LECTURE NOTES[1]
ENEE 620
RANDOM PROCESSES IN COMMUNICATION AND CONTROL

Armand M. Makowski [2]

[2]Department of Electrical and Computer Engineering, and Institute for Systems Research, University of Maryland, College Park, MD 20742. E-mail: armand@isr.umd.edu. Phone: (301) 405-6844

Chapter 1

Limits in $\mathbb{R}$

We begin with a few standard definitions. We refer to a mapping $a : \mathbb{N}_0 \to \mathbb{R}$ as a ($\mathbb{R}$-valued) sequence; sometimes we also use the notation $\{a_n,\ n = 1, 2, \ldots\}$.

A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges to $a^\star$ in $\mathbb{R}$ if for every $\varepsilon > 0$, there exists an integer $n^\star(\varepsilon)$ such that

$$|a_n - a^\star| \leq \varepsilon, \quad n \geq n^\star(\varepsilon). \tag{1.1}$$

We shall write $\lim_{n\to\infty} a_n = a^\star$, and refer to the scalar $a^\star$ as the *limit* of the sequence.

Sometimes it is desirable to make sense of the situations where values of the sequence become either unboundedly large or unboundely negative, in which case we shall write $\lim_{n\to\infty} a_n = \infty$ and $\lim_{n\to\infty} a_n = -\infty$, respectively. A precise definition of such occurences is as follows: We write $\lim_{n\to\infty} a_n = \infty$ to signify that for every $M > 0$, there exists a finite integer $n^\star(M)$ in $\mathbb{N}_0$ such that

$$a_n > M, \quad n \geq n^\star(M). \tag{1.2}$$

It is natural to define $\lim_{n\to\infty} a_n = -\infty$ as $\lim_{n\to\infty} (-a_n) = \infty$.

If there exists $a^\star$ in $\mathbb{R} \cup \{\pm\infty\}$ such that $\lim_{n\to\infty} a_n = a^\star$, we shall simply say that the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ *converges* or *is convergent* (without any reference to its limit). Sometimes we shall also say that the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges *in* $\mathbb{R}$ to indicate that the limit $a^\star$ is an element of $\mathbb{R}$ (thus finite).

Applying the definition (1.1) requires that the limit be *known*. Often this information is not available, and yet the need remains to check whether the sequence converges. The notion of *Cauchy sequence*, which is instrumental in that respect, is built around the following observation: If the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges to $a^\star$ in $\mathbb{R}$, then for every $\varepsilon > 0$, there exists a finite integer $n^\star(\varepsilon)$ such that (1.1) holds, hence for $n, m \geq n^\star(\varepsilon)$ we have

$$|a_n - a_m| \leq |a_n - a^\star| + |a^\star - a_m| \leq \varepsilon + \varepsilon = 2\varepsilon.$$

This observation is turned into the following definition.

A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is said to be a *Cauchy sequence* if for every $\varepsilon > 0$, there exists an integer $n^\star(\varepsilon)$ such that

$$|a_n - a_m| \leq \varepsilon, \quad m, n \geq n^\star(\varepsilon). \tag{1.3}$$

As observed earlier, a convergent sequence $a : \mathbb{N}_0 \to \mathbb{R}$ in $\mathbb{R}$ is always a Cauchy sequence. It is a deep fact concerning the topological properties of $\mathbb{R}$ that being a Cauchy sequence is sufficient for convergence of the sequence in $\mathbb{R}$.

Theorem 1.0.1 *(Cauchy criterion) A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is convergent in $\mathbb{R}$ if and only if it is a Cauchy sequence.*

This provides a criterion for convergence which does not require knowledge of the limit.

1.1 Accumulation points

Since not all sequences converge, it is important to understand how can non-convergence occur. To that end, consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. A *subsequence* of the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is any sequence of the form $\mathbb{N}_0 \to \mathbb{R} : k \to a_{n_k}$ where

$$n_k < n_{k+1}, \quad k = 1, 2, \ldots$$

This forces $\lim_{k\to\infty} n_k = \infty$.

An *accumulation point* for the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is defined as any $a^\star$ in $\mathbb{R} \cup \{\pm\infty\}$ such that

$$\lim_{k\to\infty} a_{n_k} = a^\star$$

for *some* subsequence $\mathbb{N}_0 \to \mathbb{R} : k \to a_{n_k}$.

A convergent sequence $a : \mathbb{N}_0 \to \mathbb{R}$ has exactly one accumulation point, namely its limit. In fact, were the sequence *not* be convergent, it must necessarily have distinct accumulation points, in which case there is a smallest and a largest. The next definition formalizes this observation.

Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. The quantities

$$\bar{A} = \limsup_{n\to\infty} A_n = \inf_{n\geq 1}\left(\sup_{m\geq n} a_m\right)$$

and

$$\underline{\mathbf{A}} = \liminf_{n\to\infty} A_n, = \sup_{n\geq 1}\left(\inf_{m\geq n} a_m\right)$$

are known as the *limsup* and *liminf* of the sequence $a : \mathbb{N}_0 \to \mathbb{R}$.

The following notation is found to be convenient to discuss liminf and limsup quantities: For each $n = 1, 2, \ldots$, we define the quantities

$$\bar{A}_n = \sup_{m\geq n} a_m \quad \text{and} \quad \underline{\mathbf{A}}_n = \inf_{m\geq n} a_m \tag{1.4}$$

Note that $\underline{\mathbf{A}}_n \leq \bar{A}_n$, and the sequences $n \to \bar{A}_n$ and $n \to \underline{\mathbf{A}}_n$ are non-increasing and non-decreasing, respectively. Therefore, $\bar{A} = \lim_{n\to\infty} \bar{A}_n$ and $\underline{\mathbf{A}} = \lim_{n\to\infty} \underline{\mathbf{A}}_n$ both exist, but are possibly infinite. Moreover, we always have $\underline{\mathbf{A}} \leq \bar{A}$.

Theorem 1.1.1 *Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. If it converges to $a^\star$, then $\bar{A} = \underline{\mathbf{A}} = a^\star$. Conversely, if $\bar{A} = \underline{\mathbf{A}} = a^\star$ for some $a^\star$ in $\mathbb{R} \cup \{\pm\infty\}$, then the sequence converges to $a^\star$.*

Note that if $a, b : \mathbb{N}_0 \to \mathbb{R}$ are two sequences such that

$$a_n \leq b_n, \quad n = 1, 2, \ldots$$

then $\bar{A} \leq \bar{B}$ and $\underline{\mathbf{A}} \leq \underline{\mathbf{B}}$. The following arguments will often be made on the basis of this observation: Consider a sequence $\{p_n,\ n = 1, 2, \ldots\}$ where for each $n = 1, 2, \ldots$, p_n is the probability of some event so that

$$0 \leq p_n \leq 1, \quad n = 1, 2, \ldots \tag{1.5}$$

If we show that

$$1 \leq \liminf_{n\to\infty} p_n, \tag{1.6}$$

then we necessarily have convergence of the sequence with $\lim_{n\to\infty} p_n = 1$: Indeed, we always have $\limsup_{n\to\infty} p_n \leq 1$ as a result of (1.5), whence

$$\liminf_{n\to\infty} p_n = \limsup_{n\to\infty} p_n = 1$$

upon using (1.6). In a similar vein, if we show $\limsup_{n\to\infty} p_n = 0$, then we necessarily have convergence of the sequence with $\lim_{n\to\infty} p_n = 0$.

1.2 Two important facts

In addition to the Cauchy convergence criterion, here are two facts that are often found useful in studying convergence. A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is said to be *non-decreasing* (resp. *non-increasing*) if

$$a_n \leq a_{n+1} \quad \text{(resp. } a_{n+1} \leq a_n\text{)}, \quad n = 1, 2, \ldots$$

A sequence that is either non-decreasing or non-increasing is called a *monotone* sequence.

Theorem 1.2.1 *A monotone sequence $a : \mathbb{N}_0 \to \mathbb{R}$ always converges and we have $\lim_{n\to\infty} a_n = \sup(a_n,\ n = 1, 2, \ldots)$ (resp. $\lim_{n\to\infty} a_n = \inf(a_n,\ n = 1, 2, \ldots)$) if the sequence is non-decreasing (resp. non-increasing).*

A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ is said to be *bounded* if there exists some $B > 0$ such that

$$\sup(|a_n|,\ n = 1, 2, \ldots) \leq B.$$

Theorem 1.2.2 *(Bolzano-Weierstrass) For any bounded sequence $a : \mathbb{N}_0 \to \mathbb{R}$, there exists a convergent subsequence $\mathbb{N}_0 \to \mathbb{R} : k \to a_{n_k}$ with $\lim_{k\to\infty} a_{n_k} = a^\star$ for some $a^\star$ in $\mathbb{R}$.*

1.3 Cesaro convergence

With any sequence $a : \mathbb{N}_0 \to \mathbb{R}$ we associate the *Cesaro* sequence $\bar{a} : \mathbb{N}_0 \to \mathbb{R}$ given by

$$\bar{a}_n = \frac{1}{n}(a_1 + \ldots + a_n), \quad n = 1, 2, \ldots$$

Theorem 1.3.1 *(Cesaro convergence) If the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges to $a^\star$, then the Cesaro sequence $\bar{a} : \mathbb{N}_0 \to \mathbb{R}$ also converges with limit $a^\star$.*

Proof. First we assume the convergent sequence $a : \mathbb{N}_0 \to \mathbb{R}$ to have a finite limit $a^\star$ in $\mathbb{R}$. Note that

$$\bar{a}_n - a^\star = \frac{1}{n} \sum_{k=1}^{n} (a_k - a^\star), \quad n = 1, 2, \ldots$$

Now, for every $\varepsilon > 0$, there exists an integer $n^\star(\varepsilon)$ such that

$$|a_n - a^\star| \le \frac{\varepsilon}{2}, \quad n \ge n^\star(\varepsilon).$$

On that range, with $B(\varepsilon) = \sum_{k=1}^{n^\star(\varepsilon)} |a_k - a^\star|$, we have

$$\begin{aligned} |\bar{a}_n - a^\star| &\le \frac{1}{n} \sum_{k=1}^{n} |a_k - a^\star| \\ &= \frac{1}{n} \sum_{k=1}^{n^\star(\varepsilon)} |a_k - a^\star| + \frac{1}{n} \sum_{k=n^\star(\varepsilon)+1}^{n} |a_k - a^\star| \\ &\le \frac{B(\varepsilon)}{n} + \frac{n - n^\star(\varepsilon)}{n} \cdot \varepsilon \\ &\le \frac{B(\varepsilon)}{n} + \varepsilon \qquad (1.7) \end{aligned}$$

Since $\lim_{n\to\infty} \frac{1}{n} = 0$, it follows that for every $\varepsilon > 0$, there exists a finite integer $n^{\star\star}(\varepsilon)$ such that

$$\frac{1}{n} < \frac{\varepsilon}{B(\varepsilon)}, \quad n \ge n^{\star\star}(\varepsilon).$$

Just take $n^{\star\star}(\varepsilon) = \lceil \frac{B(\varepsilon)}{\varepsilon} \rceil$. As a result,

$$|\bar{a}_n - a^\star| \le \varepsilon + \varepsilon = 2\varepsilon, \quad n \ge \max(n^\star(\varepsilon), n^{\star\star}(\varepsilon))$$

and the proof is now complete since ε is arbitrary. We leave it as an exercise to show the result when $a^\star = \pm\infty$. ■

However, the converse is not true: Take the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ given by

$$a_n = (-1)^n, \quad n = 1, 2, \ldots$$

This sequence does not converge and yet $\lim_{n\to\infty} \bar{a}_n = 0$. This example nicely illustrate the smoothing effect of averaging. It might be tempting to conjecture that such averaging always produces a convergent sequence. However, this is not so as the following example shows: Consider the sequence $a \to \mathbb{R}$ given by

$$a_n = (-1)^k, \qquad \begin{matrix} 2^{2^k} \leq n < 2^{2^{k+1}} \\ k = 0, 1, \ldots \end{matrix}$$

It is plain that $\liminf_{n\to\infty} a_n = -1$ while $\limsup_{n\to\infty} a_n = 1$, and so the sequence $a \to \mathbb{R}$ does not converge. However, it is also not Cesaro convergent.

1.4 Series

Starting with a sequence $a : \mathbb{N}_0 \to \mathbb{R}$, we define the partial sums

$$s_n = a_1 + \ldots + a_n, \quad n = 1, 2, \ldots$$

where s_n is known as the n^{th} *partial sum*. We refer to the sequence $s : \mathbb{N}_0 \to \mathbb{R} : n \to s_n$ as the sequence of partial sums associated with the sequence $a : \mathbb{N}_0 \to \mathbb{R}$. It is customary to say that the series $\sum_{n=1}^{\infty} a_n$ converges if the sequence $s : \mathbb{N}_0 \to \mathbb{R}$ converges to some $s^\star$ in $\mathbb{R}$, in which case we often write $\sum_{n=1}^{\infty} a_n$ as its limit. This amounts to the following: For every $\varepsilon > 0$ there exists a finite integer $n^\star(\varepsilon)$ such that

$$|s_n - s^\star| < \varepsilon, \quad n \geq n^\star(\varepsilon).$$

The series $s : \mathbb{N}_0 \to \mathbb{R}$ is said to be *absolutely convergent* if the series associated with the sequence of absolute values $\mathbb{N}_0 \to \mathbb{R}_+ : n \to |a_n|$ does itself converge in $\mathbb{R}$.

A series which is absolutely convergent is also convergent in the usual sense since

$$\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k|, \qquad \begin{matrix} m = n+1, \ldots \\ n = 1, 2, \ldots \end{matrix}$$

However, the converse is not true as is easily seen through the example

$$a_n = \frac{(-1)^n}{n}, \quad n = 1, 2, \ldots$$

A series which is convergent in the usual sense but not absolutely convergent is said to be *conditionally* convergent.

