$$f_{X_1X_2}(x, y) = e^{-\lambda}\frac{\lambda^x}{x!}\cdot\frac{1}{1+x}I\{x = 0, 1, \ldots\}I_{(0,1+x)}(y).$$

The marginal p.d.f. of X_1, is

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

The marginal p.d.f. of X_2 is

$$f_{X_2}(y) = \frac{1}{\lambda}\sum_{n=0}^{\infty}\left(1 - e^{-\lambda}\sum_{j=0}^{n}\frac{\lambda^j}{j!}\right)I_{(n,n+1)}(y).$$

■

Example 1.14. Suppose that X, Y are positive random variables, having a joint p.d.f.

$$f_{XY}(x, y) = \frac{1}{y}\lambda e^{-\lambda y}I_{(0,y)}(x), \quad 0 < y < \infty,\ 0 < x < y,\ 0 < \lambda < \infty.$$

The marginal p.d.f. of X is

$$f_X(x) = \lambda\int_x^{\infty}\frac{1}{y}e^{-\lambda y}dy$$
$$= \lambda E_1(\lambda x),$$

where $E_1(\xi) = \int_{\xi}^{\infty}\frac{1}{u}e^{-u}du$ is called the exponential integral, which is finite for all $\xi > 0$. Thus, according to (1.6.62), for $x_0 > 0$,

$$f_{Y|X}(y \mid x_0) = \frac{\frac{1}{y}e^{-\lambda y}I_{(x_0,\infty)}(y)}{E_1(\lambda x_0)}.$$

Finally, for $x_0 > 0$,

$$E\{Y \mid X = x_0\} = \frac{\int_{x_0}^{\infty} e^{-\lambda y} dy}{E_1(\lambda x_0)} = \frac{e^{-\lambda x_0}}{\lambda E_1(\lambda x_0)}.$$

■

Example 1.15. In this example we show a distribution function whose m.g.f., M, exists only on an interval $(-\infty, t_0)$. Let

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0, \end{cases}$$

where $0 < \lambda < \infty$. The m.g.f. is

$$M(t) = \lambda \int_0^{\infty} e^{tx - \lambda x} dx = \frac{\lambda}{\lambda - t} = \left(1 - \frac{t}{\lambda}\right)^{-1}, \quad -\infty < t < \lambda.$$

The integral in $M(t)$ is ∞ if $t \geq \lambda$. Thus, the domain of convergence of M is $(-\infty, \lambda)$. ■

Example 1.16. Let

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } (1 - p) \end{cases}$$

$i = 1, \ldots, n$. We assume also that $X_1, \ldots, X_n$ are independent. We wish to derive the p.d.f. of $S_n = \sum_{i=1}^{n} X_i$. The p.g.f. of S_n is, due to independence, when $q = 1 - p$,

$$E\{t^{S_n}\} = E\left\{t^{\sum_{i=1}^{n} X_i}\right\} = \prod_{i=1}^{n} E\{t^{X_i}\} = (pt + q)^n, \quad -\infty < t < \infty.$$

Since all X_i have the same distribution. Binomial expansion yields

$$E\{t^{S_n}\} = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} t^j.$$

Since two polynomials of degree n are equal for **all** t only if their coefficients are equal, we obtain

$$P\{S_n = j\} = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, \ldots, n.$$

The distribution of S_n is called the **binomial distribution**. ■

Example 1.17. In Example 1.13 Part C, the conditional p.d.f. of X_2 given $\{X_1 = x\}$ is

$$f_{X_2|X_1}(y \mid x) = \frac{1}{1+x} I_{(0,1+x)}(y).$$

This is called the **uniform distribution** on $(0, 1+x)$. It is easy to find that

$$E\{Y \mid X = x\} = \frac{1+x}{2}$$

and

$$V\{Y \mid X = x\} = \frac{(1+x)^2}{12}.$$

Since the p.d.f. of X is

$$P\{X = x\} = e^{-\lambda} \frac{\lambda x}{x!}, \quad x = 0, 1, 2, \ldots$$

the law of iterated expectation yields

$$\begin{aligned} E\{Y\} &= E\{E\{Y \mid X\}\} \\ &= E\left\{\frac{1}{2} + \frac{1}{2}X\right\} \\ &= \frac{1}{2} + \frac{\lambda}{2}, \end{aligned}$$

since $E\{X\} = \lambda$.

The law of total variance yields

$$\begin{aligned}
V\{Y\} &= V\{E\{Y \mid X\}\} + E\{V\{Y \mid X\}\} \\
&= V\left\{\frac{1}{2} + \frac{1}{2}X\right\} + E\left\{\frac{(1+X)^2}{12}\right\} \\
&= \frac{1}{4}V\{X\} + \frac{1}{12}E\{1 + 2X + X^2\} \\
&= \frac{1}{4}\lambda + \frac{1}{12}(1 + 2\lambda + \lambda(1+\lambda)) \\
&= \frac{1}{12}(1+\lambda)^2 + \frac{\lambda}{3}.
\end{aligned}$$

To verify these results, prove that $E\{X\} = \lambda$, $V\{X\} = \lambda$ and $E\{X^2\} = \lambda(1+\lambda)$. We also used the result that $V\{a + bX\} = b^2 V\{X\}$. ■

Example 1.18. Let X_1, X_2, X_3 be uncorrelated random variables, having the same variance σ^2, i.e.,

$$\Sigma = \sigma^2 I.$$

Consider the linear transformations

$$\begin{aligned}
Y_1 &= X_1 + X_2, \\
Y_2 &= X_1 + X_3,
\end{aligned}$$

and

$$Y_3 = X_2 + X_3.$$

In matrix notation

$$\mathbf{Y} = A\mathbf{X},$$

where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

The variance–covariance matrix of **Y**, according to (1.8.30) is

$$\begin{aligned}V[\mathbf{Y}] &= A\Sigma A' \\ &= \sigma^2 AA' \\ &= \sigma^2 \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.\end{aligned}$$

From this we obtain that correlations of Y_i, Y_j for $i \neq j$ and $\rho_{ij} = \dfrac{1}{2}$. ■

Example 1.19. We illustrate here convergence in distribution.

A. Let $X_1, X_2, \ldots$ be random variables with distribution functions

$$F_n(x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{1}{n} + \left(1 - \dfrac{1}{n}\right)\left(1 - e^{-x}\right), & \text{if } x \geq 0. \end{cases}$$

$X_n \xrightarrow{d} X$, where the distribution of X is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0. \end{cases}$$

B. X_n are random variables with

$$F_n(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-nx}, & x \geq 0 \end{cases}$$

and $F(x) = I\{x \geq 0\}$. $X_n \xrightarrow{d} X$. Notice that $F(x)$ is discontinuous at $x = 0$. But, for all $x \neq 0$ $\lim_{n\to\infty} F_n(x) = F(x)$.

C. $\mathbf{X}_n$ are random vectors, i.e.,

$$\mathbf{X}_n = (X_{1n}, X_{2n}), \quad n \geq 1.$$

The function $I_x(a, b)$, for $0 < a, b < \infty$, $0 \leq x \leq 1$, is called the incomplete beta function ratio and is given by

$$I_x(a, b) = \frac{\displaystyle\int_0^x u^{a-1}(1 - u)^{b-1} du}{B(a, b)},$$

where $B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1} du$. In terms of these functions, the marginal distribution of X_{1n} and X_{2n} are

$$F_{1n}(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{n} + \left(1 - \frac{1}{n}\right) I_x(a, b), & 0 \le x \le 1 \\ 1, & 1 < x \end{cases}$$

and

$$F_{2n}(y) = \begin{cases} 0, & y < 0 \\ \left(1 - \frac{1}{n}\right) I_y(a, b), & 0 \le y < 1 \\ 1, & 1 \le y \end{cases}$$

where $0 < a, b < \infty$. The joint distribution of (X_{1n}, X_{2n}) is $F_n(x, y) = F_{1n}(x)F_{2n}(y)$, $n \ge 1$. The random vectors $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, where $F(\mathbf{x})$ is

$$F(x, y) = \begin{cases} 0, & x < 0 \text{ or } y < 0 \\ I_x(a, b)I_y(a, b), & 0 \le x, y \le 1 \\ I_x(a, b), & 0 \le x \le 1, y > 1 \\ I_y(a, b), & 1 < x, 0 \le y \le 1 \\ 1, & 1 < x, 1 < y. \end{cases}$$

■

Example 1.20. Convergence in probability.

Let $\mathbf{X}_n = (X_{1n}, X_{2n})$, where $X_{i,n}$ $(i = 1, 2)$ are independent and have a distribution

$$F_n(x) = \begin{cases} 0, & x < 0 \\ nx, & 0 < x < \frac{1}{n} \\ 1, & \frac{1}{n} \le x. \end{cases}$$

Fix an $\epsilon > 0$ and let $N(\epsilon) = \left[\frac{2}{\epsilon}\right]$, then for every $n > N(\epsilon)$,

$$P[(X_{1,n}^2 + X_{2,n}^2)^{1/2} < \epsilon] = 1.$$

Thus, $\mathbf{X}_n \xrightarrow{p} \mathbf{0}$. ■

Example 1.21. Convergence in mean square.

Let $\{X_n\}$ be a sequence of random variables such that

$$E\{X_n\} = 1 + \frac{a}{n}, \quad 0 < a < \infty \text{ and}$$

$$V\{X_n\} = \frac{b}{n}, \qquad 0 < b < \infty.$$

Then, $X_n \xrightarrow{2} 1$, as $n \to \infty$. Indeed, $E\{(X_n - 1)^2\} = \frac{a^2}{n^2} + \frac{b}{n} \to 0$, as $n \to \infty$. ■

Example 1.22. Central Limit Theorem.

A. Let $\{X_n\}$, $n \geq 1$ be a sequence of i.i.d. random variables, $P\{X_n = 1\} = P\{X_n = -1\} = \frac{1}{2}$. Thus, $E\{X_n\} = 0$ and $V\{X_n\} = 1$, $n \geq 1$. Thus $\sqrt{n}\,\bar{X}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{d} N(0, 1)$. It is interesting to note that for these random variables, when $S_n = \sum_{i=1}^{n} X_i$, $\frac{1}{\sqrt{n}} S_n \xrightarrow{d} N(0, 1)$, while $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$.

B. Let $\{X_n\}$ be i.i.d, having a rectangular p.d.f.

$$f(x) = 1_{(0,1)}(x).$$

In this case, $E\{X_1\} = \frac{1}{2}$ and $V\{X_1\} = \frac{1}{12}$. Thus,

$$\sqrt{n}\,\frac{\bar{X}_n - \frac{1}{2}}{\sqrt{\frac{1}{12}}} \xrightarrow[n\to\infty]{d} N(0, 1).$$

Notice that if $n = 12$, then if $S_{12} = \sum_{i=1}^{12} X_i$, then $S_{12} - 6$ might have a distribution close to that of $N(0, 1)$. Early simulation programs were based on this. ■

Example 1.23. Application of Lyapunov's Theorem.