When the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ assumes only non-negative values, i.e., $a_n \geq 0$ for all $n = 1, 2, \ldots$, then the sequence $s : \mathbb{N}_0 \to \mathbb{R}_+$ of partial sums is non-decreasing, so that $\lim_{n\to\infty} s_n$ always exists, possibly infinite. When this limit is finite, it is easy to establish the following fact.

Lemma 1.4.1 *For any sequence $a : \mathbb{N}_0 \to \mathbb{R}$ whose sequence of partial sums converges in $\mathbb{R}$, we have $\lim_{n\to\infty} a_n = 0$*

Proof. Since the sequence of partial sums $s : \mathbb{N}_0 \to \mathbb{R}$ converges in $\mathbb{R}$, it is a Cauchy sequence: For every $\varepsilon > 0$, there exists a finite integer $n^\star(\varepsilon)$ such that

$$|s_n - s_m| \leq \varepsilon, \quad n, m \geq n^\star(\varepsilon).$$

Selecting $m = n + 1$ with $n \geq n^\star(\varepsilon)$, we get $|a_{n+1}| = |s_n - s_{n+1}| \leq \varepsilon$ whenever $n \geq n^\star(\varepsilon)$, and the conclusion $\lim_{n\to\infty} a_n = 0$ follows. ■

Many tests exist to check the convergence of series. The most basic one is the Comparison Test given next.

Theorem 1.4.1 *(Comparison Test) Consider two sequences $a, b : \mathbb{N}_0 \to \mathbb{R}_+$ such that*

$$0 \leq a_n \leq b_n, \quad n = 1, 2, \ldots$$

If $\sum_{n=1}^\infty b_n$ converges in $\mathbb{R}$, then $\sum_{n=1}^\infty a_n$ also converges n $\mathbb{R}$ with

$$0 \leq \sum_{n=1}^\infty a_n \leq \sum_{n=1}^\infty b_n.$$

On the other hand, if $\sum_{n=1}^\infty a_n = \infty$, then we necessarily have $\sum_{n=1}^\infty b_n = \infty$.

Geometric series play a pivotal role in determining the convergence of series through the Comparison Test. The *geometric* series with reason ρ is the series associated with the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ given by

$$a_n = \rho^n, \quad n = 1, 2, \ldots$$

It well known that

$$s_n = a_1 + \ldots + a_n = \begin{cases} \frac{\rho}{1-\rho}(1-\rho^n) & \text{if } \rho \neq 1 \\ \\ n & \text{if } \rho = 1 \end{cases}$$

Therefore,

$$\lim_{n\to\infty} s_n = \frac{\rho}{1-\rho} \quad \text{if } |\rho| < 1.$$

This observation constitutes the basis for two criteria for convergence of series, namely the criteria of Cauchy and d' Alembert, also known as the Root Test and Ratio Test, respectively.

Theorem 1.4.2 *(Ratio Test) Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. Assume that the limit*

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = R$$

exists (possibly infinite). Then,

$$\sum_{n=1}^{\infty} |a_n| < \infty \quad \textit{if } R < 1 \tag{1.8}$$

and

$$\sum_{n=1}^{\infty} |a_n| = \infty \quad \textit{if } 1 < R. \tag{1.9}$$

Theorem 1.4.3 *(Root Test) Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. Assume that the limit*

$$\lim_{n\to\infty} \sqrt[n]{|a_n|} = R$$

exists. Then,

$$\sum_{n=1}^{\infty} |a_n| < \infty \quad \textit{if } R < 1 \tag{1.10}$$

and

$$\sum_{n=1}^{\infty} |a_n| = \infty \quad \textit{if } 1 < R. \tag{1.11}$$

1.5 Power series

In a number of places we shall need to understand the behavior of series that belong to the class of *power series*. With any sequence $a : \mathbb{N}_0 \to \mathbb{R}$ we associate the *formal* power series

$$\sum_{n=1}^{\infty} a_n z^n, \quad z \in \mathbb{C}.$$

A natural question arises as to when such formal series are in fact convergent. In particular, we define the *domain of convergence* of the power series as the set $\mathcal{C}$ given by

$$\mathcal{C} = \{z \in \mathbb{C} : \sum_{n=0}^{\infty} |a_n||z|^n < \infty\}.$$

This region is determined by the asymptotic behavior of the sequence $a : \mathbb{N}_0 \to \mathbb{R}$. This is the content of the following well-known result which is a consequence of the Root Test (applied to the sequence $\{a_n z^n,\ n = 0, 1, \ldots\}$).

Theorem 1.5.1 *With*

$$R = \limsup_{n\to\infty} \sqrt[n]{|a_n|},$$

we have

$$\sum_{n=1}^{\infty} |a_n||z|^n < \infty \quad \text{if } |z| < R^{-1}$$

and

$$\sum_{n=1}^{\infty} |a_n||z|^n = \infty \quad \text{if } R^{-1} < |z|$$

The open disk $\{z \in \mathbb{C} : |z| < R^{-1}\}$ is therefore contained in $\mathcal{C}$.

Chapter 2

Probability distribution functions and their transforms

A number of developments concerning rvs and their probability distribution functions are sometimes best handled through transforms associated with them. There are a number of such transforms with varying ranges of applications. Here we focus mainly on the notion of characteristic function.

2.1 Definitions

With any rv $X : \Omega \to \mathbb{R}$, we associate its *characteristic function* $\Phi_X : \mathbb{R} \to \mathbb{C}$ given by

$$\Phi_X(\theta) := \mathbb{E}\left[e^{i\theta X}\right], \quad \theta \in \mathbb{R}. \tag{2.1}$$

The definition (2.1) is well posed since for each θ in $\mathbb{R}$, the rvs $\Omega \to \mathbb{R} : \omega \to \cos(\theta X(\omega))$ and $\Omega \to \mathbb{R} : \omega \to \sin(\theta X(\omega))$ are both bounded. As a result, their expected values $\mathbb{E}[\cos(\theta X)]$ and $\mathbb{E}[\sin(\theta X)]$ are well defined. This fact allows us to make sense of (2.1) in the usual way by linearity through

$$\mathbb{E}\left[e^{i\theta X}\right] = \mathbb{E}[\cos(\theta X) + i\sin(\theta X)] = \mathbb{E}[\cos(\theta X)] + i\mathbb{E}[\sin(\theta X)].$$

Characteristic functions are akin to Fourier transforms. In fact, if the rv X admits a probability density function $f_X : \mathbb{R} \to \mathbb{R}_+$, then

$$\Phi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f_X(x)dx, \quad \theta \in \mathbb{R}.$$

Much of the discussion makes use of the elementary relation

$$e^{i\theta x} - 1 = \int_0^x i\theta e^{i\theta s} ds, \quad x, \theta \in \mathbb{R} \tag{2.2}$$

so that the bounds

$$\left| e^{i\theta x} - 1 \right| \leq \int_0^x \left| i\theta e^{i\theta s} \right| ds \leq |\theta| x \tag{2.3}$$

hold.[1] Obviously, the characteristic function Φ_X of the rv X is determined by its probability distribution function $F_X : \mathbb{R} \to [0, 1]$.

2.2 Basic properties

Here are some simple properties.

Theorem 2.2.1 *Consider a rv $X : \Omega \to \mathbb{R}$ with characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ given by (2.1). It satisfies the following properties:*

(i) Boundedness: We have

$$|\Phi_X(\theta)| \leq \Phi_X(0) = 1 \quad \theta \in \mathbb{R}. \tag{2.4}$$

(ii) Uniform continuity on $\mathbb{R}$: We have

$$\lim_{\delta \to 0} \sup \left(|\Phi_X(\theta + \delta) - \Phi_X(\theta)| , \quad \theta \in \mathbb{R} \right) = 0. \tag{2.5}$$

(iii) Positive semi-definiteness: For every $n = 1, 2, \ldots$, we have

$$\sum_{k=1}^{n} \sum_{\ell=1}^{n} \Phi_X(\theta_k - \theta_\ell) z_k z_\ell^\star \geq 0 \tag{2.6}$$

with arbitrary $z_1, \ldots, z_n$ in $\mathbb{C}$.

---

[1] With ab in $\mathbb{R}$, we have

$$|a + ib| = \sqrt{a^2 + b^2} \leq |a| + |b|.$$

Proof. (i) It is plain that $\Phi_X(0) = 1$. Next,

$$|\Phi_X(\theta)| \leq \mathbb{E}\left[\left|e^{i\theta X}\right|\right] = 1, \quad \theta \in \mathbb{R}.$$

(ii) Fix θ and δ in $\mathbb{R}$. Since

$$e^{i(\theta+\delta)X} - e^{i\theta X} = e^{i\theta X}\left(e^{i\delta X} - 1\right),$$

it follows that

$$\begin{aligned} |\Phi_X(\theta+\delta) - \Phi_X(\theta)| &= \left|\mathbb{E}\left[e^{i(\theta+\delta)X}\right] - \mathbb{E}\left[e^{i\theta X}\right]\right| \\ &= \left|\mathbb{E}\left[\left(e^{i\delta X} - 1\right)e^{i\theta X}\right]\right| \\ &\leq \mathbb{E}\left[\left|\left(e^{i\delta X} - 1\right)e^{i\theta X}\right|\right] \\ &= \mathbb{E}\left[\left|e^{i\delta X} - 1\right|\right], \end{aligned}$$

so that

$$\sup\left(|\Phi_X(\theta+\delta) - \Phi_X(\theta)|, \quad \theta \in \mathbb{R}\right) \leq \mathbb{E}\left[\left|e^{i\delta X} - 1\right|\right]. \tag{2.7}$$

Uniform continuity follows if we can show that

$$\lim_{\delta\to 0} \mathbb{E}\left[\left|e^{i\delta X} - 1\right|\right] = 1.$$

This last statement is a simple consequence of the Bounded Convergence Theorem. ■

(iii) Fix $n = 1, 2, \ldots$ and pick arbitrary $z_1, \ldots, z_n$ in $\mathbb{C}$: It is plain that

$$\begin{aligned} &\sum_{k=1}^n \sum_{\ell=1}^n \Phi_X(\theta_k - \theta_\ell) z_k z_\ell^\star \\ &\quad = \sum_{k=1}^n \sum_{\ell=1}^n \mathbb{E}\left[e^{j(\theta_k-\theta_\ell)X}\right] z_k z_\ell^\star \\ &\quad = \mathbb{E}\left[\sum_{k=1}^n \sum_{\ell=1}^n e^{j(\theta_k-\theta_\ell)X} z_k z_\ell^\star\right] \\ &\quad = \mathbb{E}\left[\sum_{k=1}^n \sum_{\ell=1}^n e^{j\theta_k X} e^{-\theta_\ell X} z_k z_\ell^\star\right] \end{aligned}$$

$$= \mathbb{E}\left[\left(\sum_{k=1}^{n} e^{j\theta_k X} z_k\right)\left(\sum_{\ell=1}^{n} e^{j\theta_\ell X} z_\ell\right)^{\star}\right]$$

$$= \mathbb{E}\left[\left|\sum_{k=1}^{n} e^{j\theta_k X} z_k\right|^2\right] \geq 0. \tag{2.8}$$

■

2.3 Bochner's Theorem

Sometimes a function $\Phi : \mathbb{R} \to \mathbb{C}$ arises in the discussion, and it is imperative to know whether it is the characteristic function of some rv. The terminology given next should facilitate the discussion.

A function $\Phi : \mathbb{R} \to \mathbb{C}$ is said to be a *characteristic function* if there exists a rv $X : \Omega :\to \mathbb{R}$ such that

$$\Phi(\theta) = \mathbb{E}\left[e^{\theta X}\right] = \Phi_X(\theta), \quad \theta \in \mathbb{R}.$$

Alternatively, a function $\Phi : \mathbb{R} \to \mathbb{C}$ is said to be a characteristic function if there exists a probability distribution $F : \mathbb{R} \to [0, 1]$ such that

$$\Phi(\theta) = \int_{\mathbb{R}} e^{itx} dF(x), \quad \theta \in \mathbb{R}.$$

Not every function $\Phi : \mathbb{R} :\to \mathbb{C}$ is a characteristic function. That much is clear from the basic properties derived in Theorem 2.2.1. Interestingly enough the three properties given there turn out to be sufficient. This is a consequence of a deep result of Harmonic Analysis, known as the Bochner-Herglotz Theorem [**?**, Thm. 6.5.2, p. 179].

Theorem 2.3.1 *A function $\Phi : \mathbb{R} \to \mathbb{C}$ is a characteristic function if it is (i) bounded with $|\Phi(\theta)| \leq \Phi(0) = 1$ for all θ in $\mathbb{R}$; (ii) uniformly continuous on $\mathbb{R}$; and (iii) positive semi-definite.*

The property of positive semi-definiteness already implies the boundedness property (i). It also implies uniform continuity if $\Phi : \mathbb{R} \to \mathbb{C}$ is continuous at $\theta = 0$ [**?**, Thm. 6.5.1, p. 178]. This gives rise to the following sharp characterizzation.

Theorem 2.3.2 *A function $\Phi : \mathbb{R} \to \mathbb{C}$ is a characteristic function if and only if it is positive semi-definite and continuous at $\theta = 0$ with $\Phi(0) = 1$.*

2.4 Easy analytical facts

We begin with a simple fact that will prove useful in a number of places.

Theorem 2.4.1 *Fix x and θ in $\mathbb{R}$. For each $k = 1, 2, \ldots$, the expansion*

$$e^{i\theta x} = \sum_{\ell=0}^{k} \frac{1}{\ell!} (i\theta x)^{\ell} + R_k(x;\theta) \tag{2.9}$$

holds with the remainder term given by

$$R_k(x;\theta) = (i\theta)^k \int_0^x \frac{(x-t)^{k-1}}{(k-1)!} \left(e^{i\theta t} - 1\right) dt. \tag{2.10}$$

Proof. The proof proceed by induction: Throughout θ and x in $\mathbb{R}$ are scalars held fixed.

Basis step For $k = 1$, we use (2.2) to get

$$\begin{aligned}
e^{i\theta x} - 1 &= \int_0^x i\theta e^{i\theta t} dt \\
&= \int_0^x i\theta \left(e^{i\theta t} - 1\right) dt + \int_0^x i\theta dt \\
&= i\theta x + i\theta \int_0^x \left(e^{i\theta t} - 1\right) dt \\
&= i\theta x + R_1(x;\theta) \qquad (2.11)
\end{aligned}$$

by direct inspection.

Induction step Now assume that (2.10)-(2.10) holds for some $k = 1, 2, \ldots$. It is plain that

$$\begin{aligned}
&\int_0^x \frac{(x-t)^{k-1}}{(k-1)!}\left(e^{i\theta t}-1\right) dt \\
&= \int_0^x \frac{(x-t)^{k-1}}{(k-1)!}\left(\int_0^t i\theta e^{i\theta s} ds\right) dt \\
&= \int_0^x \left(\int_0^t \frac{(x-t)^{k-1}}{(k-1)!} i\theta e^{i\theta s} ds\right) dt \\
&= \int_0^x \left(\int_s^x \frac{(x-t)^{k-1}}{(k-1)!} i\theta e^{i\theta s} dt\right) ds \\
&= \int_0^x \left(\int_s^x \frac{(x-t)^{k-1}}{(k-1)!} dt\right) i\theta e^{i\theta s} ds \\
&= \int_0^x i\theta \frac{(x-s)^k}{k!} e^{i\theta s} ds \qquad (2.12)
\end{aligned}$$

since

$$\int_s^x \frac{(x-t)^{k-1}}{(k-1)!} dt = \left[-\frac{(x-t)^k}{k!}\right]_s^x = \frac{(x-s)^k}{k!}, \quad 0 \le s \le x.$$

Therefore, we have

$$\begin{aligned}
R_k(x;\theta) &= (i\theta)^k \int_0^x \frac{(x-t)^{k-1}}{(k-1)!}\left(e^{i\theta t}-1\right) dt \\
&= (i\theta)^{k+1} \int_0^x \frac{(x-s)^k}{k!} e^{i\theta s} ds \\
&= (i\theta)^{k+1} \int_0^x \frac{(x-s)^k}{k!}\left(e^{i\theta s}-1\right) ds + (i\theta)^{k+1} \int_0^x \frac{(x-s)^k}{k!} ds \\
&= R_{k+1}(x;\theta) + (i\theta)^{k+1} \frac{x^{k+1}}{(k+1)!} \qquad (2.13)
\end{aligned}$$

and the proof of the induction step is now completed. ■

2.5 Characteristic functions and moments

Since the probability distribution function of the rv X can be recovered from its characteristic function, it is not unreasonable to expect that there might be simple ways to recover moments whenever they exist and are finite. This is explored below.

Consider a rv $X : \Omega \to \mathbb{R}$ with characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ given by (2.1). Fix θ in $\mathbb{R}$. It follows from Theorem 2.4.1 that

$$e^{i\theta X} - \sum_{\ell=0}^{k} \frac{1}{\ell!} (i\theta X)^{\ell} = R_k(X;\theta) \tag{2.14}$$

Therefore, if the rv X has a finite moment of order k for some $k = 1, 2, \ldots$, the expectation

$$\mathbb{E}\left[R_k(X;\theta)\right]$$

exists and is well defined since all the moments of X of order $\ell = 1, 2, \ldots, k$ exist and are finite. Thus, the relationship

$$\mathbb{E}\left[e^{i\theta X}\right] = \sum_{\ell=0}^{k} \frac{1}{\ell!} (i\theta)^{\ell}\, \mathbb{E}\left[X^{\ell}\right] + \mathbb{E}\left[R_k(X;\theta)\right] \tag{2.15}$$

does hold. This suggests the following result.

Theorem 2.5.1 *Consider a rv $X : \Omega \to \mathbb{R}$ with characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ given by (2.1). If $\mathbb{E}\left[|X|^n\right] < \infty$ for some $n = 1, 2, \ldots$, then for each $k = 1, 2, \ldots, n$, the characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ is everywhere k^{th} differentiable with*

$$\frac{d^k}{d\theta^k}\Phi_X(\theta) = \mathbb{E}\left[(iX)^k e^{i\theta X}\right], \quad \theta \in \mathbb{R}. \tag{2.16}$$

Proof. If $k = 1$. Fix θ in $\mathbb{R}$ and for each $h \neq 0$ note that

$$\begin{aligned} \Phi_X(\theta + h) - \Phi_X(\theta) &= \mathbb{E}\left[e^{i\theta X}\left(e^{ihX} - 1\right)\right] \\ &= \mathbb{E}\left[e^{i\theta X} \int_0^X ihe^{iht} dt\right] \end{aligned} \tag{2.17}$$

so that

$$\frac{1}{h}\left(\Phi_X(\theta+h)-\Phi_X(\theta)\right)=\mathbb{E}\left[e^{i\theta X}\int_0^X ie^{iht}dt\right].$$

The bound

$$\left|e^{i\theta X}\int_0^X ie^{iht}dt\right| = \left|e^{i\theta X}\right|\left|\int_0^X ie^{iht}dt\right| \leq |X| \tag{2.18}$$

holds *uniformly* in $h \neq 0$, whence

$$\lim_{h\to 0}\left(e^{i\theta X}\int_0^X ie^{iht}dt\right) = (iX)\,e^{i\theta X}$$

by the Bounded Convergence Theorem. We now conclude that

$$\begin{aligned}
\lim_{h\to 0}\frac{1}{h}\left(\Phi_X(\theta+h)-\Phi_X(\theta)\right) &= \lim_{h\to 0}\mathbb{E}\left[e^{i\theta X}\int_0^X ie^{iht}dt\right]. \\
&= \mathbb{E}\left[\lim_{h\to 0}\left(e^{i\theta X}\int_0^X ie^{iht}dt\right)\right] \\
&= \mathbb{E}\left[(iX)\,e^{i\theta X}\right] \qquad (2.19)
\end{aligned}$$

by the Dominated Convergence Theorem and the conclusion (2.16) holds for $k = 1$.

If $k \geq 2$, we proceed by induction: The basis step was just established. To establish the induction step, assume that for each $\ell = 1, \ldots, k-1$, the characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ is everywhere ℓ^{th} differentiable with

$$\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta) = \mathbb{E}\left[(iX)^\ell\, e^{i\theta X}\right], \quad \theta \in \mathbb{R}. \tag{2.20}$$

Under the assumption $\mathbb{E}\left[|X|^k\right] < \infty$, we shall now show that the characteristic function $\Phi_X : \mathbb{R} \to \mathbb{C}$ is everywhere $(\ell+1)^{rst}$ differentiable with

$$\frac{d^{\ell+1}}{d\theta^{\ell+1}}\Phi_X(\theta) = \mathbb{E}\left[(iX)^{\ell+1}\, e^{i\theta X}\right], \quad \theta \in \mathbb{R}. \tag{2.21}$$

Indeed, for every $h \neq 0$, we have

$$\begin{aligned}
\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta+h) - \frac{d^\ell}{d\theta^\ell}\Phi_X(\theta) &= \mathbb{E}\left[(iX)^\ell\left(e^{i(\theta+h)X}-e^{i\theta X}\right)\right] \\
&= \mathbb{E}\left[(iX)^\ell\, e^{i\theta X}\left(e^{ihX}-1\right)\right] \\
&= \mathbb{E}\left[(iX)^\ell\, e^{i\theta X}\int_0^X ihe^{iht}dt\right]
\end{aligned}$$

so that

$$\frac{1}{h}\left(\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta+h)-\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta)\right)=\mathbb{E}\left[(iX)^\ell e^{i\theta X}\int_0^X ie^{iht}dt\right]$$

Again we see that

$$\left|(iX)^\ell e^{i\theta X}\int_0^X ie^{iht}dt\right|\leq |X|^{\ell+1}$$

uniformly in $h \neq 0$ with $\mathbb{E}\left[|X|^{\ell+1}\right]<\infty$ by assumption. Invoking the Dominated Convergence Theorem we conclude that

$$\begin{aligned}
&\lim_{h\to 0}\frac{1}{h}\left(\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta+h)-\frac{d^\ell}{d\theta^\ell}\Phi_X(\theta)\right)\\
&= \lim_{h\to 0}\mathbb{E}\left[(iX)^\ell e^{i\theta X}\int_0^X ie^{iht}dt\right]\\
&= \mathbb{E}\left[(iX)^\ell e^{i\theta X}\lim_{h\to 0}\int_0^X ie^{iht}dt\right]\\
&= \mathbb{E}\left[(iX)^{\ell+1} e^{i\theta X}\right], \qquad (2.22)
\end{aligned}$$

and this establishes (2.21) holds. This concludes the induction step as we have now shown that (2.20) holds for $\ell = 1, \ldots, k$.

■

Chapter 3

Convergence of random variables

In Chapter 1 we reviewed basic notions of convergence in $\mathbb{R}$. In the present chapter we turn to developing a convergence theory for sequences of rvs.

Before we proceed several remarks are in order:

Basic points: (i) Compatibility with convergence in $\mathbb{R}$ (ii) Dual perspective on rvs: mappings vs. probability distributions!

3.1 Almost sure convergence

Consider rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$.

The sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ *converges almost surely* (a.s.) to the rv X if $\mathbb{P}[C] = 1$ where C is the event

$$C = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}.$$

We shall write $\lim_{n\to\infty} X_n = X$ a.s. Sometimes the qualifier "almost sure(ly)" is replaced by the qualifier "with probability one" (often abbreviated as wp 1), in which case we write $\lim_{n\to\infty} X_n = X$ wp 1.

It is easy to see that the convergence set C is indeed an event in $\mathcal{F}$ since

$$C = \cap_{k=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \left[|X_n - X| \leq \frac{1}{k} \right].$$

The following notation will prove convenient in what follows: Pick $\varepsilon > 0$ arbitrary, and for each $n = 1, 2, \ldots$, define the events

$$A_n(\varepsilon) = [|X_n - X| \leq \varepsilon]$$

and

$$\begin{aligned} B_n(\varepsilon) &= \cap_{m \geq n} A_m(\varepsilon) \\ &= [|X_n - X| \leq \varepsilon,\ m = n, n+1, \ldots]\,. \end{aligned} \tag{3.1}$$

Theorem 3.1.1 *The sequence of rvs* $\{X_n,\ n = 1, 2, \ldots\}$ *converges a.s. to the rv* X *if and only if*

$$\mathbb{P}\left[B_\infty(\varepsilon)\right] = 1, \quad \varepsilon > 0 \tag{3.2}$$

with

$$B_\infty(\varepsilon) = \cup_{n=1}^\infty B_n(\varepsilon). \tag{3.3}$$

Proof. With this notation, the characterization of C given earlier can now be expressed in the more compact form

$$C = \cap_{k=1}^\infty B_\infty(k^{-1}).$$

Note also that $B_\infty(\varepsilon') \subseteq B_\infty(\varepsilon)$ whenever $0 < \varepsilon' < \varepsilon$. Hence, by the continuity property of $\mathbb{P}$ under monotone limits, we get

$$\mathbb{P}\left[C\right] = \lim_{k \to \infty} \mathbb{P}\left[B_\infty(k^{-1})\right]. \tag{3.4}$$

This last convergence being monotonically decreasing as k increases, we conclude that $\mathbb{P}\left[C\right] = 1$ if and only if

$$\mathbb{P}\left[B_\infty(k^{-1})\right] = 1, \quad k = 1, 2, \ldots.$$

The conclusion follows since for every $\varepsilon > 0$ there exists a positive integer k such that $(k+1)^{-1} \leq \varepsilon \leq k^{-1}$ with $B_\infty((k+1)^{-1}) \subseteq B_\infty(\varepsilon) \subseteq B_\infty(k^{-1})$. ■

This simple observation paves the way for the following simple criterion for a.s. convergence.

Theorem 3.1.2 *The sequence of rvs* $\{X_n,\ n = 1, 2, \ldots\}$ *converges a.s. to the rv* X *if for every* $\varepsilon > 0$*, it holds that*

$$\sum_{n=1}^{\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] < \infty. \tag{3.5}$$

Proof. Pick $\varepsilon > 0$. Note that $B_\infty(\varepsilon) = \liminf_{n\to\infty} A_n(\varepsilon)$, or equivalently, $B_\infty(\varepsilon)^c = \limsup_{n\to\infty} A_n(\varepsilon)^c$. The first part of the Borel-Cantelli Lemma now yields $\mathbb{P}\left[B_\infty(\varepsilon)^c\right] = 0$ provided

$$\sum_{n=1}^{\infty} \mathbb{P}\left[A_n(\varepsilon)^c\right] < \infty.$$

This is equivalent to $\mathbb{P}\left[B_\infty(\varepsilon)\right] = 1$ provided (3.5) holds, and the proof is completed by invoking Theorem 3.1.1. ■

The sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ is said to be *completely convergent* to the rv X if for every $\varepsilon > 0$, we have

$$\sum_{n=1}^{\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] < \infty. \tag{3.6}$$

By Theorem 3.1.2 we see that complete convergence implies a.s. convergence. But complete convergence is only a sufficient condition for a.s. convergence, and not a necessary condition. The next example shows that the converse does not hold.

A.s. convergence does not imply complete convergence

Take $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ and $\mathbb{P}$ is Lebesgue measure λ. Define the rvs $\{X_n,\ n = 1, 2, \ldots\}$ to be

$$X_n = \begin{cases} 1 & \text{if } 0 \le \omega \le 1 - \frac{1}{n} \\ \\ 0 & \text{if } 1 - \frac{1}{n} < \omega \le 1 \end{cases}$$

for every $n = 1, 2, \ldots$. Fix ω in $[0,)$. It is plain that $\lim_{n\to\infty} X_n(\omega) = 0$, and the sequence $\{X_n,\ n = 1, 2, \ldots\}$ converges a.s. to the rv $X \equiv 0$. However, for every ε in $(0, 1)$, we get

$$\mathbb{P}\left[|X_n| > \varepsilon\right] = \frac{1}{n}, \quad n = 1, 2, \ldots$$

whence (3.6) fails since $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$.