Let $\{X_n\}$ be a sequence of independent random variables, with distribution functions

$$F_n(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp\{-x/n\}, & x \geq 0 \end{cases}$$

$n \geq 1$. Thus, $E\{X_n\} = n$, $V\{X_n\} = n^2$, and $E\{X_n^3\} = 6n^3$. Thus, $B_n^2 = \sum_{k=1}^{n} k^2 = \dfrac{n(n+1)(2n+1)}{6}$, $n \geq 1$. In addition,

$$\sum_{k=1}^{n} E\{X_k^3\} = 6 \sum_{k=1}^{n} k^3 = O(n^4).$$

Thus,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} E\{X_k^3\}}{B_n^3} = 0.$$

It follows from Lyapunov's Theorem that

$$\sqrt{6}\, \frac{\sum_{k=1}^{n} (X_k - k)}{\sqrt{n(n+1)(2n+1)}} \xrightarrow{d} N(0, 1).$$

■

Example 1.24. Variance stabilizing transformation.

Let $\{X_n\}$ be i.i.d. binary random variables, such that $P\{X_n = 1\} = p$, and $P\{X_n = 0\} = 1 - p$. It is easy to verify that $\mu = E\{X_1\} = p$ and $V\{X_1\} = p(1-p)$. Hence, by the CLT, $\sqrt{n}\, \dfrac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{d} N(0, 1)$, as $n \to \infty$. Consider the transformation

$$g(\bar{X}_n) = 2 \sin^{-1} \sqrt{\bar{X}_n}.$$

The derivative of $g(x)$ is

$$g^{(1)}(x) = \frac{2}{\sqrt{1-x}} \cdot \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{x(1-x)}}.$$

Hence $V\{X_1\}(g^{(1)}(p))^2 = 1$.

It follows that

$$\sqrt{n}(2 \sin^{-1}(\sqrt{\bar{X}_n}) - 2 \sin^{-1}(\sqrt{p})) \xrightarrow{d} N(0, 1).$$

$g^{(2)}(x) = -\frac{1}{2} \frac{1-2x}{(x(1-x))^{3/2}}$. Hence, by the delta method,

$$E\{g(\bar{X}_n)\} \cong 2\sin^{-1}(\sqrt{p}) - \frac{1-2p}{4n(p(1-p))^{1/2}}.$$

This approximation is very ineffective if p is close to zero or close to 1. If p is close to $\frac{1}{2}$, the second term on the right-hand side is close to zero. ■

Example 1.25. A. Let $X_1, X_2, \ldots$ be i.i.d. random variables having a finite variance $0 < \sigma^2 < \infty$. Since $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$, we say that $\bar{X}_n - \mu = O_p\left(\frac{1}{\sqrt{n}}\right)$ as $n \to \infty$. Thus, if $c_n \nearrow \infty$ but $c_n = o(\sqrt{n})$, then $c_n(\bar{X}_n - \mu) \xrightarrow{p} 0$. Hence $\bar{X}_n - \mu = o_p(c_n)$, as $n \to \infty$.

B. Let $X_1, X_2, \ldots, X_n$ be i.i.d. having a common exponential distribution with p.d.f.

$$f(x; \mu) = \begin{cases} 0, & \text{if } x < 0 \\ \mu e^{-\mu x}, & \text{if } x \geq 0 \end{cases}$$

$0 < \mu < \infty$. Let $Y_n = \min[X_i, i = 1, \ldots, n]$ be the first order statistic in a random sample of size n (see Section 2.10). The p.d.f. of Y_n is

$$f_n(y; \mu) = \begin{cases} 0, & \text{if } y < 0 \\ n\mu e^{-n\mu y}, & \text{if } y \geq 0. \end{cases}$$

Thus $nY_n \sim X_1$ for all n. Accordingly, $Y_n = O_p\left(\frac{1}{n}\right)$ as $n \to \infty$. It is easy to see that $\sqrt{n}\, Y_n \xrightarrow{p} 0$. Indeed, for any given $\epsilon > 0$,

$$P\{\sqrt{n}\, Y_n > \epsilon\} = e^{-\sqrt{n}\,\mu\epsilon} \to 0 \quad \text{as } n \to \infty.$$

Thus, $Y_n = o_p\left(\frac{1}{\sqrt{n}}\right)$ as $n \to \infty$. ■

PART III: PROBLEMS

Section 1.1

1.1.1 Show that $A \cup B = B \cup A$ and $AB = BA$.

1.1.2 Prove that $A \cup B = A \cup B\bar{A}$, $(A \cup B) - AB = A\bar{B} \cup \bar{A}B$.

1.1.3 Show that if $A \subset B$ then $A \cup B = B$ and $A \cap B = A$.

1.1.4 Prove DeMorgan's laws, i.e., $\overline{A \cup B} = \bar{A} \cap \bar{B}$ or $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

1.1.5 Show that for every $n \geq 2$, $\overline{\left(\bigcup_{i=1}^{n} A_i\right)} = \bigcap_{i=1}^{n} \bar{A}_i$.

1.1.6 Show that if $A_1 \subset \cdots \subset A_N$ then $\sup_{1 \leq n \leq N} A_n = A_N$ and $\inf_{1 \leq n \leq N} A_n = A_1$.

1.1.7 Find $\lim_{n \to \infty} \left[0, 1 - \frac{1}{n}\right)$.

1.1.8 Find $\lim_{n \to \infty} \left(0, \frac{1}{n}\right)$.

1.1.9 Show that if $\mathcal{D} = \{A_1, \ldots, A_k\}$ is a partition of $\mathcal{S}$ then, for every B, $B = \bigcup_{i=1}^{n} A_i B$.

1.1.10 Prove that $\underline{\lim}_{n \to \infty} A_n \subset \overline{\lim}_{n \to \infty} A_n$.

1.1.11 Prove that $\bigcup_{n=1}^{\infty} A_n = \lim_{n \to \infty} \bigcup_{j=1}^{n} A_j$ and $\bigcap_{n=1}^{\infty} A_n = \lim_{n \to \infty} \bigcap_{j=1}^{n} A_j$.

1.1.12 Show that if $\{A_n\}$ is a sequence of pairwise disjoint sets, then $\lim_{n \to \infty} \bigcup_{j=n}^{\infty} A_j = \phi$.

1.1.13 Prove that $\overline{\lim}_{n \to \infty} (A_n \cup B_n) = \overline{\lim}_{n \to \infty} A_n \cup \overline{\lim}_{n \to \infty} B_n$.

1.1.14 Show that if $\{a_n\}$ is a sequence of nonnegative real numbers, then $\sup_{n \geq 1} [0, a_n) = [0, \sup_{n \geq 1} a_n)$.

1.1.15 Let $A \triangle B = A\bar{B} \cup B\bar{A}$ (symmetric difference). Let $\{A_n\}$ be a sequence of disjoint events; define $B_1 = A_1$, $B_{n+1} = B_n \triangle A_{n+1}$, $n \geq 1$. Prove that lim $B_n = \bigcup_{n=1}^{\infty} A_n$.

1.1.16 Verify

(i) $A \triangle B = \bar{A} \triangle \bar{B}$.

(ii) $C = A \triangle B$ if and only if $A = B \triangle C$.

(iii) $\left(\bigcup_{n=1}^{\infty} A_n\right) \triangle \left(\bigcup_{n=1}^{\infty} B_n\right) \subset \bigcup_{n=1}^{\infty} (A_n \triangle B_n).$

1.1.17 Prove that $\overline{\overline{\lim}_{n\to\infty} A_n} = \underline{\lim}_{n\to\infty} \bar{A}_n$.

Section 1.2

1.2.1 Let $\mathcal{A}$ be an algebra over $\mathcal{S}$. Show that if $A_1, A_2 \in \mathcal{A}$ then $A_1 A_2 \in \mathcal{A}$.

1.2.2 Let $\mathcal{S} = \{-, \ldots, -2, -1, 0, 1, 2, \ldots\}$ be the set of all integers. A set $A \subset \mathcal{S}$ is called symmetric if $A = -A$. Prove that the collection $\mathcal{A}$ of all symmetric subsets of $\mathcal{S}$ is an algebra.

1.2.3 Let $\mathcal{S} = \{-, \ldots, -2, -1, 0, 1, 2 \ldots\}$. Let $\mathcal{A}_1$ be the algebra of symmetric subsets of $\mathcal{S}$, and let $\mathcal{A}_2$ be the algebra generated by sets $A_n = \{-2, -1, i_1, \ldots, i_n\}$, $n \geq 1$, where $i_j \geq 0$, $j = 1, \ldots, n$.
(i) Show that $\mathcal{A}_3 = \mathcal{A}_1 \cap \mathcal{A}_2$ is an algebra.
(ii) Show that $\mathcal{A}_4 = \mathcal{A}_1 \cup \mathcal{A}_2$ is **not** an algebra.

1.2.4 Show that if $\mathcal{A}$ is a σ-field, $A_n \subset A_{n+1}$, for all $n \geq 1$, then $\lim_{n\to\infty} \bar{A}_n \in \mathcal{A}$.

Section 1.3

1.3.1 Let $F(x) = P\{(-\infty, x]\}$. Verify
(a) $P\{(a, b]\} = F(b) - F(a)$.
(b) $P\{(a, b)\} = F(b-) - F(a)$.
(c) $P\{[a, b)\} = F(b-) - F(a-)$.

1.3.2 Prove that $P\{A \cup B\} = P\{A\} + P\{B\bar{A}\}$.

1.3.3 A point (X, Y) is chosen in the unit square. Thus, $\mathcal{S} = \{(x, y) : 0 \leq x, y \leq 1\}$. Let $\mathcal{B}$ be the Borel σ-field on $\mathcal{S}$. For a Borel set B, we define

$$P\{B\} = \int\int_B dx dy.$$

Compute the probabilities of

$$B = \left\{(x, y) : x > \frac{1}{2}\right\}$$
$$C = \left\{(x, y) : x^2 + y^2 \leq 1\right\}$$
$$D = \{(x, y) : x + y \leq 1\}$$

$P\{D \cap B\}$, $P\{D \cap C\}$, $P\{C \cap B\}$.