3.2 Convergence in probability

Consider rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$.

The sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ *converges in probability* to the rv X if for every $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] = 0.$$

We shall write $X_n \xrightarrow{P} {}_n X$.

Convergence in probability admits the following Cauchy criterion.

Theorem 3.2.1 *(Cauchy criterion for convergence in probability) The sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in probability if and only if for every $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} \left(\sup_{m\geq n} \mathbb{P}\left[|X_n - X_m| > \varepsilon\right] \right) = 0. \tag{3.7}$$

A.s. convergence is a stronger notion of convergence than convergence in probability.

Theorem 3.2.2 *Almost sure convergence implies convergence in probability: If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges a.s. to the rv X, then it also converges in probability to the rv X.*

Proof. Pick $\varepsilon > 0$ arbitrary. We have $B_n(\varepsilon) \subseteq A_n(\varepsilon)$ for each $n = 1, 2, \ldots$, whence

$$\mathbb{P}\left[B_n(\varepsilon)\right] \leq \mathbb{P}\left[A_n(\varepsilon)\right], \quad n = 1, 2, \ldots$$

The sets $\{B_n(\varepsilon),\ n = 1, 2, \ldots\}$ being non-decreasing, we readily conclude that $\lim_{n\to\infty} \mathbb{P}\left[B_n(\varepsilon)\right] = \mathbb{P}\left[B_\infty(\varepsilon)\right]$ with $B_\infty(\varepsilon)$ defined at (3.3). It is now plain that

$$\mathbb{P}\left[B_\infty(\varepsilon)\right] = \lim_{n\to\infty} \mathbb{P}\left[B_n(\varepsilon)\right] \leq \liminf_{n\to\infty} \mathbb{P}\left[A_n(\varepsilon)\right].$$

By Theorem 3.1.1 the a.s. convergence of the sequence $\{X_n,\ n = 1, 2, \ldots\}$ implies $\mathbb{P}\left[B_\infty(\varepsilon)\right] = 1$, and this immediately implies $\liminf_{n\to\infty} \mathbb{P}\left[A_n(\varepsilon)\right] = 1$. Thus, $\lim_{n\to\infty} \mathbb{P}\left[A_n(\varepsilon)\right] = 1$, and the sequence $\{X_n,\ n = 1, 2, \ldots\}$ converges in probability. ■

Here is an example of a sequence which converges in probability but does not converge almost surely:

A counterexample

Take $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ and $\mathbb{P}$ is Lebesgue measure λ. Define the rvs $\{X_n,\ n = 1, 2, \ldots\}$ as follows: For each $n = 1, 2, \ldots$, there exists a unique integer $k = 0, 1, \ldots$ such that $2^k \leq n < 2^{k+1}$ so that $n = 2^k + m$ for some unique $m = 0, \ldots, 2^k - 1$. Define

$$X_n = \begin{cases} 1 & \text{if } \omega \in I_n \\ \\ 0 & \text{if } \omega \notin I_n \end{cases}$$

where $I_n = (m2^{-k}, (m + 1)2^{-k})$.

The set Ω_b of boundary points

$$\Omega_b = \left\{m2^{-k},\ m = 0, \ldots, 2^k,\ k = 0, 1, \ldots\right\}$$

is countable, hence $\mathbb{P}\left[\Omega_b\right] = 0$. With ω not in Ω_b we note that $X_n(\omega) = 0$ and $X_n(\omega) = 1$ infinitely often, so that $\liminf_{n\to\infty} X_n(\omega) = 0 < \limsup_{n\to\infty} X_n(\omega) = 1$. The sequence $\{X_n,\ n = 1, 2, \ldots\}$ therefore does not converge a.s.. However, with $X = 0$, we have $\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] = 0$ for every $\varepsilon > 0$ since

$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] = \begin{cases} \mathbb{P}\left[I_n\right] & \text{if } 0 < \varepsilon < 1 \\ \\ 0 & \text{if } 1 \geq \varepsilon. \end{cases}$$

The sequence $\{X_n,\ n = 1, 2, \ldots\}$ indeed converges in probability.

Yet, despite this counterexample which shows that a.s. convergence is strictly stronger than convergence in probability, there is a partial converse in the following sense.

Theorem 3.2.3 *Convergence in probability implies almost sure convergence but only along a subsequence: If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in probability to the rv X, then there exists a (deterministic) subsequence $\mathbb{N}_0 \to \mathbb{N}_0$ with*

$$n_k < n_{k+1}, \quad k = 1, 2, \ldots$$

such that the subsequence of rvs $\{X_{n_k},\ k = 1, 2, \ldots\}$ converges almost surely to X.

Proof. The assumed convergence in probability of the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ to the rv X amounts to

$$\lim_{n\to\infty} \mathbb{P}\left[|X - X_n| > \varepsilon\right] = 0, \quad \varepsilon > 0.$$

More precisely, fix $\varepsilon > 0$. Then, for every $\delta > 0$ there exists a positive integer $n^\star(\varepsilon, \delta)$ such that

$$\mathbb{P}\left[|X - X_n| > \varepsilon\right] \leq \delta, \quad n \geq n^\star(\varepsilon, \delta).$$

We now use this observation (with $\varepsilon = k^{-1}$ and $\delta = 2^{-k}$) as follows: For each $k = 1, 2, \ldots$, there exists a positive integer n_k such that

$$\mathbb{P}\left[|X - X_n| > k^{-1}\right] \leq 2^{-k}, \quad n \geq n_k.$$

It is always possible to select n_k as any positive integer satisfying

$$\max\left(n^\star(\varepsilon, \delta), n_{k-1}\right) < n_k$$

with the convention $n_0 = 0$. This construction guarantees $n_k < n_{k+1}$ for all $k = 1, 2, \ldots$.

Pick $\varepsilon > 0$ and introduce the integer $k(\varepsilon) = \lfloor \varepsilon^{-1} \rfloor$. With the quantities just introduced we have

$$\begin{aligned} &\sum_{k=1}^{\infty} \mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] \\ &= \sum_{k=1,2,\ldots:\ k^{-1}>\varepsilon} \mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] + \sum_{k=1,2,\ldots:\ k^{-1}\leq\varepsilon} \mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] \\ &\leq k(\varepsilon) + \sum_{k=k(\varepsilon)}^{\infty} \mathbb{P}\left[|X_{n_k} - X| > k^{-1}\right] \\ &\leq k(\varepsilon) + \sum_{k=k(\varepsilon)}^{\infty} 2^{-k} \end{aligned}$$

and the conclusion $\sum_{k=1}^{\infty} \mathbb{P}\left[|X_{n_k} - X| > \varepsilon\right] < \infty$ follows. The desired a.s. convergence of the sequence of rvs $\{X_{n_k},\ k = 1, 2, \ldots\}$ is now a consequence of Theorem 3.1.2. ■

3.3 Convergence in the r^{th} mean

Consider rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$.

With $r \geq 1$, the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges to the rv X *in the r^{th} mean* if the rvs $\{X_n,\ n = 1, 2, \ldots\}$ satisfy

$$\mathbb{E}\left[|X_n|^r\right] < \infty, \quad n = 1, 2, \ldots \tag{3.8}$$

and

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0. \tag{3.9}$$

We shall write $X_n \xrightarrow{L^r}{}_n X$. The case $r = 2$ is often used in applications where it is referred as *mean-square* convergence. The case $r = 1$ also occurs with some regularity, and is referred as *mean* convergence.

It follows from (3.9) that $\mathbb{E}\left[|X_n - X|^r\right] < \infty$ for all n sufficiently large, whence the rv X necessarily has a finite moment of order r by virtue of Minkowski's inequality under (3.8).

Convergence in the r^{th} mean also admits a Cauchy criterion which is given next.

Theorem 3.3.1 *(Cauchy criterion for r^{th} mean convergence) With $r \geq 1$, the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in the r^{th} mean if and only if*

$$\lim_{n\to\infty} \left(\sup_{m\geq n} \mathbb{E}\left[|X_n - X_m|^r\right] \right) = 0. \tag{3.10}$$

Convergence in the r^{th} mean becomes more stringent as r increases. This is not surprising if we recall that for any rv $\xi : \Omega \to \mathbb{R}$, we have

$$\mathbb{E}\left[|\xi|^s\right] \leq 1 + \mathbb{E}\left[|\xi|^r\right], \quad 1 \leq s < r$$

as a result of the trivial identity $x^s \leq 1 + x^r$ for $x \geq 0$.

Theorem 3.3.2 *With $1 \leq s < r$, convergence in the r^{th} mean implies convergence in the s^{th}mean: If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in the r^{th} mean to the rv X, then the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ also converges in the s^{th} mean to the rv X.*

Proof. This is a simple consequence of Lyapounov's inequality

$$\mathbb{E}\left[|X_n - X|^s\right]^{\frac{1}{s}} \leq \mathbb{E}\left[|X_n - X|^r\right]^{\frac{1}{r}}, \quad n = 1, 2, \ldots$$

■

Next, we compare r^{th} mean convergence to convergence in probability.

Theorem 3.3.3 *Convergence in the r^{th} mean implies convergence in probability: If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in r^{th} mean to the rv X for some $r \geq 1$, then it also converges in probability to the rv X.*

Proof. Pick $\varepsilon > 0$ arbitrary. Markov's inequality yields

$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\varepsilon^r}, \quad n = 1, 2, \ldots$$

so that $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$ as soon as $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0$. ■

The converse is more delicate as the next example already illustrates; see also Section 3.5.

(Counter)examples

Consider a collection of rvs $\{X_n,\ n = 1, 2, \ldots\}$ such that

$$X_n = \begin{cases} 0 & \text{with probability } 1 - n^{-\alpha} \\ \\ n^{\beta} & \text{with probability } n^{-\alpha} \end{cases}$$

for each $n = 1, 2, \ldots$ where $\alpha > 0$ and $\beta > 0$. Thus,

$$\mathbb{P}[|X_n| > \varepsilon] = n^{-\alpha}, \quad n = 1, 2, \ldots$$

as soon as $0 < \varepsilon \leq 1$ so that $X_n \xrightarrow{P}{}_n 0$.

On the other hand, with $r \geq 1$, we find

$$\mathbb{E}[|X_n|^r] = 0\left(1 - n^{-\alpha}\right) + n^{r\beta} n^{-\alpha} = n^{r\beta - \alpha}, \quad n = 1, 2, \ldots$$

so that

$$\lim_{n\to\infty} \mathbb{E}[|X_n|^r] = \begin{cases} 0 & \text{if } r\beta < \alpha \\ 1 & \text{if } r\beta = \alpha \\ \infty & \text{if } r\beta > \alpha. \end{cases}$$

It is now plain that $X_n \xrightarrow{L^r}{}_n 0$ when $r\beta < \alpha$ but no such conclusion can be reached when $r\beta \geq \alpha$.

We close this section with a simple observation, based on Theorem 3.1.2, which allows us to determine a.s. convergence in the presence of convergence in the r^{th} mean.

Theorem 3.3.4 *If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in r^{th} mean to the rv X for some $r \geq 1$, then it also converges almost surely to the rv X whenever the condition*

$$\sum_{n=1}^{\infty} \mathbb{E}[|X_n - X|^r] < \infty$$

holds.

Proof. By Markov's inequality, we have

$$\mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\varepsilon^r}, \quad n = 1, 2, \ldots$$

for every $\varepsilon > 0$, whence

$$\sum_{n=1}^{\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] \leq \frac{1}{\varepsilon^r} \sum_{n=1}^{\infty} \mathbb{E}\left[|X_n - X|^r\right],$$

and the conclusion is immediate by Theorem 3.1.2. ■

3.4 Convergence in distribution

For any rv $X : \Omega \to \mathbb{R}$, its probability distribution function $F_X : \mathbb{R} \to [0, 1]$ satisfies the following properties: (i) it is non-decreasing; (ii) it has left-limit and is right-continuous at every point; and (iii) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

Let $\mathcal{C}(F_X)$ denote the set of points in $\mathbb{R}$ where $F_X : \mathbb{R} \to [0, 1]$ is continuous, i.e.,

$$\mathcal{C}(F_X) = \{x \in \mathbb{R} : \ F_X(x-) = F_X(x)\}.$$

The complement $\mathcal{C}(F_X)^c$ of $\mathcal{C}(F_X)$ in $\mathbb{R}$ consists of the points where $F_X : \mathbb{R} \to [0, 1]$ is not continuous.

Theorem 3.4.1 *For any rv $X : \Omega \to \mathbb{R}$, its probability distribution function $F_X : \mathbb{R} \to [0, 1]$ has the property that $\mathcal{C}(F_X)$ is a countable subset of $\mathbb{R}$.*

Proof. For each $n = 1, 2, \ldots$, let $\mathcal{D}_n$ denote the collection of points of discontinuity in $\mathcal{C}(F_X)^c$ whose discontinuity jump lies in the interval $(\frac{1}{n+1}, \frac{1}{n}]$, i.e.,

$$\mathcal{D}_n \equiv \left\{ x \in \mathcal{C}(F_X)^c : \ \frac{1}{n+1} < F_X(x) - F_X(x-) \leq \frac{1}{n} \right\}$$

Noting that

$$|\mathcal{D}_n| \cdot \frac{1}{n+1} \leq \sum_{x \in \mathcal{D}_n} \left(F_X(x) - F_X(x-)\right) \leq 1,$$

it follows that $|\mathcal{D}_n| \leq n+1$. The desired result is now immediate since $\mathcal{C}(F_X)^c = \cup_{n=1}^{\infty} \mathcal{D}_n$. ■

The sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ *converges in distribution* to the rv X if

(3.11) $$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad x \in \mathcal{C}(F_X)..$$

We shall write $X_n \Longrightarrow_n X$ or $X_n \xrightarrow{\mathcal{L}}{}_n X$. Some authors refer to this mode of convergence as *convergence in law* or as *weak convergence*.

As this mode of convergence involves *only* the probability distribution functions, it is sometimes convenient to define this notion without any reference to the rvs (viewed as mappings): The sequence of probability distribution functions $\{F_n,\ n = 1, 2, \ldots\}$ *converges in distribution* to the probability distribution function F if

(3.12) $$\lim_{n\to\infty} F_n(x) = F(x), \quad x \in \mathcal{C}(F_X)..$$

We shall write $F_n \Longrightarrow_n F$ or $F_n \xrightarrow{\mathcal{L}}{}_n F$.

At this point the reader may wonder as to why the definition of distributional convergence requires the convergence (3.11) only on the set of points of continuity of the limit. This is best seen on the following example.

The importance of discontinuity points

Consider the two sequences of rvs $\{X_n,\ n = 1, 2, \ldots\}$ and $\{X'_n,\ n = 1, 2, \ldots\}$ given by

$$X_n = -\frac{1}{n} \quad \text{and} \quad X'_n = \frac{1}{n}, \quad n = 1, 2, \ldots$$

defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Both sequences converge as *deterministic* sequences with $\lim_{n\to\infty} X_n(\omega) = 0$ and $\lim_{n\to\infty} Y_n(\omega) = 0$ for every ω in Ω. Yet it is easy to check that

$$\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \quad \text{and} \quad \lim_{n\to\infty} F_{X'_n}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Theorem 3.4.2 *Convergence in probability implies convergence in distribution: If the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in probability to the rv X, then it also converges in distribution.*

Proof. Fix $n = 1, 2, \ldots$ and pick x in $\mathbb{R}$. With $\varepsilon > 0$, we note that

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}[X_n \leq x] \\ &= \mathbb{P}[X_n \leq x, X \leq x+\varepsilon] + \mathbb{P}[X_n \leq x, x+\varepsilon < X] \\ &\leq \mathbb{P}[X \leq x+\varepsilon] + \mathbb{P}[|X_n - X| > \varepsilon] \\ &= F_X(x+\varepsilon) + \mathbb{P}[|X_n - X| > \varepsilon] \qquad (3.13) \end{aligned}$$

In a similar way, we find

$$\begin{aligned} F_X(x-\varepsilon) &= \mathbb{P}[X \leq x-\varepsilon] \\ &= \mathbb{P}[X \leq x-\varepsilon, X_n \leq x] + \mathbb{P}[X \leq x-\varepsilon, x < X_n] \\ &\leq \mathbb{P}[X_n \leq x] + \mathbb{P}[|X_n - X| > \varepsilon] \\ &= F_{X_n}(x) + \mathbb{P}[|X_n - X| > \varepsilon] \qquad (3.14) \end{aligned}$$

Let n go to infinity in these inequalities. Under the assumed convergence in probability, we find

$$\limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x+\varepsilon)$$

and

$$F_X(x-\varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x)$$

Picking x to be a point of continuity for F_X, we get

$$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = \lim_{\varepsilon\downarrow 0} F_X(x-\varepsilon) = F_X(x).$$

Therefore,

$$\begin{aligned} \limsup_{n\to\infty} F_{X_n}(x) &= \lim_{\varepsilon\downarrow 0}\left(\limsup_{n\to\infty} F_{X_n}(x)\right) \\ &\leq \lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) \\ &= F_X(x) \end{aligned}$$

and

$$\begin{aligned} F_X(x) &= \lim_{\varepsilon \downarrow 0} F_X(x - \varepsilon) \\ &\leq \lim_{\varepsilon \downarrow 0} \left(\liminf_{n \to \infty} F_{X_n}(x) \right) \\ &= \liminf_{n \to \infty} F_{X_n}(x) \end{aligned}$$

whence $\liminf_{n\to\infty} F_{X_n}(x) = \limsup_{n\to\infty} F_{X_n}(x) = F_X(x)$. It follows that

$$\liminf_{n \to \infty} F_{X_n}(x) = F_X(x), \quad \mathcal{C}(F_X).$$

■

Although weak convergence is weaker than convergence in probability, there is one situation where they are equivalent.

Theorem 3.4.3 *With c a scalar in $\mathbb{R}$, the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in probability to the degenerate rv $X = c$ if and only if the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in distribution to the degenerate rv $X = c$.*

Proof. Assume that the sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ converges in distribution to the degenerate rv $X = c$. Fix $\varepsilon > 0$. For every $n = 1, 2, \ldots$, we observe that

$$\begin{aligned} \mathbb{P}[|X_n - X| \leq \varepsilon] &= \mathbb{P}[|X_n - c| \leq \varepsilon] \\ &= \mathbb{P}[c - \varepsilon \leq X_n \leq c + \varepsilon] \\ &= \mathbb{P}[X_n \leq c + \varepsilon] - \mathbb{P}[X_n < c - \varepsilon] \\ &= F_{X_n}(c + \varepsilon) - F_{X_n}((c - \varepsilon)-) \qquad (3.15) \end{aligned}$$

so that

$$\begin{aligned} \mathbb{P}[|X_n - X| > \varepsilon] &= 1 - F_{X_n}(c + \varepsilon) + F_{X_n}((c - \varepsilon)-) \\ &\leq 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon). \qquad (3.16) \end{aligned}$$

Recall that $F_X(x) = 0$ (resp. $F_X(x) = 1$) if $x < c$ (resp. $c \leq x$) so that the only point of discontinuity of F_X is located at $x = c$. Thus, under the assumed convergence in distribution, we have $\lim_{n\to\infty} F_{X_n}(c + \varepsilon) = 1$ and $\lim_{n\to\infty} F_{X_n}(c - \varepsilon) = 0$, whence $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$ as desired. ■

3.5 Uniform integrability

If a rv X has finite first moment, we know that

$$\lim_{B \to \infty} \mathbb{E}\left[\mathbf{1}\left[|X| > B\right] |X|\right] = 0. \tag{3.17}$$

This is a simple consequence of the Dominated Convergence Theorem (since $Y_B \leq |X|$ where $Y_B = \mathbf{1}\left[|X| > B\right] |X|$ for all $B > 0$). Thus, for every $\varepsilon > 0$, there exists $B^\star(\varepsilon) > 0$ such that

$$\mathbb{E}\left[\mathbf{1}\left[|X| > B\right] |X|\right] \leq \varepsilon, \quad B \geq B^\star(\varepsilon). \tag{3.18}$$

As we consider a collection of rvs $\{X_n,\ n = 1, 2, \ldots\}$ with finite first moments, we can certainly assert the following: For each $n = 1, 2, \ldots$ and every $\varepsilon > 0$, there exists $B^\star(\varepsilon; n) > 0$ such that

$$\mathbb{E}\left[\mathbf{1}\left[|X_n| > B\right] |X_n|\right] \leq \varepsilon, \quad B \geq B^\star(\varepsilon; n). \tag{3.19}$$

This is a direct consequence of (3.18). However, sometimes it is required that this condition holds *uniformly* with respect to $n = 1, 2, \ldots$ in that $B^\star(\varepsilon; n)$ can be selected independently of n. This leads to the following stronger notion of integrability for a sequence of rvs.

The collection of rvs $\{X_n,\ n = 1, 2, \ldots\}$ is said to be *uniformly integrable* if

$$\lim_{B \to \infty} \left(\sup_{n=1,2,\ldots} \mathbb{E}\left[\mathbf{1}\left[|X_n| > B\right] |X_n|\right] \right) = 0. \tag{3.20}$$

In other words, for every $\varepsilon > 0$, there exists $B^\star(\varepsilon) > 0$ such that

$$\sup_{n=1,2,\ldots} \mathbb{E}\left[\mathbf{1}\left[|X_n| > B\right] |X_n|\right] \leq \varepsilon, \quad B \geq B^\star(\varepsilon). \tag{3.21}$$

Interest in this notion arises from the need to have an easy characterization of situations where interchange between limits and expectation can take place.

Theorem 3.5.1 *Consider a collection of rvs* $\{X, X_n,\ n = 1, 2, \ldots\}$ *such that* $\lim_{n \to \infty} X_n = X$ *a.s. If the collection of rvs* $\{X_n,\ n = 1, 2, \ldots\}$ *is uniformly integrable, then* $\mathbb{E}\left[|X|\right] < \infty$ *and*

$$\lim_{n \to \infty} \mathbb{E}\left[X_n\right] = \mathbb{E}\left[X\right] \tag{3.22}$$

3.6 Weak convergence via characteristic functions

Weak convergence of a sequence of rvs can be characterized through the limiting behavior of the corresponding sequence of characteristic functions.

Theorem 3.6.1 *The sequence of rvs* $\{X_n, n = 1, 2, \ldots\}$ *converges weakly to the rv* X *if and only if*

$$\lim_{n\to\infty} \Phi_{X_n}(\theta) = \Phi_X(\theta), \quad \theta \in \mathbb{R}.$$

This result suggests the following strategy: Consider the limit

$$\Phi(\theta) = \lim_{n\to\infty} \Phi_{X_n}(\theta), \quad \theta \in \mathbb{R} \tag{3.23}$$

and identify the rv X whose characteristic function coincides with $\Phi : \mathbb{R} \to \mathbb{C}$. However, a word of caution is in order as the limit (3.23) may not necessarily define the characteristic function of a rv as can be seen from the following example.

The limit of characteristic functions is not always a characteristic function — For each $n = 1, 2, \ldots$, the rv X_n is the uniform rv on the interval $(-n, n)$. Easy calculations show that

$$\Phi_{X_n}(\theta) = \int_{-n}^{n} \frac{e^{i\theta x}}{2n} dx = \begin{cases} \frac{\sin(n\theta)}{n} & \text{if } \theta \neq 0 \\ \\ 1 & \text{if } \theta = 0, \end{cases} \tag{3.24}$$

so that

$$\Phi(\theta) = \lim_{n\to\infty} \Phi_{X_n}(\theta) = \begin{cases} 0 & \text{if } \theta \neq 0 \\ \\ 1 & \text{if } \theta = 0. \end{cases}$$

Obviously, there are no rv X whose characteristic function coincides with the limit.

This difficulty can be remedied with the help of the next result by simply checking *continuity* at $\theta = 0$ for the limit (3.23). This is a consequence of the Bochner-Herglotz Theorem.

Theorem 3.6.2 *Consider a sequence of rvs* $\{X_n, n = 1, 2, \ldots\}$ *such that the limits*

$$\Phi(\theta) = \lim_{n\to\infty} \Phi_{X_n}(\theta), \quad \theta \in \mathbb{R}$$

all exist. If $\Phi : \mathbb{R} \to \mathbb{C}$ *is continuous at* $\theta = 0$*, then it is the characteristic function of some rv* X*, and* $X_n \Longrightarrow_n X$*.*

Proof. For each $n = 1, 2, \ldots$, the function $\Phi_{X_n} : \mathbb{R} \to \mathbb{C}$ is a characteristic function. Therefore, by Theorem 2.3.1 it is (i) bounded with $|\Phi_{X_n}(\theta)| \leq \Phi_{X_n}(0) = 1$ for all θ in $\mathbb{R}$; (ii) uniformly continuous on $\mathbb{R}$; and (iii) positive semi-definite. Properties (i) and (iii) are clearly inherited by the limit $\Phi : \mathbb{R} \to \mathbb{C}$. Therefore, by Theorem 2.3.2 the assumed continuity of Φ implies that it is a characteristic function, i.e., there exists a rv X such that $\Phi = \Phi_X$. Invoking Theorem 3.6.1 we conclude that $X_n \Longrightarrow_n X$. ■

3.7 Weak convergence via the Skorokhod representation

Consider a collection $\{F, F_n,\ n = 1, 2, \ldots\}$ of probability distribution functions on $\mathbb{R}$.

Theorem 3.7.1 *If the sequence of probability distribution functions* $\{F_n,\ n = 1, 2, \ldots\}$ *converges weakly to* F*, then there exists a probability triple* $(\Omega^\star, \mathcal{F}^\star, \mathbb{P}^\star)$ *and a collection of* $\mathbb{R}$*-valued rvs* $\{X^\star, X_n^\star,\ n = 1, 2, \ldots\}$ *all defined on* $\Omega^\star$ *with the following properties:*

(i) We have

$$F(x) = \mathbb{P}^\star[X^\star \leq x], \quad x \in \mathbb{R} \tag{3.25}$$

and

$$F_n(x) = \mathbb{P}^\star[X_n^\star \leq x] \quad \begin{array}{l} x \in \mathbb{R} \\ n = 1, 2, \ldots \end{array} \tag{3.26}$$

(ii) The rvs $\{X_n^\star,\ n = 1, 2, \ldots\}$ *converges a.s. to* $X^\star$ *(under* $\mathbb{P}^\star$*), i.e.,*

$$\mathbb{P}^\star\left[\left\{\omega^\star \in \Omega^\star :\ \lim_{n\to\infty} X_n^\star(\omega^\star) = X^\star(\omega^\star)\right\}\right] = 1$$

3.8 Functional characterization of convergence in distribution

The following equivalent characterizations of distributional convergence have many use.