1.3.4 Let $\mathcal{S} = \{x : 0 \le x < \infty\}$ and $\mathcal{B}$ the Borel σ-field on $\mathcal{S}$, generated by the sets $[0, x)$, $0 < x < \infty$. The probability function on $\mathcal{B}$ is $P\{B\} = \lambda \int_B e^{-\lambda x} dx$, for some $0 < \lambda < \infty$. Compute the probabilities

- **(i)** $P\{X \le 1/\lambda\}$.
- **(ii)** $P\left\{\frac{1}{\lambda} \le X \le \frac{2}{\lambda}\right\}$.
- **(iii)** Let $B_n = \left[0, \left(1 + \frac{1}{n}\right)/\lambda\right)$. Compute $\lim_{n\to\infty} P\{B_n\}$ and show that it is equal to $P\left\{\lim_{n\to\infty} B_n\right\}$.

1.3.5 Consider an experiment in which independent trials are conducted sequentially. Let R_i be the result of the ith trial. $P\{R_i = 1\} = p$, $P\{R_i = 0\} = 1 - p$. The trials stop when $(R_1, R_2, \ldots, R_N)$ contains exactly two 1s. Notice that in this case, the number of trials N is random. Describe the sample space. Let w_n be a point of $\mathcal{S}$, which contains exactly n trials. $w_n = \{(i_1, \ldots, i_{n-1}, 1)\}$, $n \ge 2$, where $\sum_{j=1}^{n-1} i_j = 1$. Let $E_n = \{(i_1, \ldots, i_{n-1}, 1) : \sum_{j=1}^{n-1} i_j = 1\}$.

- **(i)** Show that $\mathcal{D} = \{E_2, E_3, \ldots\}$ is a countable partition of $\mathcal{S}$.
- **(ii)** Show that $P\{E_n\} = (n - 1)p^2 q^{n-2}$, where $0 < p < 1$, $q = 1 - p$, and prove that $\sum_{n=2}^{\infty} P\{E_n\} = 1$.
- **(iii)** What is the probability that the experiment will require at least 5 trials?

1.3.6 In a parking lot there are 12 parking spaces. What is the probability that when you arrive, assuming cars fill the spaces at random, there will be four adjacent spaces vacant, while all other spaces filled?

Section 1.4

1.4.1 Show that if A and B are independent, then $\bar{A}$ and $\bar{B}$, A and $\bar{B}$, $\bar{A}$ and B are independent.

1.4.2 Show that if three events are mutually independent, then if we replace any event with its complement, the new collection is still mutually independent.

1.4.3 Two digits are chosen from the set $\mathcal{P} = \{0, 1, \ldots, 9\}$, without replacement. The order of choice is immaterial. The probability function assigns every possible set of two the same probability. Let $A_i (i = 0, \ldots, 9)$ be the event that the chosen set contains the digit i. Show that for any $i \ne j$, A_i and A_j are **not** independent.

1.4.4 Let $A_1, \ldots, A_n$ be mutually independent events. Show that

$$P\left\{\bigcup_{i=1}^{n} A_i\right\} = 1 - \prod_{i=1}^{n} P\{\bar{A}_i\}.$$

1.4.5 If an event A is independent of itself, then $P\{A\} = 0$ or $P(A) = 1$.

1.4.6 Consider the random walk model of Example 1.2.

(i) What is the probability that after n steps the particle will be on a positive integer?

(ii) Compute the probability that after $n = 7$ steps the particle will be at $x = 1$.

(iii) Let p be the probability that in each trial the particle goes one step to the right. Let A_n be the event that the particle returns to the origin after n steps. Compute $P\{A_n\}$ and show, by using the Borel–Cantelli Lemma, that if $p \neq \frac{1}{2}$ then $P\{A_n, i.o.\} = 0$.

1.4.7 Prove that

(i) $\sum_{k=0}^{n} \binom{n}{k} = 2^k$.

(ii) $\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$.

(iii) $\sum_{k=0}^{n} k\binom{M}{k}\binom{N-M}{n-k} = n\frac{M}{N}\binom{N}{n} = M\binom{N-1}{n-1}$.

1.4.8 What is the probability that the birthdays of $n = 12$ randomly chosen people will fall in 12 different calendar months?

1.4.9 A stick is broken at random into three pieces. What is the probability that these pieces can form a triangle?

1.4.10 There are $n = 10$ particles and $m = 5$ cells. Particles are assigned to the cells at random.

(i) What is the probability that each cell contains at least one particle?

(ii) What is the probability that all 10 particles are assigned to the first 3 cells?

Section 1.5

1.5.1 Let F be a discrete distribution concentrated on the jump points $-\infty < \xi_1 < \xi_2 < \cdots < \infty$. Let $p_i = dF(\xi_i)$, $i = 1, 2, \ldots$. Define the function

$$U(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0. \end{cases}$$

(i) Show that, for all $-\infty < x < \infty$

$$F(x) = \sum_{i=1}^{\infty} p_i U(x - \xi_i)$$
$$= \sum_{i=1}^{\infty} p_i I(\xi_i \leq x).$$

(ii) For $h > 0$, define

$$D_h U(x) = \frac{1}{h}[U(x+h) - U(x)] = \frac{1}{h}[I(x \geq -h) - I(x \geq 0)].$$

Show that

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} p_i D_h U(x - \xi_i) dx = 1 \quad \text{for all } h > 0.$$

(iii) Show that for any continuous function $g(x)$, such that $\sum_{i=1}^{\infty} p_i |g(\xi_i)| < \infty$,

$$\lim_{h \to 0} \int_{-\infty}^{\infty} \sum_{i=1}^{\infty} p_i g(x) D_h U(x - \xi_i) dx = \sum_{i=1}^{\infty} p_i g(\xi_i).$$

1.5.2 Let X be a random variable having a discrete distribution, with jump points $\xi_i = i$, and $p_i = dF(\xi_i) = e^{-2}\frac{2^i}{i!}$, $i = 0, 1, 2, \ldots$. Let $Y = X^3$. Determine the p.d.f. of Y.

1.5.3 Let X be a discrete random variable assuming the values $\{1, 2, \ldots, n\}$ with probabilities

$$p_i = \frac{2i}{n(n+1)}, \quad i = 1, \ldots, n.$$

(i) Find $E\{X\}$.

(ii) Let $g(X) = X^2$; find the p.d.f. of $g(X)$.

1.5.4 Consider a discrete random variable X, with jump points on $\{1, 2, \ldots\}$ and p.d.f.

$$f_X(n) = \frac{c}{n^2}, \quad n = 1, 2, \ldots$$

where c is a normalizing constant.

(i) Does $E\{X\}$ exist?

(ii) Does $E\{X/\log X\}$ exist?

1.5.5 Let X be a discrete random variable whose distribution has jump points at $\{x_1, x_2, \ldots, x_k\}$, $1 \leq k \leq \infty$. Assume also that $E\{|X|\} < \infty$. Show that for any linear transformation $Y = \alpha + \beta x$, $\beta \neq 0$, $-\infty < \alpha < \infty$, $E\{Y\} = \alpha + \beta E\{X\}$. (The result is trivially true for $\beta = 0$).

1.5.6 Consider two discrete random variables (X, Y) having a joint p.d.f.

$$f_{XY}(j, n) = \frac{e^{-\lambda}}{j!(n-j)!}\left(\frac{p}{1-p}\right)^j (\lambda(1-p))^n, \quad j = 0, 1, \ldots, n,$$
$$n = 0, 1, 2, \ldots.$$

(i) Find the marginal p.d.f. of X.

(ii) Find the marginal p.d.f. of Y.

(iii) Find the conditional p.d.f. $f_{X|Y}(j \mid n)$, $n = 0, 1, \ldots$.

(iv) Find the conditional p.d.f. $f_{Y|X}(n \mid j)$, $j = 0, 1, \ldots$.

(v) Find $E\{Y \mid X = j\}$, $j = 0, 1, \ldots$.

(vi) Show that $E\{Y\} = E\{E\{Y \mid X\}\}$.

1.5.7 Let X be a discrete random variable, $X \in \{0, 1, 2, \ldots\}$ with p.d.f.

$$f_X(n) = e^{-n} - e^{-(n+1)}, \quad n = 0, 1, \ldots.$$

Consider the partition $\mathcal{D} = \{A_1, A_2, A_3\}$, where

$$A_1 = \{w : X(w) < 2\},$$
$$A_2 = \{w : 2 \leq X(w) < 4\},$$
$$A_3 = \{w : 4 \leq X(w)\}.$$

(i) Find the conditional p.d.f.

$$f_{X|\mathcal{D}}(x \mid A_i), \quad i = 1, 2, 3.$$

(ii) Find the conditional expectations $E\{X \mid A_i\}$, $i = 1, 2, 3$.

(iii) Specify the random variable $E\{X \mid \mathcal{D}\}$.

1.5.8 For a given λ, $0 < \lambda < \infty$, define the function $P(j;\lambda) = e^{-\lambda}\sum_{l=0}^{j}\frac{\lambda^l}{l!}$.

(i) Show that, for a fixed nonnegative integer j, $F_j(x)$ is a distribution function, where

$$F_j(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - P(j-1;x), & \text{if } x \geq 0 \end{cases}$$

and where $P(j;0) = I\{j \geq 0\}$.

(ii) Show that $F_j(x)$ is absolutely continuous and find its p.d.f.

(iii) Find $E\{X\}$ according to $F_j(x)$.

1.5.9 Let X have an absolutely continuous distribution function with p.d.f.

$$f(x) = \begin{cases} 3x^2, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find $E\{e^{-X}\}$.

Section 1.6

1.6.1 Consider the absolutely continuous distribution

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } 1 \leq x \end{cases}$$

of a random variable X. By considering the sequences of simple functions

$$X_n(w) = \sum_{i=1}^{n}\frac{i-1}{n}I\left\{\frac{i-1}{n} \leq X(w) < \frac{i}{n}\right\}, \quad n \geq 1$$

and

$$X_n^2(w) = \sum_{i=1}^{n}\left(\frac{i-1}{n}\right)^2 I\left\{\frac{i-1}{n} \leq X(w) < \frac{i}{n}\right\}, \quad n \geq 1,$$

show that

$$\lim_{n\to\infty} E\{X_n\} = \int_0^1 x\,dx = \frac{1}{2}$$

and

$$\lim_{n\to\infty} E\{X_n^2\} = \int_0^1 x^2 dx = \frac{1}{3}.$$

1.6.2 Let X be a random variable having an absolutely continuous distribution F, such that $F(0) = 0$ and $F(1) = 1$. Let f be the corresponding p.d.f.

(i) Show that the Lebesgue integral

$$\int_0^1 x P\{dx\} = \lim_{n\to\infty} \sum_{i=1}^{2^n} \frac{i-1}{2^n} \left[F\left(\frac{i}{2^n}\right) - F\left(\frac{i-1}{2^n}\right)\right].$$

(ii) If the p.d.f. f is continuous on (0, 1), then

$$\int_0^1 x P\{dx\} = \int_0^1 x f(x) dx,$$

which is the Riemann integral.