Theorem 3.8.1 *Consider the* $\mathbb{R}$*-valued rvs* $\{X, X_n,\ n = 1, 2, \ldots\}$ *defined on some probability triple* $(\Omega, \mathcal{F}, \mathbb{P})$*. The following three statements are equivalent:*

(i) The rvs $\{X_n,\ n = 1, 2, \ldots\}$ *converge in distribution to the rv* X*, i.e.,*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X x), \quad x \in \mathcal{C}(F_X).$$

(ii) For every bounded continuous mapping $g : \mathbb{R} \to \mathbb{R}$*, it holds that*

$$\lim_{n\to\infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right]. \tag{3.27}$$

(iii) The characteristic functions converge in the sense that

$$\lim_{n\to\infty} \Phi_{X_n}(\theta) = \Phi_X(\theta), \quad \theta \in \mathbb{R}. \tag{3.28}$$

Proof. It follows from Theorem 3.7.1 that (i) implies the validity of (ii): Indeed, with the notation used in that result, consider the probability triple $(\Omega^\star, \mathcal{F}^\star, \mathbb{P}^\star)$ and the $\mathbb{R}$-valued rvs $\{X^\star, X_n^\star,\ n = 1, 2, \ldots\}$ all defined on $\Omega^\star$ such that

$$\mathbb{P}\left[X \leq x\right] = \mathbb{P}^\star[X^\star \leq x], \quad x \in \mathbb{R} \tag{3.29}$$

and

$$\mathbb{P}\left[X_n \leq x\right] = \mathbb{P}^\star[X_n^\star \leq x] \quad \begin{array}{l} x \in \mathbb{R} \\ n = 1, 2, \ldots \end{array} \tag{3.30}$$

with

$$\mathbb{P}^\star\left[\omega^\star \in \Omega^\star :\ \lim_{n\to\infty} X_n^\star(\omega^\star) = X^\star(\omega^\star)\right] = 1.$$

Pick a mapping $g : \mathbb{R} \to \mathbb{R}$ which is continuous and bounded - Set

$$B_g \equiv \sup_{x\in\mathbb{R}} |g(x)| < \infty.$$

Obviously,

$$\mathbb{E}\left[g(X)\right] = \mathbb{E}^\star\left[g^\star(X^\star)\right] \quad \text{and} \quad \mathbb{E}\left[g(X_n)\right] = \mathbb{E}^\star\left[g^\star(X_n^\star)\right], \quad n = 1, 2, \ldots$$

It is plain that

$$\lim_{n\to\infty} g(X_n^\star) = g(X^\star) \quad \mathbb{P}^\star\text{-a.s.}$$

by the continuity of g, with

$$|g(X_n^\star(\omega^\star)| \leq B_g, \quad \begin{array}{c}\omega^\star \in \Omega^\star \\ n = 1, 2, \ldots\end{array}$$

Invoking the Dominated Convergence Theorem we readily conclude that

$$\lim_{n\to\infty} \mathbb{E}^\star\left[g^\star(X_n^\star)\right] = \mathbb{E}^\star\left[g^\star(X^\star)\right].$$

This completes the proof of the validity of (ii). The proof that (ii) implies (i) is omitted.

The equivalence of (i) and (iii) is just Theorem **??**. Note that (iii) is a simple consequence of (ii) since for every θ in $\mathbb{R}$ the mappings $x \to \cos(\theta x)$ and $x \to \sin(\theta x)$ are bounded and continuous on $\mathbb{R}$. ∎

An immediate consequence of Theorem is the following continuity result for weak convergence.

Theorem 3.8.2 *Consider the $\mathbb{R}$-valued rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. If the rvs $\{X_n,\ n = 1, 2, \ldots\}$ converge in distribution to the rv X, then the $\mathbb{R}$-valued rvs $\{h(X_n),\ n = 1, 2, \ldots\}$ converge in distribution to the rv $h(X)$ for any continuous mapping $h : \mathbb{R} \to \mathbb{R}$, namely*

$$h(X_n) \Longrightarrow_n h(X).$$

Proof. The proof follows by a simple application of Theorem 3.10: Pick a bounded continuous mapping $g : \mathbb{R} \to \mathbb{R}$. Given the continuous mapping $h : \mathbb{R} \to \mathbb{R}$, we note that the mapping $g \circ h : \mathbb{R} \to \mathbb{R}$ given by

$$g \circ h(x) = g(h(x)), \quad x \in \mathbb{R}$$

is also a bounded continuous mapping $\mathbb{R} \to \mathbb{R}$. Therefore, by Part (ii) of Theorem 3.10 we conclude from the assumed convergence $X_n \Longrightarrow_n X$ that

$$\lim_{n\to\infty} \mathbb{E}\left[g \circ h(X_n)\right] = \mathbb{E}\left[g \circ h(X)\right].$$

or equivalently,

$$\lim_{n\to\infty} \mathbb{E}\left[g(h(X_n))\right] = \mathbb{E}\left[g(h(X))\right].$$

Invoking one more time Part (ii) of Theorem 3.10 we now conclude that $h(X_n) \Longrightarrow_n h(X)$ as desired. ■

3.9 Weak convergence of discrete rvs

In this section we consider a collection of *discrete* rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ with

$$\mathbb{P}\left[X \in S\right] = \mathbb{P}\left[X_n \in S\right] = 1, \quad n = 1, 2, \ldots$$

where $S = \{a_i, i \in I\}$ is a countable subset of $\mathbb{Z}$.

Theorem 3.9.1 *The sequence of discrete rvs $X_n \Longrightarrow_n X$ converges weakly to the rv X if and only if*

$$\lim_{n\to\infty} \mathbb{P}\left[X_n = a_i\right] = \mathbb{P}\left[X = a_i\right], \quad i \in I.$$

Proof. Assume first that $X_n \Longrightarrow_n X$. Pick a a point of disconitnuity for F_X. By assumption a is an element of $\mathbb{Z}$, so that $\varepsilon > 0$ can be selected so that both $a \pm \varepsilon$ are not in $\mathbb{Z}$, whence

$$\lim_{n\to\infty} \mathbb{P}\left[X_n \leq a \pm \varepsilon\right] = \mathbb{P}\left[X \leq a \pm \varepsilon\right]. \tag{3.31}$$

Note however that

$$\mathbb{P}\left[X_n \leq a - \varepsilon\right] = \mathbb{P}\left[X_n \leq a + \varepsilon\right] + \mathbb{P}\left[X_n = a\right], \quad n = 1, 2, \ldots \tag{3.32}$$

and

$$\mathbb{P}\left[X \leq a - \varepsilon\right] = \mathbb{P}\left[X \leq a + \varepsilon\right] + \mathbb{P}\left[X = a\right]. \tag{3.33}$$

since the probability distribution functions are piecewise constant with jumps only at points in $\mathbb{Z}$.

Let n go to infinity in (3.32). It is plain from (3.31) that $\lim_{n\to\infty} \mathbb{P}[X_n = a]$ exists and is given by

$$\lim_{n\to\infty} \mathbb{P}[X_n = a] = \mathbb{P}[X \leq a + \varepsilon] - \mathbb{P}[X \leq a - \varepsilon] = \mathbb{P}[X = a]$$

where the last equality follows from (3.33).

Conversely, assume that

$$\lim_{n\to\infty} \mathbb{P}[X_n = a] = \mathbb{P}[X = a], \quad a \notin \mathcal{C}(F_X) \tag{3.34}$$

With a Borel subset B in $\mathbb{R}$, we shall show that

$$\lim_{n\to\infty} \mathbb{P}[X_n \in B] = \mathbb{P}[X \in B]. \tag{3.35}$$

This will immediately imply $X_n \Longrightarrow_n X$ upon specializing B to sets of the form $B = (-\infty, x]$ with x in $\mathcal{C}(F_X)$. To do so, fix $n = 1, 2, \ldots$ and pick A an arbitrary positive integer A:

We see that

$$\begin{aligned} &\mathbb{P}[X_n \in B] \\ &= \mathbb{P}[|X_n| \leq A, X_n \in B] + \mathbb{P}[|X_n| > A, X_n \in B] \\ &= \sum\nolimits_{a \in \mathbb{Z} \cap B : |a| \leq A} \mathbb{P}[X_n = a] + \mathbb{P}[|X_n| > A, X_n \in B] \end{aligned} \tag{3.36}$$

while

$$\begin{aligned} &\mathbb{P}[X \in B] \\ &= \mathbb{P}[|X| \leq A, X \in B] + \mathbb{P}[|X| > A, X \in B] \\ &= \sum\nolimits_{a \in \mathbb{Z} \cap B : |a| \leq A} \mathbb{P}[X = a] + \mathbb{P}[|X| > A, X \in B]. \end{aligned} \tag{3.37}$$

Substracting we conclude that

$$\begin{aligned} &|\mathbb{P}[X_n \in B] - \mathbb{P}[X \in B]| \\ &\leq \sum\nolimits_{a \in \mathbb{Z} \cap B : |a| \leq A} |\mathbb{P}[X_n = a] - \mathbb{P}[X = a]| + \mathbb{P}[|X_n| > A] + \mathbb{P}[|X| > A]. \end{aligned}$$

Let n go to infinity in this last inequality: Using (3.34) we get

$$\lim_{n\to\infty} \sum\nolimits_{a \in \mathbb{Z} \cap B : |a| \leq A} |\mathbb{P}[X_n = a] - \mathbb{P}[X = a]| = 0$$

since this sum has at most $2A + 1$ terms, while

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left[|X_n| > A\right] &= \lim_{n\to\infty} \left(1 - \mathbb{P}\left[|X_n| \le A\right]\right) \\ &= 1 - \mathbb{P}\left[|X| \le A\right] = \mathbb{P}\left[|X| \le A\right] \end{aligned} \tag{3.38}$$

by a similar argument. Collecting these facts we obtain

$$\limsup_{n\to\infty} \left|\mathbb{P}\left[X_n \in B\right] - \mathbb{P}\left[X \in B\right]\right| \le 2\mathbb{P}\left[|X| > A\right].$$

Now let A go to infinity in this last inequality and note that

$$\lim_{A\to\infty} \left(\limsup_{n\to\infty} \left|\mathbb{P}\left[X_n \in B\right] - \mathbb{P}\left[X \in B\right]\right|\right) = 0$$

and the desired conclusion (3.35) follows since the left handside does not depend on A ■

In the more restrictive setting where $S \subseteq \mathbb{N}$, probability generating functions can be defined, and the following analog of Theorem 3.6.1 holds.

Theorem 3.9.2 *The sequence of $\mathbb{N}$-valued rvs $\{X_n, n = 1, 2, \ldots\}$ converges weakly to the rv X if and only if*

$$\lim_{n\to\infty} G_{X_n}(z) = G_X(z), \quad |z| \le 1.$$

3.10 Convergence in higher dimensions

The discussion so far has been in the context of $\mathbb{R}$-valued rvs. We now outline the corresponding theory for $\mathbb{R}^p$-valued rvs with $p \ge 1$. The first observation is that the three first modes of convergence, namely a.s. convergence, convergence in probability and convergence in the r^{th} mean are "metric" notions in the following sense: The rvs $\{X_n,\ n = 1, 2, \ldots\}$

- converge a.s. to the rv X if

$$\lim_{n\to\infty} |X_n - X| = 0 \quad a.s.$$

- converge in probability to the rv X if

$$\lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] = 0, \quad \varepsilon > 0$$

- converge in the r^{th} mean (for some $r \geq 1$) to the rv X if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

They are all expressed in terms of the *distance* $|X_n - X|$ of X_n to X.

In $\mathbb{R}^p$ there are a number of ways to define the distance between two vectors. Here we limit ourselves to metrics that are induced by norms, so that distance is measured by

$$d(x, y) = \|x - y\|, \quad x, y \in \mathbb{R}^p$$

where $\|\cdot\| : \mathbb{R}^p \to \mathbb{R}_+$ is a norm. Therefore, a natural to define the modes of convergence for $\mathbb{R}^p$-valued rvs as follows:

Consider any norm $\|\cdot\| : \mathbb{R}^p \to \mathbb{R}_+$. The $\mathbb{R}^p$-valued rvs $\{X_n,\ n = 1, 2, \ldots\}$

- converge a.s. to the rv X if

$$\lim_{n\to\infty} \|X_n - X\| = 0 \quad a.s.$$

- converge in probability to the rv X if

$$\lim_{n\to\infty} \mathbb{P}\left[\|X_n - X\| > \varepsilon\right] = 0, \quad \varepsilon > 0$$

- converge in the r^{th} mean (for some $r \geq 1$) to the rv X if

$$\lim_{n\to\infty} \mathbb{E}\left[\|X_n - X\|^r\right] = 0.$$

Note that all norms on $\mathbb{R}^p$ are equivalent in the following sense: If $\|\cdot\|_a : \mathbb{R}^p \to \mathbb{R}_+$ and $\|\cdot\|_b : \mathbb{R}^p \to \mathbb{R}_+$ are two different norms, then there exists constants $c_{a|b} > 0$ and $C_{a|b} >$ such that

$$c_{a|b}\|x\|_a \leq \|x\|_b \leq C_{a|b}\|x\|_a, \quad x \in \mathbb{R}^p.$$

Norms often used in applications include

- The Euclidean norm (or L_1-norm):

$$\|x\|_2 = \sqrt{\sum_{k=1}^{p} |x_k|^2}, \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p$$

- The L_1-norm:

$$\|x\|_1 = \sum_{k=1}^{p} |x_k|, \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p$$

- The Manhattan norm

$$\|x\|_\infty = \max(|x_k|,\ k = 1, \ldots, p), \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p$$

However when it comes to convergence in distribution matters are quite different because this notion does not rely on a notion of proximity in the range of the rvs under consideration. Furthermore, probability distribution functions on $\mathbb{R}^p$ are more cumbersome to characterize. So instead of using the definition given in Section 3.4 we instead rely on the equivalence given in Theorem 3.10

The sequence of $\mathbb{R}^p$-valued rvs $\{X_n,\ n = 1, 2, \ldots\}$ *converges in distribution* to the $\mathbb{R}^p$-valued rv X if for every bounded continuous mapping $g : \mathbb{R}^p \to \mathbb{R}$, it holds that

$$\lim_{n\to\infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right]. \tag{3.39}$$

Here as well we shall write $X_n \Longrightarrow_n X$ or $X_n \xrightarrow{\mathcal{L}}{}_n X$. Some authors also refer to this mode of convergence as *convergence in law* or as *weak convergence*.

Theorem 3.10 has the following multi-dimensional analog.