1.6.3 Let X, Y be independent identically distributed random variables and let $E\{X\}$ exist. Show that

$$E\{X \mid X+Y\} = E\{Y \mid X+Y\} = \frac{X+Y}{2} \text{ a.s.}$$

1.6.4 Let $X_1, \ldots, X_n$ be i.i.d. random variables and let $E\{X_1\}$ exist. Let $S_n = \sum_{j=1}^{n} X_j$. Then, $E\{X_1 \mid S_n\} = \frac{S_n}{n}$, a.s.

1.6.5 Let

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{4}, & \text{if } x = 0 \\ \frac{1}{4} + \frac{1}{2}x^3, & \text{if } 0 < x < 1 \\ 1, & \text{if } 1 \le x. \end{cases}$$

Find $E\{X\}$ and $E\{X^2\}$.

1.6.6 Let $X_1, \dots, X_n$ be Bernoulli random variables with $P\{X_i = 1\} = p$. If $n = 100$, how large should p be so that $P\{S_n < 100\} < 0.1$, when $S_n = \sum_{i=1}^{n} X_i$?

1.6.7 Prove that if $E\{|X|\} < \infty$, then, for every $A \in \mathcal{F}$,

$$E\{|X| I_A(X)\} \leq E\{|X|\}.$$

1.6.8 Prove that if $E\{|X|\} < \infty$ and $E\{|Y|\} < \infty$, then $E\{X + Y\} = E\{X\} + E\{Y\}$.

1.6.9 Let $\{X_n\}$ be a sequence of i.i.d. random variables with common c.d.f.

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-x}, & \text{if } x \geq 0. \end{cases}$$

Let $S_n = \sum_{i=1}^{n} X_i$.

(i) Use the Borel–Cantelli Lemma to show that $\lim_{n \to \infty} S_n = \infty$ a.s.

(ii) What is $\lim_{n \to \infty} E\left\{\frac{S_n}{1 + S_n}\right\}$?

1.6.10 Consider the distribution function F of Example 1.11, with $\alpha = .9$, $\lambda = .1$, and $\mu = 1$.

(i) Determine the lower quartile, the median, and the upper quartile of $F_{ac}(x)$.

(ii) Tabulate the values of $F_{\rm d}(x)$ for $x = 0, 1, 2, \dots$ and determine the lower quartile, median, and upper quartile of $F_{\rm d}(x)$.

(iii) Determine the values of the median and the interquartile range IQR of $F(x)$.

(iv) Determine $P\{0 < X < 3\}$.

1.6.11 Consider the Cauchy distribution with p.d.f.

$$f(x; \mu, \sigma) = \frac{1}{\pi\sigma} \cdot \frac{1}{1 + (x - \mu)^2/\sigma^2}, \quad -\infty < x < \infty,$$

with $\mu = 10$ and $\sigma = 2$.

(i) Write the formula of the c.d.f. $F(x)$.

(ii) Determine the values of the median and the interquartile range of $F(x)$.

1.6.12 Let X be a random variable having the p.d.f. $f(x) = e^{-x}$, $x \geq 0$. Determine the p.d.f. and the median of

(i) $Y = \log X$,

(ii) $Y = \exp\{-X\}$.

1.6.13 Let X be a random variable having a p.d.f. $f(x) = \frac{1}{\pi}, -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. Determine the p.d.f. and the median of

(i) $Y = \sin X$,

(ii) $Y = \cos X$,

(iii) $Y = \tan X$.

1.6.14 Prove that if $E\{|X|\} < \infty$ then

$$E\{X\} = -\int_{-\infty}^{0} F(x)dx + \int_{0}^{\infty} (1 - F(x))dx.$$

1.6.15 Apply the result of the previous problem to derive the expected value of a random variable X having an exponential distribution, i.e.,

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$

1.6.16 Prove that if $F(x)$ is symmetric around η, i.e.,

$$F(\eta - x) = 1 - F(\eta + x-), \qquad \text{for all } 0 \leq x < \infty,$$

then $E\{X\} = \eta$, provided $E\{|X|\} < \infty$.

Section 1.7

1.7.1 Let (X, Y) be random variables having a joint p.d.f.

$$f_{XY}(x, y) = \begin{cases} 1, & \text{if } -1 < x < 1, 0 < y < 1 - |x| \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the marginal p.d.f. of Y.

(ii) Find the conditional p.d.f. of X given $\{Y = y\}$, $0 < y < 1$.

1.7.2 Consider random variables $\{X, Y\}$. X is a discrete random variable with jump points $\{0, 1, 2, \ldots\}$. The marginal p.d.f. of X is $f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, \ldots$, $0 < \lambda < \infty$. The conditional distribution of Y given $\{X = x\}$, $x \geq 1$, is

$$F_{Y|X}(y \mid x) = \begin{cases} 0, & y < 0 \\ y/x, & 0 \leq y \leq x \\ 1, & x < y. \end{cases}$$

When $\{X = 0\}$

$$F_{Y|X}(y \mid 0) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0. \end{cases}$$

(i) Find $E\{Y\}$.

(ii) Show that the c.d.f. of Y has discontinuity at $y = 0$, and $F_Y(0) - F_Y(0-) = e^{-\lambda}$.

(iii) For each $0 < y < \infty$, $F'_Y(y) = f_Y(y)$, where $\int_0^\infty f_Y(y)dy = 1 - e^{-\lambda}$. Show that, for $y > 0$,

$$f_Y(y) = \sum_{n=1}^{\infty} I\{n - 1 < y < n\} e^{-\lambda} \sum_{x=n}^{\infty} \frac{1}{x} \cdot \frac{\lambda^x}{x!},$$

and prove that $\int_0^\infty f_Y(y)dy = 1 - e^{-\lambda}$.

(iv) Derive the conditional p.d.f. of X given $\{Y = y\}$, $0 < y < \infty$, and find $E\{X \mid Y = y\}$.

1.7.3 Show that if X, Y are independent random variables, $E\{|X|\} < \infty$ and $E\{|Y| < \infty\}$, then $E\{XY\} = E\{X\}E\{Y\}$. More generally, if g, h are integrable, then if X, Y are independent, then

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\}.$$

1.7.4 Show that if X, Y are independent, absolutely continuous, with p.d.f. f_X and f_Y, respectively, then the p.d.f. of $T = X + Y$ is

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t - x)dx.$$

[f_T is the **convolution** of f_X and f_Y.]

Section 1.8

1.8.1 Prove that if $E\{|X|^r\}$ exists, $r \geq 1$, then $\lim_{a\to\infty} (a)^r P\{|X| \geq a\} = 0$.

1.8.2 Let X_1, X_2 be i.i.d. random variables with $E\{X_1^2\} < \infty$. Find the correlation between X_1 and $T = X_1 + X_n$.

1.8.3 Let $X_1, \ldots, X_n$ be i.i.d. random variables; find the correlation between X_1 and the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1.8.4 Let X have an absolutely continuous distribution with p.d.f.

$$f_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{\lambda^m}{(m-1)!} x^{m-1} e^{-\lambda x}, & \text{if } x \geq 0 \end{cases}$$

where $0 < \lambda < \infty$ and m is an integer, $m \geq 2$.

(i) Derive the m.g.f. of X. What is its domain of convergence?

(ii) Show, by differentiating the m.g.f. $M(t)$, that $E\{X^r\} = \dfrac{m(m+1)\cdots(m+r-1)}{\lambda^r}$, $r \geq 1$.

(iii) Obtain the first four central moments of X.

(iv) Find the coefficients of skewness β_1 and kurtosis β_2.

1.8.5 Let X have an absolutely continuous distribution with p.d.f.

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

(i) What is the m.g.f. of X?

(ii) Obtain $E\{X\}$ and $V\{X\}$ by differentiating the m.g.f.

1.8.6 Random variables X_1, X_2, X_3 have the covariance matrix

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Find the variance of $Y = 5x_1 - 2x_2 + 3x_3$.

1.8.7 Random variables $X_1, \ldots, X_n$ have the covariance matrix

$$\Sigma = I + J,$$

where J is an $n \times n$ matrix of 1s. Find the variance of $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

1.8.8 Let X have a p.d.f.

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

Find the characteristic function ϕ of X.

1.8.9 Let $X_1, \ldots, X_n$ be i.i.d., having a common characteristic function ϕ. Find the characteristic function of $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$.

1.8.10 If ϕ is a characteristic function of an absolutely continuous distribution, its p.d.f. is

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt.$$

Show that the p.d.f. corresponding to

$$\phi(t) = \begin{cases} 1 - |t|, & |t| \leq 1 \\ 0, & |t| > 1 \end{cases}$$

is

$$f(x) = \frac{1 - \cos x}{\pi x^2}, \quad |x| \leq \frac{\pi}{2}.$$

1.8.11 Find the m.g.f. of a random variable whose p.d.f. is

$$f_X(x) = \begin{cases} \dfrac{a - |x|}{a^2}, & \text{if } |x| \leq a \\ 0, & \text{if } |x| > a, \end{cases}$$

$0 < a < \infty$.

1.8.12 Prove that if ϕ is a characteristic function, then $|\phi(t)|^2$ is a characteristic function.

1.8.13 Prove that if ϕ is a characteristic function, then

(i) $\lim_{|t|\to\infty} \phi(t) = 0$ if X has an absolutely continuous distribution.

(ii) $\limsup_{|t|\to\infty} |\phi(t)| = 1$ if X is discrete.

1.8.14 Let X be a discrete random variable with p.d.f.

$$f(x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!}, & x = 0, 1, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Find the p.g.f. of X.

Section 1.9

1.9.1 Let F_n, $n \geq 1$, be the c.d.f. of a discrete uniform distribution on $\left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$. Show that $F_n(x) \xrightarrow{d} F(x)$, as $n \to \infty$, where

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x \leq 1 \\ 1, & \text{if } 1 < x. \end{cases}$$

1.9.2 Let $B(j; n, p)$ denote the c.d.f. of the binomial distribution with p.d.f.

$$b(j; n, p) = \binom{n}{j} p^j (1-p)^{n-j}, \quad j = 0, 1, \ldots, n,$$

where $0 < p < 1$. Consider the sequence of binomial distributions

$$F_n(x) = B\left([x]; n, \frac{1}{2n}\right) I\{0 \leq x \leq n\} + I\{x > n\}, \quad n \geq 1.$$

What is the weak limit of $F_n(x)$?

1.9.3 Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random variables such that $V\{X_1\} = \sigma^2 < \infty$, and $\mu = E\{X_1\}$. Use Chebychev's inequality to prove that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$ as $n \to \infty$.

1.9.4 Let $X_1, X_2, \ldots$ be a sequence of binary random variables, such that $P\{X_n = 1\} = \frac{1}{n}$, and $P\{X_n = 0\} = 1 - \frac{1}{n}$, $n \geq 1$.

(i) Show that $X_n \xrightarrow{r} 0$ as $n \to \infty$, for any $r \geq 1$.

(ii) Show from the definition that $X_n \xrightarrow{p} 0$ as $n \to \infty$.

(iii) Show that if $\{X_n\}$ are independent, then $P\{X_n = 1, i.o.\} = 1$. Thus, $X_n \not\to 0$ a.s.

1.9.5 Let $\epsilon_1, \epsilon_2, \ldots$ be independent r.v., such that $E\{\epsilon_n\} = \mu$ and $V\{\epsilon_n\} = \sigma^2$ for all $n \geq 1$. Let $X_1 = \epsilon_1$ and for $n \geq 2$, let $X_n = \beta X_{n-1} + \epsilon_n$, where $-1 < \beta < 1$. Show that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{2} \frac{\mu}{1-\beta}$, as $n \to \infty$.

1.9.6 Prove that convergence in the rth mean, for some $r > 0$ implies convergence in the sth mean, for all $0 < s < r$.

1.9.7 Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. random variables having a common rectangular distribution $R(0, \theta)$, $0 < \theta < \infty$. Let $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Let $\epsilon > 0$. Show that $\sum_{n=1}^{\infty} P_\theta\{X_{(n)} < \theta - \epsilon\} < \infty$. Hence, by the Borel–Cantelli Lemma, $X_{(n)} \xrightarrow{\text{a.s.}} \theta$, as $n \to \infty$. The $R(0, \theta)$ distribution is

$$F_\theta(x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{x}{\theta}, & \text{if } 0 \le x \le \theta \\ 1, & \text{if } \theta < x \end{cases}$$

where $0 < \theta < \infty$.

1.9.8 Show that if $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$, then $P\{w : X(w) \neq Y(w)\} = 0$.

1.9.9 Let $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$, $P\{w : X(w) \neq Y(w)\} = 0$. Then, for every $\epsilon > 0$,

$$P\{|X_n - Y_n| \ge \epsilon\} \to 0, \quad \text{as} \quad n \to \infty.$$

1.9.10 Show that if $X_n \xrightarrow{d} C$ as $n \to \infty$, where C is a constant, then $X_n \xrightarrow{p} C$.

1.9.11 Let $\{X_n\}$ be such that, for any $p > 0$, $\sum_{n=1}^{\infty} E\{|X_n|^p\} < \infty$. Show that $X_n \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

1.9.12 Let $\{X_n\}$ be a sequence of i.i.d. random variables. Show that $E\{|X_1|\} < \infty$ if and only if $\sum_{n=1}^{\infty} P\{|X_1| > \epsilon \cdot n\} < \infty$. Show that $E|X_1| < \infty$ if and only if $\dfrac{X_n}{n} \xrightarrow{\text{a.s.}} 0$.

Section 1.10

1.10.1 Show that if X_n has a p.d.f. f_n and X has a p.d.f. $g(x)$ and if $\int |f_n(x) - g(x)|dx \to 0$ as $n \to \infty$, then $\sup_B |P_n\{B\} - P\{B\}| \to 0$ as $n \to \infty$, for all Borel sets B. (Ferguson, 1996, p. 12).

1.10.2 Show that if $\mathbf{a}'\mathbf{X}_n \xrightarrow{d} \mathbf{a}'\mathbf{X}$ as $n \to \infty$, for all vectors $\mathbf{a}$, then $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ (Ferguson, 1996, p. 18).

1.10.3 Let $\{X_n\}$ be a sequence of i.i.d. random variables. Let $Z_n = \sqrt{n}(\bar{X}_n - \mu)$, $n \geq 1$, where $\mu = E\{X_1\}$ and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $V\{X_1\} < \infty$. Show that $\{Z_n\}$ is tight.

1.10.4 Let $B(n, p)$ designate a discrete random variable, having a binomial distribution with parameter (n, p). Show that $\left\{B\left(n, \frac{1}{2n}\right)\right\}$ is tight.

1.10.5 Let $P(\lambda)$ designate a discrete random variable, which assumes on $\{0, 1, 2, \ldots\}$ the p.d.f. $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, \ldots, 0 < \lambda < \infty$. Using the continuity theorem prove that $B(n, p_n) \xrightarrow{d} P(\lambda)$ if $\lim_{n\to\infty} np_n = \lambda$.

1.10.6 Let $X_n \sim B\left(n, \frac{1}{2n}\right)$, $n \geq 1$. Compute $\lim_{n\to\infty} E\{e^{-X_n}\}$.

Section 1.11

1.11.1 (Khinchin WLLN). Use the continuity theorem to prove that if $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. random variables, then $\bar{X}_n \xrightarrow{p} \mu$, where $\mu = E\{X_1\}$.

1.11.2 (Markov WLLN). Prove that if $X_1, X_2, \ldots, X_n, \ldots$ are independent random variables and if $\mu_k = E\{X_k\}$ exists, for all $k \geq 1$, and $E|X_k - \mu_k|^{1+\delta} < \infty$ for some $\delta > 0$, all $k \geq 1$, then $\frac{1}{n^{1+\delta}}\sum_{k=1}^{n} E|X_k - \mu_k|^{1+\delta} \to 0$ as $n \to \infty$ implies that $\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k) \xrightarrow{p} 0$ as $n \to \infty$.

1.11.3 Let $\{\mathbf{X}_n\}$ be a sequence of random vectors. Prove that if $\bar{\mathbf{X}}_n \xrightarrow{d} \boldsymbol{\mu}$ then $\bar{\mathbf{X}}_n \xrightarrow{p} \boldsymbol{\mu}$, where $\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$ and $\boldsymbol{\mu} = E\{\mathbf{X}_1\}$.

1.11.4 Let $\{X_n\}$ be a sequence of i.i.d. random variables having a common p.d.f.

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{\lambda^m}{(m-1)!}x^{m-1}e^{-\lambda x}, & \text{if } x \geq 0, \end{cases}$$

where $0 < \lambda < \infty, m = 1, 2, \ldots$. Use Cantelli's Theorem (Theorem 1.11.1) to prove that $\bar{X}_n \xrightarrow{\text{a.s.}} \frac{m}{\lambda}$, as $n \to \infty$.

1.11.5 Let $\{X_n\}$ be a sequence of independent random variables where

$$X_n \sim R(-n, n)/n$$

and $R(-n, n)$ is a random variable having a uniform distribution on $(-n, n)$, i.e.,

$$f_n(x) = \frac{1}{2n} 1_{(-n,n)}(x).$$

Show that $\bar{X}_n \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. [Prove that condition (1.11.6) holds].

1.11.6 Let $\{X_n\}$ be a sequence of i.i.d. random variables, such that $|X_n| \le C$ a.s., for all $n \ge 1$. Show that $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$ as $n \to \infty$, where $\mu = E\{X_1\}$.

1.11.7 Let $\{X_n\}$ be a sequence of independent random variables, such that

$$P\{X_n = \pm 1\} = \frac{1}{2}\left(1 - \frac{1}{2^n}\right)$$

and

$$P\{X_n = \pm n\} = \frac{1}{2} \cdot \frac{1}{2^n}, \quad n \ge 1.$$

Prove that $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$.

Section 1.12

1.12.1 Let $X \sim P(\lambda)$, i.e.,

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$

Apply the continuity theorem to show that

$$\frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow{d} N(0, 1), \text{ as } \lambda \to \infty.$$

1.12.2 Let $\{X_n\}$ be a sequence of i.i.d. discrete random variables, and $X_1 \sim P(\lambda)$. Show that

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} \xrightarrow{d} N(0, 1), \text{ as } n \to \infty.$$

What is the relation between problems 1 and 2?

1.12.3 Let $\{X_n\}$ be i.i.d., binary random variables, $P\{X_n = 1\} = P\{X_n = 0\} = \frac{1}{2}$, $n \geq 1$. Show that

$$\frac{\sum_{i=1}^{n} i X_i - \frac{n(n+1)}{4}}{B_n} \xrightarrow{d} N(0, 1), \quad \text{as } n \to \infty,$$

where $B_n^2 = \frac{n(n+1)(2n+1)}{24}$, $n \geq 1$.

1.12.4 Consider a sequence $\{X_n\}$ of independent discrete random variables, $P\{X_n = n\} = P\{X_n = -n\} = \frac{1}{2}$, $n \geq 1$. Show that this sequence satisfies the CLT, in the sense that

$$\frac{\sqrt{6}\, S_n}{\sqrt{n(n+1)(2n+1)}} \xrightarrow{d} N(0, 1), \quad \text{as } n \to \infty.$$

1.12.5 Let $\{X_n\}$ be a sequence of i.i.d. random variables, having a common absolutely continuous distribution with p.d.f.

$$f(x) = \begin{cases} \dfrac{1}{2|x| \log^2 |x|}, & \text{if } |x| < \dfrac{1}{e} \\ 0, & \text{if } |x| \geq \dfrac{1}{e}. \end{cases}$$

Show that this sequence satisfies the CLT, i.e.,

$$\sqrt{n}\, \frac{\bar{X}_n}{\sigma} \xrightarrow{d} N(0, 1), \quad \text{as } n \to \infty,$$

where $\sigma^2 = V\{X\}$.

1.12.6 (**i**) Show that

$$\frac{(G(1, n) - n)}{\sqrt{n}} \xrightarrow{d} N(0, 1), \quad \text{as } n \to \infty$$

where $G(1, n)$ is an absolutely continuous random variable with a p.d.f.

$$g_n(x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{1}{(n-1)!} x^{n-1} e^{-x}, & x \geq 0. \end{cases}$$

(**ii**) Show that, for large n,

$$g_n(n) = \frac{1}{(n-1)!} n^{n-1} e^{-n} \approx \frac{1}{\sqrt{2\pi}\,\sqrt{n}}.$$

Or

$$n! \approx \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n} \text{ as } n \to \infty.$$

This is the famous **Stirling approximation**.

Section 1.13

1.13.1 Let $X_n \sim R(-n, n)$, $n \geq 1$. Is the sequence $\{X_n\}$ uniformly integrable?

1.13.2 Let $Z_n = \dfrac{X_n - n}{\sqrt{n}} \sim N(0, 1)$, $n \geq 1$. Show that $\{Z_n\}$ is uniformly integrable.

1.13.3 Let $\{X_1, X_2, \ldots, X_n, \ldots\}$ and $\{Y_1, Y_2, \ldots, Y_n, \ldots\}$ be two independent sequences of i.i.d. random variables. Assume that $0 < V\{X_1\} = \sigma_x^2 < \infty$, $0 < V\{Y_1\} = \sigma_y^2 < \infty$. Let $f(x, y)$ be a continuous function on R^2, having continuous partial derivatives. Find the limiting distribution of $\sqrt{n}(f(\bar{X}_n, \bar{Y}_n) - f(\xi, \eta))$, where $\xi = E\{X_1\}$, $\eta = E\{Y_1\}$. In particular, find the limiting distribution of $R_n = \bar{X}_n/\bar{Y}_n$, when $\eta > 0$.