Theorem 3.10.1 *Consider the $\mathbb{R}^p$-valued rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the rvs $\{X_n,\ n = 1, 2, \ldots\}$ converge in distribution to the rv X if and only if*

$$\lim_{n\to\infty} \Phi_{X_n}(\theta) = \Phi_X(\theta), \quad \theta \in \mathbb{R}. \tag{3.40}$$

This amounts to

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\theta' X_n}\right] = \mathbb{E}\left[e^{i\theta' X}\right], \quad \theta \in \mathbb{R}.$$

In the same way that Theorem implied Theorem 3.8.2, we readily see that Theorem 3.10.1has the following important consequence.

Theorem 3.10.2 *Consider the $\mathbb{R}^p$-valued rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. If the rvs $\{X_n,\ n = 1, 2, \ldots\}$ converge in distribution to the rv X, then the $\mathbb{R}^q$-valued rvs $\{h(X_n),\ n = 1, 2, \ldots\}$ converge in distribution to the $\mathbb{R}^q$-valued rv $h(X)$ for any continuous mapping $h : \mathbb{R}^p \to \mathbb{R}^q$, namely*

$$h(X_n) \Longrightarrow_n h(X).$$

Chapter 4

The classical limit theorems

The setting of the next four sections is as follows: The rvs $\{X_n,\ n = 1, 2, \ldots\}$ are rvs defined on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. With this sequence we associate the sums

$$S_n = \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots$$

Two types of results will be discussed: The first class of results are known as Laws of Large Numbers; they deal with the convergence of the sample averages

$$\bar{S}_n = \frac{1}{n} \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots$$

The second class of results are called Central Limit Theorems and provide a rate of convergence in the Laws Large Numbers.

4.1 Weak Laws of Large Numbers (I)

Laws of Large Numbers come in two types which are distinguished by the mode of convergence used. When convergence in probability is used, we refer to such results as weak Laws of Large Numbers. The most basic such results is given first.

Theorem 4.1.1 *Assume the rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ to be i.i.d. rvs with $\mathbb{E}\left[|X|^2\right] < \infty$. Then,*

$$\frac{S_n}{n} \xrightarrow{L^2}{}_n \mathbb{E}\left[X\right], \qquad (4.1)$$

whence

$$\frac{S_n}{n} \xrightarrow{P}{}_n \mathbb{E}[X]. \tag{4.2}$$

Proof. For each $n = 1, 2, \ldots$, we note that

$$\begin{aligned} \mathbb{E}\left[\left|\frac{S_n}{n} - \mathbb{E}[X]\right|^2\right] &= \mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X])\right|^2\right] \\ &= \frac{1}{n^2} \cdot \mathrm{Var}[S_n] \end{aligned} \tag{4.3}$$

with

$$\begin{aligned} \mathrm{Var}[S_n] &= \sum_{k=1}^{n}\sum_{\ell=1}^{n} \mathrm{Cov}[X_k, X_\ell] \\ &= \sum_{k=1}^{n} \mathrm{Var}[X_k] \\ &= n\mathrm{Var}[X] \end{aligned} \tag{4.4}$$

since

$$\mathrm{Cov}[X_k, X_\ell] = \delta(k;\ell)\mathrm{Var}[X], \quad k, \ell = 1, \ldots, n$$

under the enforced independence assumptions.

As a result,

$$\mathbb{E}\left[\left|\frac{S_n}{n} - \mathbb{E}[X]\right|^2\right] = \frac{n\mathrm{Var}[X]}{n^2} = \frac{\mathrm{Var}[X]}{n^2}$$

and the desired conclusions follow. ■

4.2 Weak Laws of Large Numbers (II)

A careful inspection of the proof of Theorem 4.1.1 suggests a more general result. Assume that the rvs $\{X_n,\ n = 1, 2, \ldots\}$ are second-order rvs. For each $n =$

$1, 2, \ldots$, we note that

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k])\right|^2\right] = \frac{\mathrm{Var}[S_n]}{n^2}.$$

By computations similar to the ones used in the proof of Theorem 4.1.1, we find

$$\begin{aligned} \mathrm{Var}[S_n] &= \mathrm{Var}[\sum_{k=1}^{n} X_k] \\ &= \sum_{k=1}^{n}\sum_{\ell=1}^{n} \mathrm{Cov}[X_k, X_\ell] \\ &= \sum_{k=1}^{n} \mathrm{Var}[X_k] + \sum_{k,\ell=1,\ k\neq\ell} \mathrm{Cov}[X_k, X_\ell], \end{aligned} \tag{4.5}$$

whence

$$\begin{aligned} &\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k])\right|^2\right] \\ &= \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}[X_k] + \frac{1}{n^2}\sum_{k,\ell=1,\ k\neq\ell} \mathrm{Cov}[X_k, X_\ell]. \end{aligned} \tag{4.6}$$

Theorem 4.2.1 *Consider a collection* $\{X_n,\ n = 1, 2, \ldots\}$ *of second-order rvs such that*

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}[X_k] = 0. \tag{4.7}$$

We have

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]) \xrightarrow{L^2}{}_n 0 \tag{4.8}$$

and

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k]) \xrightarrow{P}{}_n \tag{4.9}$$

whenever either one of the following conditions holds:

(i) The rvs $\{X_n,\ n = 1, 2, \ldots\}$ *are uncorrelated*

(ii) The rvs $\{X_n,\ n = 1, 2, \ldots\}$ *are negatively correlated, i.e.,*

$$\mathrm{Cov}[X_k, X_\ell] \leq 0, \qquad \begin{matrix} k \neq \ell \\ k, \ell = 1, \ldots, n \end{matrix}$$

(iii) The rvs $\{X_n,\ n = 1, 2, \ldots\}$ *satisfy the condition*

$$\lim_{n \to \infty} \frac{1}{n^2} \sum_{k,\ell=1,\ k \neq \ell} \mathrm{Cov}[X_k, X_\ell] = 0. \tag{4.10}$$

This result is often applied when the rvs $\{X_n,\ n = 1, 2, \ldots\}$ have identical means and variances, namely there exist μ and $\sigma^2 > 0$ such that

$$\mathbb{E}\left[X_n\right] = \mu \quad \text{and} \quad \mathrm{Var}[X_n] = \sigma^2, \quad n = 1, 2, \ldots$$

In that case, condition (4.7) is automatically satisfied and the convergence statements take the simpler form

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{L^2}{}_n \mu \tag{4.11}$$

and

$$\frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow{P}{}_n \mu \tag{4.12}$$

4.3 The classical Weak Law of Large Numbers (III)

As we now show, the finiteness of the second moment of X can be dropped.

Theorem 4.3.1 *Assume the rvs* $\{X, X_n,\ n = 1, 2, \ldots\}$ *to be i.i.d. rvs with* $\mathbb{E}\left[|X|\right] < \infty$*. Then, we have*

$$\frac{S_n}{n} \xrightarrow{P}{}_n \mathbb{E}\left[X\right]. \tag{4.13}$$

Proof. Fix $n = 1, 2, \ldots$ and θ in $\mathbb{R}$. Note that

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta\left(\frac{S_n}{n}-\mathbb{E}[X]\right)}\right] &= \mathbb{E}\left[e^{i\frac{\theta}{n}\sum_{k=1}^{n}(X_k-\mathbb{E}[X])}\right] \\
&= \mathbb{E}\left[\prod_{k=1}^{n} e^{i\frac{\theta}{n}(X_k-\mathbb{E}[X])}\right] \\
&= \prod_{k=1}^{n} \mathbb{E}\left[e^{i\frac{\theta}{n}(X_k-\mathbb{E}[X])}\right] \\
&= \left(\mathbb{E}\left[e^{i\frac{\theta}{n}(X-\mathbb{E}[X])}\right]\right)^n \qquad (4.14)
\end{aligned}$$

so that

$$\mathbb{E}\left[e^{i\theta\left(\frac{S_n}{n}-\mathbb{E}[X]\right)}\right] = \left(\mathbb{E}\left[e^{i\frac{\theta}{n}(X-\mathbb{E}[X])}\right]\right)^n .$$

As pointed out in Section 2.4, using Theorem 2.4.1 (for $k = 1$ and $x = X - \mathbb{E}[X]$), we get

$$e^{i\theta(X-\mathbb{E}[X])} = 1 + i\theta(X - \mathbb{E}[X]) + i\theta \int_0^{X-\mathbb{E}[X]} \left(e^{i\theta t} - 1\right) dt,$$

whence

$$\mathbb{E}\left[e^{i\theta(X-\mathbb{E}[X])}\right] = 1 + i\theta\mathbb{E}\left[\int_0^{X-\mathbb{E}[X]} \left(e^{i\theta t} - 1\right) dt\right]$$

upon taking expectations. Substituting θ by $\frac{\theta}{n}$, we obtain the relation

$$\mathbb{E}\left[e^{i\frac{\theta}{n}(X-\mathbb{E}[X])}\right] = 1 + \frac{i\theta}{n} \cdot C_1\left(\frac{\theta}{n}\right)$$

where

$$C_1(\theta) \equiv \mathbb{E}\left[\int_0^{X-\mathbb{E}[X]} \left(e^{i\theta t} - 1\right) dt\right].$$

It follows that

$$\mathbb{E}\left[e^{i\theta\left(\frac{S_n}{n}-\mathbb{E}[X]\right)}\right] = \left(1 + \frac{i\theta}{n} \cdot C_1\left(\frac{\theta}{n}\right)\right)^n . \qquad (4.15)$$

By Dominated Convergence, we conclude that $\lim_{n\to\infty} C_1\left(\frac{\theta}{n}\right) = 0$, whence

$$\lim_{n\to\infty}\left(\mathbb{E}\left[e^{i\frac{\theta}{n}(X-\mathbb{E}[X])}\right]\right)^n = \lim_{n\to\infty}\left(1 + \frac{i\theta}{n} \cdot C_1\left(\frac{\theta}{n}\right)\right)^n = 1.$$

It follows that $\frac{S_n}{n} - \mathbb{E}[X] \xrightarrow{P}{}_n 0$, and this conclude the proof of (4.13). ■

4.4 The Strong Law of Large Numbers

Strong Laws of Large Numbers are given are convergence statements in the a.s. sense. The classical Strong Law of Large Numbers is given next.

Theorem 4.4.1 *Assume the rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ to be i.i.d. rvs with $\mathbb{E}[|X|] < \infty$. Then,*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X] \quad a.s. \tag{4.16}$$

We give two proofs of this result under stronger assumptions on the moments of X. One proof assumes $\mathbb{E}[|X|^4] < \infty$ while the second proof is given under the condition $\mathbb{E}[|X|^2] < \infty$. A proof under the first moment condition $\mathbb{E}[|X|] < \infty$ is available in [].

Proof 1 Assume $\mathbb{E}[|X|^4] < \infty$ – Note that there is no loss in generality in assuming that $\mathbb{E}[X] = 0$ as we do from now on in this proof. The basic idea of the proof is as follows: By the Monotone Convergence Theorem it is always the case that

$$\mathbb{E}\left[\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right] = \sum_{n=1}^{\infty}\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right]$$

Therefore, if we could show that

$$\sum_{n=1}^{\infty}\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] < \infty, \tag{4.17}$$

we immediately conclude that

$$\mathbb{E}\left[\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4\right] < \infty$$

As a result,

$$\sum_{n=1}^{\infty}\left(\frac{S_n}{n}\right)^4 < \infty \quad a.s.$$

and the conclusion $\lim_{n\to\infty} \frac{S_n}{n} = 0$ a.s. is now straightforward.

In order to establish (4.17) we note that

$$\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] = \frac{1}{n^4} \cdot \mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^4\right]$$

with

$$\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^4\right] = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{\ell=1}^n \mathbb{E}\left[X_i X_j X_k X_\ell\right]. \tag{4.18}$$

Under the enforced independence assumptions it is plain (with $\mathbb{E}[X] = 0$) that $\mathbb{E}[X_i X_j X_k X_\ell] = 0$ as soon as one of the indices i, j, k, ℓ is different from all the other three, e.g., $i \notin \{j, k, \ell\}$, etc. The only cases when $\mathbb{E}[X_i X_j X_k X_\ell] \neq 0$ are as follows: (i) If $i = j = k = \ell$, then $\mathbb{E}[X_i X_j X_k X_\ell] = \mathbb{E}[X^4]$; there are n such configurations; (ii) If $\{i, j, k, \ell\}$ contains only two distinct values, say $a \neq b$ appearing as $aabb$, $abab$ and $abba$ in (4.18), then $\mathbb{E}[X_i X_j X_k X_\ell] = (\mathbb{E}[X^2])^2$; there are $3n(n-1)$ such configurations. It follows that

$$\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^4\right] = n\mathbb{E}\left[X^4\right] + 3n(n-1)(\mathbb{E}\left[X^2\right])^2,$$

whence

$$\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] = \frac{1}{n^3}\mathbb{E}\left[X^4\right] + 3\frac{n-1}{n^3}(\mathbb{E}\left[X^2\right])^2.$$

The conclusion (4.17) readily follows, and this completes the proof. ■

Proof 2 Assume $\mathbb{E}[|X|^2] < \infty$ – For each $k = 1, 2, \ldots$, we note that

$$\text{Var}\left[\frac{S_{k^2}}{k^2}\right] = \frac{\text{Var}[X]}{k^2}$$

so that

$$\sum_{k=1}^\infty \mathbb{P}\left[\left|\frac{S_{k^2}}{k^2}\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} \sum_{k=1}^\infty \frac{\text{Var}[X]}{k^2} < \infty, \quad \varepsilon > 0.$$

It follows from Theorem 3.1.2 that

$$\lim_{k\to\infty} \frac{S_{k^2}}{k^2} = \mathbb{E}[X] \quad \text{a.s.} \tag{4.19}$$

Now assume that the rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ are non-negative, i.e., $X \geq 0$ a.s. (in which case obviously $\mathbb{E}[X] \geq 0$). The case when the rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ are non-positive, i.e., $X \leq 0$ a.s., can be handed *mutatis mutants*.