1.13.4 We say that $X \sim E(\mu)$, $0 < \mu < \infty$, if its p.d.f. is

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ \mu e^{-\mu x}, & \text{if } x \geq 0. \end{cases}$$

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of i.i.d. random variables, $X_1 \sim E(\mu)$, $0 < \mu < \infty$. Let $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Compute $V\{e^{\bar{X}_n}\}$ exactly.
(b) Approximate $V\{e^{\bar{X}_n}\}$ by the delta method.

1.13.5 Let $\{X_n\}$ be i.i.d. Bernoulli random variables, i.e., $X_1 \sim B(1, p)$, $0 < p < 1$. Let $\hat{p}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ and

$$W_n = \log\left(\frac{\hat{p}_n}{1 - \hat{p}_n}\right).$$

Use the delta method to find an approximation, for large values of n, of

(i) $E\{W_n\}$

(ii) $V\{W_n\}$.

Find the asymptotic distribution of $\sqrt{n}\left(W_n - \log\left(\frac{p}{1-p}\right)\right)$.

1.13.6 Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having a common continuous distribution function $F(x)$. Let $F_n(x)$ be the empirical distribution function. Fix a value x_0 such that $0 < F_n(x_0) < 1$.

(i) Show that $nF_n(x_0) \sim B(n, F(x_0))$.

(ii) What is the asymptotic distribution of $F_n(x_0)$ as $n \to \infty$?

1.13.7 Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables having a standard Cauchy distribution. What is the asymptotic distribution of the sample median $F_n^{-1}\left(\frac{1}{2}\right)$?

PART IV: SOLUTIONS TO SELECTED PROBLEMS

1.1.5 For $n = 2$, $\overline{A_1 \cup A_2} = \bar{A}_1 \cap \bar{A}_2$. By induction on n, assume that $\overline{\bigcup_{i=1}^{k} A_i} = \bigcap_{i=1}^{k} \bar{A}_i$ for all $k = 2, \ldots, n$. For $k = n + 1$,

$$\overline{\bigcup_{k=1}^{n+1} A_i} = \overline{\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}} = \overline{\left(\bigcup_{i=1}^{n} A_i\right)} \cap \bar{A}_{n+1}$$
$$= \bigcap_{i=1}^{n} \bar{A}_i \cap \bar{A}_{n+1} = \bigcap_{i=1}^{n+1} \bar{A}_i.$$

1.1.10 We have to prove that $\left(\underline{\lim}_{n\to\infty} A_n\right) \subset \left(\overline{\lim}_{n\to\infty} A_n\right)$. For an elementary event $w \in \mathcal{S}$, let

$$I_{A_n}(w) = \begin{cases} 1, & \text{if } w \in A_n \\ 0, & \text{if } w \notin A_n. \end{cases}$$

Thus, if $w \in \underline{\lim}_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n$, there exists an integer $K(w)$ such that

$$\prod_{n \geq K(w)}^{\infty} I_{A_n}(w) = 1.$$

Accordingly, for all $n \geq 1$, $w \in \bigcup_{k=n}^{\infty} A_k$. Here $w \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \overline{\lim}_{n\to\infty} A_n$.

1.1.15 Let $\{A_n\}$ be a sequence of disjoint events. For all $n \geq 1$, we define

$$\begin{aligned} B_n &= B_{n-1} \Delta A_n \\ &= B_{n-1}\bar{A}_n \cup \bar{B}_{n-1} A_n \end{aligned}$$

and

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= \bar{A}_1 A_2 \cup A_1 \bar{A}_2 \\ B_3 &= \overline{(\bar{A}_1 A_2 \cup A_1 \bar{A}_2)} A_3 \cup (\bar{A}_1 A_2 \cup A_1 \bar{A}_2)\bar{A}_3 \\ &= (\overline{\bar{A}_1 A_2} \cap \overline{A_1 \bar{A}_2}) A_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 \bar{A}_3 \\ &= (A_1 \cup \bar{A}_2)(\bar{A}_1 \cup A_2) A_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 \bar{A}_3 \\ &= A_1 A_2 A_3 \cup \bar{A}_1 \bar{A}_2 A_3 \cup \bar{A}_1 A_2 \bar{A}_3 \cup A_1 \bar{A}_2 \bar{A}_3. \end{aligned}$$

By induction on n we prove that, for all $n \geq 2$,

$$B_n = \left(\bigcap_{j=1}^{n} A_j \right) \cup \left(\bigcup_{i=1}^{n} A_i \left(\bigcap_{j \neq i} \bar{A}_j \right) \right) = \bigcup_{i=1}^{n} A_i.$$

Hence $B_n \subset B_{n+1}$ for all $n \geq 1$ and $\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} A_n$.

1.2.2 The sample space $\mathcal{S} = \mathbb{Z}$, the set of all integers. A is a symmetric set in $\mathcal{S}$, if $A = -A$. Let $\mathcal{A} = \{\text{collection of all symmetric sets}\}$. $\phi \in \mathcal{A}$. If $A \in \mathcal{A}$ then $\bar{A} \in \mathcal{A}$. Indeed $-\bar{A} = -\mathcal{S} - (-A) = \mathcal{S} - A = \bar{A}$. Thus, $\bar{A} \in \mathcal{A}$. Moreover, if $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$. Thus, $\mathcal{A}$ is an algebra.

1.2.3 $\mathcal{S} = \mathbb{Z}$. Let $\mathcal{A}_1 = \{\text{generated by symmetric sets}\}$. $\mathcal{A}_2 = \{$generated by $(-2, -1, i_1, \ldots, i_n)$, $n \geq 1$, $i_j \in \mathbb{N}\ \forall j = 1, \ldots, n\}$. Notice that if $A = (-2, -1, i_1, \ldots, i_n)$ then $\bar{A} = \{(\cdots, -4, -3, \mathbb{N} - (i_1, \ldots, i_n))\} \in \mathcal{A}_2$, and $\mathcal{S} = A \cup \bar{A}_1 \in \mathcal{A}_2$. $\mathcal{A}_2$ is an algebra. $\mathcal{A}_3 = \mathcal{A}_1 \cap \mathcal{A}_2$. If $B \in \mathcal{A}_3$ it must be symmetric and also $B \in \mathcal{A}_2$. Thus, $B = (-2, -1, 1, 2)$ or $B = (\cdots, -4, -3, 3, 4, \ldots)$. Thus, B and $\bar{B}$ are in $\mathcal{A}_3$, so $\mathcal{S} = (B \cup \bar{B}) \in \mathcal{A}_3$ and so is ϕ. Thus, $\mathcal{A}$ is an algebra.

Let $\mathcal{A}_4 = \mathcal{A}_1 \cup \mathcal{A}_2$. Let $A = \{-2, -1, 3, 7\}$ and $B = \{-3, 3\}$. Then $A \cup B = \{-3, -2, -1, 3; 7\}$. But $A \cup B$ does not belong to $\mathcal{A}_1$ neither to $\mathcal{A}_2$. Thus $A \cup B \notin \mathcal{A}_4$. $\mathcal{A}_4$ is not an algebra.

1.3.5 The sample space is

$$\mathcal{S} = \{(i_1, \ldots, i_{n-1}, 1) : \sum_{j=1}^{n-1} i_j = 1, \quad n \geq 2\}.$$

(i) Let $E_n = \left\{(i_1, \ldots, i_{n-1}, 1) : \sum_{j=1}^{n-1} i_j = 1\right\}$, $n = 2, 3, \ldots$. For $j \neq k$, $E_j \cap E_k = \emptyset$. Also $\bigcup_{n=l}^{\infty} E_n = \mathcal{S}$. Thus, $\mathcal{D} = \{E_2, E_3, \ldots\}$ is a countable partition of $\mathcal{S}$.

(ii) All elementary events $w_n = (i_1, \ldots, i_{n-1}, 1) \in E_n$ are equally probable and $P\{w_n\} = p^2 q^{n-2}$. There are $\binom{n-1}{1} = n - 1$ such elementary events in E_n. Thus, $P\{E_n\} = (n-1)p^2 q^{n-2}$. Moreover,

$$\begin{aligned}\sum_{n=2}^{\infty} P\{E_n\} &= p^2 \sum_{n=2}^{\infty} (n-1) q^{n-2} \\ &= p^2 \sum_{l=1}^{\infty} l q^{k-1} = 1.\end{aligned}$$

Indeed,

$$\begin{aligned}\sum_{l=1}^{\infty} l q^{l-1} &= \sum_{l=1}^{\infty} \frac{d}{dq} q^l \\ &= \frac{d}{dq}\left(\frac{q}{1-q}\right) \\ &= \frac{1}{(1-q)^2} = \frac{1}{p^2}.\end{aligned}$$

(iii) The probability that the experiment requires at least 5 trials is the probability that in the first 4 trials there is at most 1 success, which is $1 - p^2(1 + 2q + 3q^2)$.

1.4.6 Let X_n denote the position of the particle after n steps.

(i) If $n = 2k$, the particle after n steps could be, on the positive side only on even integers $2, 4, 6, \ldots, 2k$. If $n = 2k + 1$, the particle could be after n steps on the positive side only on an odd integer $1, 3, 5, \ldots, 2k + 1$.

Let p be the probability of step to the right ($0 < p < 1$) and $q = 1 - p$ of step to the left. If $n = 2k + 1$,

$$P\{X_n = 2j + 1\} = \binom{2j}{j} p^{2k+1-j} q^j, \quad j = 0, \ldots, k.$$

Thus, if $n = 2k + 1$,

$$P\{X_n > 0\} = \sum_{j=0}^{k} \binom{2j}{j} p^{2k+1-j} q^j.$$

In this solution, we assumed that all steps are independent (see Section 1.7). If $n = 2k$ the formula can be obtained in a similar manner.

(ii) $P\{X_7 = 1\} = \binom{6}{3} p^4 q^3$. If $p = \frac{1}{2}$, then $P\{X_7 = 1\} = \frac{\binom{6}{3}}{2^7} = 0.15625$.

(iii) The probability of returning to the origin after n steps is

$$P\{X_n = 0\} = \begin{cases} 0, & \text{if } n = 2k + 1 \\ \binom{2k}{k} p^k q^k, & \text{if } n = 2k. \end{cases}$$

Let $A_n = \{X_n = 0\}$. Then, $\sum_{k=0}^{\infty} P\{A_{2k+1}\} = 0$ and when $p = \frac{1}{2}$,

$$\sum_{k=0}^{\infty} P\{A_{2k}\} = \sum_{k=0}^{\infty} \binom{2k}{k} \frac{1}{2^{2k}} = \sum_{k=0}^{\infty} \frac{(2k)!}{(k!)^2} \cdot \frac{1}{4^k} = \infty.$$

Thus, by the Borel–Cantelli Lemma, if $p = \frac{1}{2}$, $P\{A_n i.o.\} = 1$. On the other hand, if $p \neq \frac{1}{2}$,

$$\sum_{k=0}^{\infty} \binom{2k}{k} (pq)^k = \frac{4pq}{\sqrt{1 - 4pq}\left(1 + \sqrt{1 - 4pq}\right)} < \infty.$$

Thus, if $p \neq \frac{1}{2}$, $P\{A_n i.o.\} = 0$.

1.5.1 $F(x)$ is a discrete distribution with jump points at $-\infty < \xi_1 < \xi_2 < \cdots < \infty$. $p_i = d(F\xi_i)$, $i = 1, 2, \ldots$. $U(x) = I(x \geq 0)$.

(i)

$$U(x - \xi_i) = I(x \geq \xi_i)$$

$$F(x) = \sum_{\xi_i \leq x} p_i = \sum_{i=1}^{\infty} p_i U(x - \xi_i).$$

(ii) For $h > 0$,

$$D_h U(x) = \frac{1}{h}[U(x+h) - U(x)].$$

$U(x+h) = 1$ if $x \geq -h$. Thus,

$$D_h U(x) = \frac{1}{h} I(-h \leq x < 0)$$

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} p_i D_h U(x - \xi_i) dx = \sum_{i=1}^{\infty} p_i \frac{1}{h} \int_{-h+x-\xi_i}^{x-\xi_i} du = \sum_{i=1}^{\infty} p_i = 1.$$

(iii)

$$\lim_{h \to 0} \int_{-\infty}^{\infty} \sum_{i=1}^{\infty} p_i g(x) D_h U(x - \xi_i) dx = \sum_{i=1}^{\infty} p_i \lim_{h \downarrow 0} \int_{\xi_i - h}^{\xi_i} \frac{g(x)}{h} dx$$

$$= \sum_{i=1}^{\infty} p_i \lim_{h \downarrow 0} \frac{G(\xi_i) - G(\xi_i - h)}{h} = \sum_{i=1}^{\infty} p_i g(\xi_i)$$

Here, $G(\xi_i) = \int_{-\infty}^{\xi_i} g(x) dx$; $\frac{d}{d\xi_i} G(\xi_i) = g(\xi_i)$.

1.5.6 The joint p.d.f. of two discrete random variables is

$$f_{X,Y}(j, n) = \frac{e^{-\lambda}}{j!(n-j)!} \left(\frac{p}{1-p}\right)^j (\lambda(1-p))^n, \; j = 0, \ldots, n \; n = 0, 1, \ldots.$$

(i) The marginal distribution of X is

$$\begin{aligned} f_X(j) &= \sum_{n=j}^{\infty} f_{X,Y}(j, n) \\ &= \frac{p^j (\lambda(1-p))^j}{(1-p)^j j!} e^{-\lambda} \sum_{n=j}^{\infty} \frac{(\lambda(1-p))^{n-j}}{(n-j)!} \\ &= e^{-\lambda p} \frac{(\lambda p)^j}{j!} e^{-\lambda(1-p)} \sum_{n=0}^{\infty} \frac{(\lambda(1-p))^n}{n!} \\ &= e^{-\lambda p} \frac{(\lambda p)^j}{j!}, \quad j = 0, 1, 2, \ldots. \end{aligned}$$

(ii) The marginal p.d.f. of Y is

$$
\begin{aligned}
f_Y(n) &= \sum_{j=0}^{n} p_{X,Y}(j,n) \\
&= e^{-\lambda}\frac{\lambda^n}{n!}\sum_{j=0}^{n}\binom{n}{j}p^j(1-p)^{n-j} \\
&= e^{-\lambda}\frac{\lambda^n}{n!}, \quad n = 0, 1, \ldots .
\end{aligned}
$$

(iii) $$p_{X|Y}(j \mid n) = \frac{f_{X,Y}(j,n)}{f_Y(n)} = \binom{n}{j}p^j(1-p)^{n-j}, \quad j = 0, \ldots, n.$$

(iv) $$
\begin{aligned}
p_{Y|X}(n \mid j) &= \frac{f_{X,Y}(j,n)}{f_X(j)} \\
&= e^{-\lambda(1-p)}\frac{(\lambda(1-p))^{n-j}}{(n-j)!}, \quad n \geq j.
\end{aligned}
$$

(v) $$E(Y \mid X = j) = j + \lambda(1-p).$$

(vi) $$
\begin{aligned}
E\{Y\} &= E\{E\{Y \mid X\}\} = \lambda(1-p) + E\{X\} \\
&= \lambda(1-p) + \lambda p = \lambda.
\end{aligned}
$$

1.5.8 $$F_j(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - P(j-1; x), & \text{if } x \geq 0, \end{cases}$$

where $j \geq 1$, and $P(j-1; x) = e^{-x}\sum_{i=0}^{j-1}\frac{x^i}{x!}$.

(i) We have to show that, for each $j \geq 1$, $F_j(x)$ is a c.d.f.

(i) $0 \leq F_j(x) \leq 1$ for all $0 \leq x < \infty$.

(ii) $F_j(0) = 0$ and $\lim_{x\to\infty} F_j(x) = 1$.

(iii) We show now that $F_j(x)$ is strictly increasing in x. Indeed, for all $x > 0$

$$
\begin{aligned}
\frac{d}{dx}F_j(x) &= -\sum_{i=0}^{j-1}\frac{d}{dx}\left(e^{-x}\frac{x^i}{i!}\right) \\
&= e^{-x} + \sum_{i=1}^{j-1}\left(e^{-x}\frac{x^i}{i!} - e^{-x}\frac{x^{i-1}}{(i-1)!}\right) \\
&= e^{-x}\frac{x^{j-1}}{(j-1)!} > 0, \quad \text{for all } 0 < x < \infty.
\end{aligned}
$$

(ii) The density of $F_j(x)$ is

$$f_j(x) = \frac{x^{j-1}}{(j-1)!}e^{-x}, \quad j \geq 1, \quad x \geq 0.$$

$F_j(x)$ is absolutely continuous.

(iii)

$$E_j\{X\} = \int_0^\infty \frac{x^j}{(j-1)!}e^{-x}dx$$
$$= j\int_0^\infty \frac{x^j}{j!}e^{-x}dx = j.$$

1.6.3 X, Y are independent and identically distributed, $E|X| < \infty$.

$$E\{X \mid X+Y\} + E\{Y \mid X+Y\} = X+Y$$
$$E\{X \mid X+Y\} = E\{Y \mid X+Y\} = \frac{X+Y}{2}.$$

1.6.9

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0. \end{cases}$$

(i) Let $A_n = \{X_n > 1\}$. The events A_n, $n \geq 1$, are independent. Also $P\{A_n\} = e^{-1}$. Hence, $\sum_{n=1}^{\infty} P\{A_n\} = \infty$. Thus, by the Borel–Cantelli Lemma, $P\{A_n i.o.\} = 1$. That is, $P\left(\lim_{n\to\infty} S_n = \infty\right) = 1$.

(ii) $\frac{S_n}{1+S_n} \geq 0$. This random variable is bounded by 1. Thus, by the Dominated Convergence Theorem, $\lim_{n\to\infty} E\left\{\frac{S_n}{1+S_n}\right\} = E\left\{\lim_{n\to\infty}\frac{S_n}{1+S_n}\right\} = 1$.

1.8.4

$$f_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{\lambda^m}{(m-1)!}x^{m-1}e^{-\lambda x}, & x \geq 0; m \geq 2. \end{cases}$$

(i) The m.g.f. of X is

$$M(t) = \frac{\lambda^m}{(m-1)!}\int_0^\infty e^{-x(\lambda-t)}x^{m-1}dx$$
$$= \frac{\lambda^m}{(\lambda-t)^m} = \left(1 - \frac{t}{\lambda}\right)^{-m}, \quad \text{for } t < \lambda.$$

The domain of convergence is $(-\infty, \lambda)$.

(ii)
$$M'(t) = \frac{m}{\lambda}\left(1 - \frac{t}{\lambda}\right)^{-m-1}$$
$$M''(t) = \frac{m(m+1)}{\lambda^2}\left(1 - \frac{t}{\lambda}\right)^{-(m+2)}$$
$$\vdots$$
$$M^{(r)}(t) = \frac{m(m+1)\cdots(m+r-1)}{\lambda^r}\left(1 - \frac{t}{\lambda}\right)^{-(m+r)}.$$

Thus, $\mu_r = M^{(r)}(t)\Big|_{t=0} = \dfrac{m(m+1)\cdots(m+r-1)}{\lambda^r}\ r \geq 1.$

(iii)
$$\mu_1 = \frac{m}{\lambda}$$
$$\mu_2 = \frac{m(m+1)}{\lambda^2}$$
$$\mu_3 = \frac{m(m+1)(m+2)}{\lambda^3}$$
$$\mu_4 = \frac{m(m+1)(m+2)(m+3)}{\lambda^4}.$$

The central moments are

$$\mu_1^* = 0$$
$$\mu_2^* = \mu_2 - \mu_1^2 = \frac{m}{\lambda^2}$$
$$\begin{aligned}\mu_3^* &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \\ &= \frac{1}{\lambda^3}(m(m+1)(m+2) - 3m^2(m+1) + 2m^3) \\ &= \frac{2m}{\lambda^3}\end{aligned}$$
$$\begin{aligned}\mu_4^* &= \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4 \\ &= \frac{1}{\lambda^4}(m(m+1)(m+2)(m+3) - 4m^2(m+1)(m+2) \\ &\quad + 6m^3(m+1) - 3m^4) \\ &= \frac{3m(m+2)}{\lambda^4}.\end{aligned}$$

(iv)
$$\beta_1 = \frac{2m}{m^{3/2}} = \frac{2}{\sqrt{m}};\quad \beta_2 = \frac{3m(m+2)}{m^2} = 3 + \frac{6}{m}.$$

1.8.11 The m.g.f. is

$$\begin{aligned} M_X(t) &= \frac{1}{a^2}\int_{-a}^{a} e^{tx}(a-|x|)dx \\ &= \frac{2(\cosh(at)-1)}{a^2t^2} \\ &= 1+\frac{1}{12}(at)^2+o(t), \quad \text{as } t \to 0. \end{aligned}$$

1.9.1

$$F_n(x) = \begin{cases} 0, & x < 0 \\ \dfrac{j}{n}, & \dfrac{j}{n} \le x < \dfrac{j+1}{n}, \ j = 0, \dots, n-1 \\ 1, & 1 \le x. \end{cases}$$

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 < x < 1 \\ 1, & 1 \le x. \end{cases}$$

All points $-\infty < x < \infty$ are continuity points of $F(x)$. $\lim\limits_{n\to\infty} F_n(x) = F(x)$, for all $x < 0$ or $x > 1$. $|F_n(x) - F(x)| \le \dfrac{1}{n}$ for all $0 \le x \le 1$. Thus $F_n(x) \xrightarrow{w} F(x)$, as $n \to \infty$.

1.9.4

$$X_n = \begin{cases} 0, & \text{w.p. } \left(1-\dfrac{1}{n}\right) \\ 1, & \text{w.p. } \dfrac{1}{n} \end{cases}, n \ge 1.$$

(i) $E\{|X_n|^r\} = \dfrac{1}{n}1 = \dfrac{1}{n}$ for all $r > 0$. Thus, $X_n \underset{n\to\infty}{\xrightarrow{r}} 0$, for all $r > 0$.

(ii) $P\{|X_n| > \epsilon\} = \dfrac{1}{n}$ for all $n \ge 1$, any $\epsilon > 0$. Thus, $X_n \underset{n\to\infty}{\xrightarrow{p}} 0$.

(iii) Let $A_n = \{w : X_n(w) = 1\}$; $P\{A_n\} = \dfrac{1}{n}$, $n \ge 0$. $\sum_{n=1}^{\infty}\dfrac{1}{n} = \infty$. Since $X_1, X_2, \dots$ are independent, by Borel–Cantelli's Lemma, $P\{X_n = 1, i.o.\} = 1$. Thus $X_n \not\to 0$ a.s.

1.9.5 $\epsilon_1, \epsilon_2, \cdots$ independent r.v.s, such that $E(\epsilon_n) = \mu$, and $V\{\epsilon_n\} = \sigma^2$. $\forall n \ge 1$.

$$\begin{aligned} X_1 &= \epsilon_1, \\ X_n &= \beta X_{n-1} + \epsilon_n = \beta(\beta X_{n-2} + \epsilon_{n-1}) + \epsilon_n \\ &= \cdots = \sum_{j=1}^{n} \beta^{n-j}\epsilon_j, \quad \forall n \ge 1, \quad |\beta| < 1. \end{aligned}$$

Thus, $E\{X_n\} = \mu \sum_{j=0}^{n-1} \beta^j \underset{n\to\infty}{\longrightarrow} \frac{\mu}{1-\beta}$.

$$\begin{aligned}\bar{X}_n &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{i}\beta^{i-j}\epsilon_j \\ &= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\sum_{i=j}^{n}\beta^{i-j} \\ &= \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\frac{1-\beta^{n-j+1}}{1-\beta}.\end{aligned}$$

Since $\{\epsilon_n\}$ are independent,

$$\begin{aligned}V\{\bar{X}_n\} &= \frac{\sigma^2}{n^2(1-\beta)^2}\sum_{j=1}^{n}(1-\beta^{n-j+1})^2 \\ &= \frac{\sigma^2}{n(1-\beta)^2}\left(1 - 2\frac{\beta(1-\beta^{n+1})}{n(1-\beta)} + \frac{\beta^2(1-\beta^{2n+1})}{n(1-\beta^2)}\right) \to 0 \text{ as } n\to\infty.\end{aligned}$$

Furthermore,

$$E\left\{\left(\bar{X}_n - \frac{\mu}{1-\beta}\right)^2\right\} = V\{\bar{X}_n\} + \left(E\{\bar{X}_n\} - \frac{\mu}{1-\beta}\right)^2$$

$$\left(E\{\bar{X}_n\} - \frac{\mu}{1-\beta}\right)^2 = \frac{\mu^2}{n^2(1-\beta)^2}(1-\beta^{n+1})^2 \to 0 \text{ as } n\to\infty.$$

Hence, $\bar{X}_n \xrightarrow{2} \frac{\mu}{1-\beta}$.

1.9.7 $X_1, X_2, \ldots$ i.i.d. distributed like $R(0,\theta)$. $X_n = \max_{1\le i\le n}(X_i)$. Due to independence,

$$F_n(x) = P\{X_{(n)} \le x\} = \begin{cases} 0, & x < 0 \\ \left(\frac{x}{\theta}\right)^n, & 0 \le x < \theta \\ 1, & \theta \le x. \end{cases}$$

Accordingly, $P\{X_{(n)} < \theta - \epsilon\} = \left(1 - \frac{\epsilon}{\theta}\right)^n$, $0 < \epsilon < \theta$. Thus, $\sum_{n=1}^{\infty} P\{X_{(n)} \le \theta - \epsilon\} < \infty$, and $P\{X_{(n)} \le \theta - \epsilon, i.o.\} = 0$. Hence, $X_{(n)} \to \theta$ a.s.

1.10.2 We are given that $\mathbf{a}'\mathbf{X}_n \xrightarrow{d} \mathbf{a}'X$ for all $\mathbf{a}$. Consider the m.g.f.s, by continuity theorem $M_{\mathbf{a}'\mathbf{X}_n}(t) = E\{e^{t\mathbf{a}'\mathbf{X}_n}\} \to E\{e^{ta=\mathbf{X}}\}$, for all t in the domain of convergence. Thus $E\{e^{(t\mathbf{a})'\mathbf{X}_n}\} \to E\{e^{(t\mathbf{a})'\mathbf{X}}\}$ for all $\boldsymbol{\beta} = t\mathbf{a}$. Thus, $\bar{X}_n \xrightarrow{d} \mathbf{X}$.

1.10.6 $X_n \sim B\left(n, \frac{1}{n}\right)$

$$
\begin{aligned}
E\{e^{-X_n}\} &= \left(\frac{1}{n}e^{-1} + 1 - \frac{1}{n}\right)^n \\
&= \left(1 - \frac{1}{n}(1 - e^{-1})\right)^n \underset{n\to\infty}{\longrightarrow} e^{-(1-e^{-1})}.
\end{aligned}
$$

Thus, $\lim_{n\to\infty} M_{X_n}(-1) = M_x(-1)$, where $X \sim P(1)$.

1.11.1 **(i)**

$$
\begin{aligned}
M_{\bar{X}_n}(t) &= \left(M_X\left(\frac{t}{n}\right)\right)^n \\
&= \left(E\left\{e^{\frac{t}{n}X}\right\}\right)^n \\
&= \left(1 + \frac{t}{n}E\{X\} + o\left(\frac{1}{n}\right)\right)^n \\
&= \left(1 + \frac{t}{n}\mu + o\left(\frac{1}{n}\right)\right)^n \underset{n\to\infty}{\longrightarrow} e^{t\mu}, \quad \forall t.
\end{aligned}
$$

$e^{t\mu}$ is the m.g.f. of the distribution

$$
F(x) = \begin{cases} 0, & x < \mu \\ 1, & x \ge \mu. \end{cases}
$$

Thus, by the continuity theorem, $\bar{X}_n \xrightarrow{d} \mu$ and, therefore, $\bar{X}_n \xrightarrow{p} \mu$, as $n \to \infty$.

1.11.5 $\{X_n\}$ are independent. For $\delta > 0$,

$$
X_n \sim R(-n, n)/n.
$$

The expected values are $E\{X_n\} = 0 \ \forall n \geq 1$.

$$\sigma_n^2 = \frac{4n^2}{12n^2} = \frac{1}{3}$$
$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty.$$

Hence, by (1.11.6), $\bar{X}_n \xrightarrow{\text{a.s.}} 0$.

1.12.1
$$M_{\frac{X-\lambda}{\sqrt{\lambda}}} = E\left\{e^{t(\frac{X-\lambda}{\sqrt{\lambda}})}\right\} = e^{-\sqrt{\lambda}t - \lambda(1-e^{t/\sqrt{\lambda}})}.$$

$$1 - \exp\{t/\sqrt{\lambda}\} = 1 - \left(1 + \frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \cdots\right)$$
$$= -\frac{t}{\sqrt{\lambda}} - \frac{t^2}{2\lambda} + \cdots .$$

Thus,

$$\sqrt{\lambda}t - \lambda(1 - e^{t\sqrt{\lambda}}) = \frac{t^2}{2} + O\left(\frac{1}{\sqrt{\lambda}}\right).$$

Hence,

$$M_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) = \exp\left\{\frac{t^2}{2} + O\left(\frac{1}{\sqrt{\lambda}}\right)\right\} \to e^{t^2/2} \quad \text{as } \lambda \to \infty.$$

$M_Z(t) = e^{t^2/2}$ is the m.g.f. of $N(0, 1)$.

1.12.3
$$P(X_n = 1) = \frac{1}{2},$$
$$P(X_n = 0) = \frac{1}{2},$$
$$E\{X_n\} = \frac{1}{2}.$$

Let $Y_n = nX_n$; $E\{Y_n\} = \frac{n}{2}$, $E|Y_n|^3 = \frac{n^3}{2}$. Notice that $\sum_{i=1}^{n} iX_i - \frac{n(n+1)}{4} = \sum_{i=1}^{n}(Y_i - \mu_i)$, where $\mu_i = \frac{i}{2} = E\{Y_i\}$. $E|Y_i - \mu_i|^3 = \frac{i^3}{8}$. Accordingly,

$$\frac{\sum_{i=1}^{n} E\{|Y_i - \mu_i|^3\}}{\sum_{i=1}^{n} E\{(Y_i - \mu_i)^2\}^{3/2}} = \frac{\frac{1}{4}n^2(n+1)^2}{\left(\frac{1}{24}n(n+1)(2n+1)\right)^{3/2}} \to 0 \quad \text{as } n \to \infty.$$

Thus, by Lyapunov's Theorem,

$$\frac{\sum_{i=1}^{n} iX_i - \frac{n(n+1)}{4}}{\left(\frac{1}{24}n(n+1)(2n+1)\right)^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

1.13.5 $\{X_n\}$ i.i.d. $B(1, p)$, $0 < p < 1$.

$$\hat{p}_n = \frac{1}{n}\sum X_i$$

$$W_n = \log\frac{\hat{p}_n}{(1 - \hat{p}_n)}.$$

(i)

$$E\{W_n\} \cong \log\frac{p}{1-p} + \frac{1}{2n}p(1-p)W''(p)$$

$$W'(p) = \frac{(1-p)(1-p+p)}{p(1-p)^2} = \frac{1}{p(1-p)}$$

$$W''(p) = -\frac{(1-2p)}{p^2(1-p)^2}.$$

Thus,

$$E\{W_n\} \cong log\left(\frac{p}{1-p}\right) - \frac{1-2p}{2np(1-p)}.$$

(ii)

$$V\{W_n\} \cong \frac{p(1-p)}{n} \cdot \frac{1}{(p(1-p))^2} = \frac{1}{np(1-p)}.$$