Fix $n = 1, 2, \ldots$. There exists a unique positive integer $k(n)$ such that

$$k(n)^2 \leq n < (k(n) + 1)^2. \tag{4.20}$$

Under the non-negativity assumption we note the inequalities

$$S_{k(n)^2} \leq S_n \leq S_{(k(n)+1)^2} \quad \text{a.s.}$$

by virtue of the fact that $X_\ell \geq 0$ a.s. for $\ell = k(n)^2, \ldots, (k(n)+1)^2 - 1$. It follows that

$$\frac{k(n)^2}{n} \cdot \left(\frac{S_{k(n)^2}}{k(n)^2}\right) \leq \frac{S_n}{n} \leq \frac{(k(n)+1)^2}{n} \cdot \left(\frac{S_{(k(n)+1)^2}}{(k(n)+1)^2}\right). \tag{4.21}$$

Using (4.20) we readily get

$$\frac{k(n)^2}{n} \leq 1 < \frac{k(n)^2}{n} + 2 \cdot \frac{k(n)}{\sqrt{n}} \cdot \frac{1}{\sqrt{n}} + \frac{1}{n} \tag{4.22}$$

It is now straightforward to conclude from the first inequality in (4.22) that

$$\limsup_{n\to\infty} \frac{k(n)^2}{n} \leq 1$$

with $\frac{k(n)}{\sqrt{n}} \leq 1$, and the second inequality in (4.22) therefore yields $1 \leq \liminf_{n\to\infty} \frac{k(n)^2}{n}$. As a result, $\lim_{n\to\infty} \frac{k(n)^2}{n} = 1$ (whence $\lim_{n\to\infty} k(n) = \infty$ as expected). Finally let n go to infinity in (4.21), and we readily get (4.16) upon combining this last conclusion with the convergence (4.19).

To complete the proof note that $\mathbb{E}[(X^\pm)^2] < \infty$ since $\mathbb{E}[|X|^2] = \mathbb{E}[(X^+)^2] + \mathbb{E}[(X^-)^2]$. Thus, it holds that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n X_k^\pm}{n} = \mathbb{E}\left[X^\pm\right] \quad a.s. \tag{4.23}$$

since the rvs $\{X^\pm, X_k^\pm,\ k = 1, 2, \ldots\}$ form an i.i.d. sequence of rvs with finite second moments. The desired result (4.16) automatically follows since

$$X_n = X_n^+ - X_n^-, \quad n = 1, 2, \ldots$$

and $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$. ∎

4.5 The Central Limit Theorem

The Central Limit Theorem completes the Law of Large Numbers, in that it provides some indication as to the rate at which convergence takes place.

Theorem 4.5.1 *Assume the rvs $\{X, X_n,\ n = 1, 2, \ldots\}$ to be i.i.d. rvs with $\mathbb{E}[|X|^2] < \infty$. Then, we have*

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}[X]\right) \Longrightarrow_n \sqrt{\mathrm{Var}[X]} \cdot U \tag{4.24}$$

where U is standard zero-mean unit-variance Gausssian rv.

Proof. Fix $n = 1, 2, \ldots$ and θ in $\mathbb{R}$. This time, as in the proof of Theorem 4.3.1 we get

$$\mathbb{E}\left[e^{i\theta\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}[X]\right)}\right] = \left(\mathbb{E}\left[e^{i\frac{\theta}{\sqrt{n}}(X - \mathbb{E}[X])}\right]\right)^n$$

under the enforced independence.

Using Theorem 2.4.1 (with $k = 2$ and $x = X - \mathbb{E}[X]$), we get

$$\begin{aligned} & e^{i\theta(X - \mathbb{E}[X])} \\ = \;& 1 + i\theta(X - \mathbb{E}[X]) - \frac{\theta^2}{2}(X - \mathbb{E}[X])^2 \\ & -\frac{\theta^2}{2}\int_0^{X-\mathbb{E}[X]} (X - \mathbb{E}[X] - t)\left(e^{i\theta t} - 1\right) dt, \end{aligned} \tag{4.25}$$

and taking expectations yields

$$\begin{aligned} & \mathbb{E}\left[e^{i\theta(X - \mathbb{E}[X])}\right] \\ = \;& 1 - \frac{\theta^2}{2} \cdot \mathrm{Var}[X] - \frac{\theta^2}{2} \cdot C_2(\theta) \end{aligned} \tag{4.26}$$

with

$$C_2(\theta) \equiv \mathbb{E}\left[\int_0^{X-\mathbb{E}[X]} (X - \mathbb{E}[X] - t)\left(e^{i\theta t} - 1\right) dt\right]. \tag{4.27}$$

Substituting θ by $\frac{\theta}{\sqrt{n}}$ in this last relation leads to

$$\mathbb{E}\left[e^{i\frac{\theta}{\sqrt{n}}(X - \mathbb{E}[X])}\right] = 1 - \frac{\theta^2}{2n} \cdot \mathrm{Var}[X] - \frac{\theta^2}{2n} \cdot C_2\left(\frac{\theta}{\sqrt{n}}\right)$$

so that

$$\mathbb{E}\left[e^{i\theta\sqrt{n}\left(\frac{S_n}{n}-\mathbb{E}[X]\right)}\right] = \left(1 - \frac{\theta^2}{2n} \cdot \mathrm{Var}\left[X\right] - \frac{\theta^2}{2n} \cdot C_2\left(\frac{\theta}{\sqrt{n}}\right)\right)^n.$$

Again, by Dominated Convergence, we obtain

$$\lim_{n\to\infty} C_2\left(\frac{\theta}{\sqrt{n}}\right) = 0$$

under the second moment condition $\mathbb{E}\left[|X|^2\right] < \infty$, whence

$$\lim_{n\to\infty} n\left(\frac{\theta^2}{2n} \cdot \mathrm{Var}\left[X\right] - \frac{\theta^2}{2n} \cdot C_2\left(\frac{\theta}{\sqrt{n}}\right)\right) = \frac{\theta^2}{2} \cdot \mathrm{Var}\left[X\right]$$

It follows that

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\theta\sqrt{n}\left(\frac{S_n}{n}-\mathbb{E}[X]\right)}\right] = e^{-\frac{\theta^2}{2}\cdot\mathrm{Var}[X]}$$

This complete the proof of (4.24). ∎

4.6 The Central Limit Theorem – An application

We are still in the setting of Theorem 4.5.1. We can rephrase (4.24) as

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}\left[X\right]\right) \le x\right] \\ (4.28) \qquad &= \mathbb{P}\left[\sqrt{\mathrm{Var}[X]} \cdot U \le x\right], \quad x \in \mathbb{R}. \end{aligned}$$

as we recall that every point in $\mathbb{R}$ is a point of continuity for the rv U (or $\sqrt{\mathrm{Var}[X]}\cdot U$).

It follows that

$$\begin{aligned} &\lim_{n\to\infty} \mathbb{P}\left[\left|\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}\left[X\right]\right)\right| \le x\right] \\ &= \mathbb{P}\left[\sqrt{\mathrm{Var}[X]} \cdot U \le x\right] - \mathbb{P}\left[\sqrt{\mathrm{Var}[X]} \cdot U \le -x\right] \\ &= \Phi\left(\frac{x}{\sqrt{\mathrm{Var}[X]}}\right) - \Phi\left(-\frac{x}{\sqrt{\mathrm{Var}[X]}}\right) \\ (4.29) \qquad &= 2\Phi\left(\frac{x}{\sqrt{\mathrm{Var}[X]}}\right) - 1, \quad x \ge 0. \end{aligned}$$

Fix $x \geq 0$ and $n = 1, 2, \ldots$: We have

$$\left| \sqrt{n} \left(\frac{S_n}{n} - \mathbb{E}[X] \right) \right| \leq x$$

if and only if

$$-x \leq \sqrt{n} \left(\frac{S_n}{n} - \mathbb{E}[X] \right) \leq x$$

if and only if

$$\mathbb{E}[X] \in \left[\frac{S_n}{n} - \frac{x}{\sqrt{n}}, \frac{S_n}{n} + \frac{x}{\sqrt{n}} \right].$$

Thus, if we think of

$$\widehat{X}_n = \frac{S_n}{n}, \quad n = 1, 2, \ldots$$

as an estimate of $\mathbb{E}[X]$ on the basis of the observations $X_1, \ldots, X_n$, then the SLLNs already tells us that the estimate is increasingly accurate as n gets large since

$$\lim_{n \to \infty} \widehat{X}_n = \mathbb{E}[X] \quad a.s.$$

The calculations above show via (4.29) that

$$\begin{aligned} & \lim_{n \to \infty} \mathbb{P} \left[\mathbb{E}[X] \in \left[\widehat{X}_n - \frac{x}{\sqrt{n}}, \widehat{X}_n + \frac{x}{\sqrt{n}} \right] \right] \\ = \; & 2\Phi \left(\frac{x}{\sqrt{\mathrm{Var}[X]}} \right) - 1, \quad x \geq 0. \end{aligned} \tag{4.30}$$

In other words, for large n, the *unknown* value $\mathbb{E}[X]$ lies in a symmetric interval centered at the estimate $\widehat{X}_n$ (obtained from the *observed* data $X_1, \ldots, X_n$) of width $\frac{2x}{\sqrt{n}}$ with a probability approximately given by

$$2\Phi \left(\frac{x}{\sqrt{\mathrm{Var}[X]}} \right) - 1,$$

the accuracy of this approximation improving with increasing n. With α in $(0, 1)$ *given*, we can ensure that

$$\mathbb{P} \left[\mathbb{E}[X] \in \left[\widehat{X}_n - \frac{x}{\sqrt{n}}, \widehat{X}_n + \frac{x}{\sqrt{n}} \right] \right] \simeq 1 - \alpha$$

for large n if we select $x \geq 0$ such that

$$2\Phi\left(\frac{x}{\sqrt{\mathrm{Var}[X]}}\right) - 1 = 1 - \alpha,$$

or equivalently,

$$\Phi\left(\frac{x}{\sqrt{\mathrm{Var}[X]}}\right) = 1 - \frac{\alpha}{2}.$$

With λ in $(0,1)$ let z_λ denote the unique solution to the nonlinear equation

$$1 - \Phi(x) = \lambda, \quad x \in \mathbb{R}.$$

Equivalently,

$$\mathbb{P}\left[U > x\right] = \lambda, \quad x \in \mathbb{R}.$$

With this notation we see that the *random* interval

$$\left[\frac{S_n}{n} - \frac{z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{Var}[X]}}{\sqrt{n}}, \frac{S_n}{n} + \frac{z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{Var}[X]}}{\sqrt{n}}\right]$$

is known as the *confidence interval* for estimating $\mathbb{E}\left[X\right]$ on the basis data $X_1, \ldots, X_n$ with confidence $(1-\alpha)\%$

Note that this analysis is predicated on knowing the variance $\mathrm{Var}[X]$. When this value is unknown, we replace $\mathrm{Var}[X]$ by the *sample variance* S_n^2 given by

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \frac{1}{n}\sum_{\ell=1}^{n} X_\ell\right)^2, \quad n = 2, 3, \ldots$$

4.7 Poisson convergence

The setting is a follows: For each $n = 1, 2, \ldots$, let $X_1(p_n), \ldots, X_n(p_n)$ denote a collection of i.i.d. Bernoulli rvs with parameters p_n in $(0,1)$. i.e.,

$$\mathbb{P}\left[X_{k,n}(p_n) = 1\right] = 1 - \mathbb{P}\left[X_{k,n}(p_n) = 0\right] = p_n, \quad k = 1, \ldots, n$$

Write

$$S_n = \sum_{k=1}^{n} X_k(p_n), \quad n = 1, 2, \ldots$$

Theorem 4.7.1 *Assume there exists $\lambda > 0$ such that*

$$\lim_{n\to\infty} np_n = \lambda. \tag{4.31}$$

Then, we have

$$S_n \Longrightarrow_n \Pi(\lambda) \tag{4.32}$$

where $\Pi(\lambda)$ denotes a Poisson rv with parameter λ.

The convergence (4.32) can be restated as

$$\lim_{n\to\infty} \mathbb{P}\left[S_n = k\right] = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0,1,\ldots \tag{4.33}$$

We give two proofs of this important result.

Proof 1 The first proof uses the characterization of weak convergence for integer-valued rvs given in Theorem 3.9.1: Fix $n = 1,2,\ldots$. Under the independence assumptions, the rv S_n is a binomial rv $\mathrm{Bin}(n;p_n)$. Thus, Fix $k = 0,1,\ldots$. For every integer n such that $k \leq n$ we have

$$\begin{aligned}
\mathbb{P}\left[S_n = k\right] &= \binom{n}{k} p_n^k \left(1-p_n\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \cdot p_n^k \left(1-p_n\right)^{n-k} \\
&= \frac{1}{k!}\left(\frac{p_n}{1-p_n}\right)^k \cdot \frac{n!}{(n-k)!} \cdot \left(1-p_n\right)^n \\
&= \frac{1}{k!}\left(\frac{np_n}{1-p_n}\right)^k \cdot \frac{n!}{n^k(n-k)!} \cdot \left(1-p_n\right)^n . \qquad (4.34)
\end{aligned}$$

It is plain that

$$\lim_{n\to\infty} \frac{n!}{n^k(n-k)!} = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k} = 1$$

while (4.31) implies

$$\lim_{n\to\infty} \left(1-p_n\right)^n = \lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^n = e^{-\lambda}$$

and

$$\lim_{n\to\infty} \frac{p_n}{1-p_n} = \lambda$$

since $\lim_{n\to\infty} p_n = 0$. Collecting we conclude to (4.33) as we make use of Theorem 3.9.1. ■

Proof 2 This second proof relies on the characterization of weak convergence for integer-valued rvs given in terms of probability generating functions: Fix $n = 1, 2, \ldots$. For each θ in $\mathbb{R}$ we get

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta S_n}\right] &= \mathbb{E}\left[e^{i\theta \sum_{k=1}^n X_k(p_n)}\right] \\
&= \mathbb{E}\left[\prod_{k=1}^n e^{i\theta X_k(p_n)}\right] \\
&= \prod_{k=1}^n \mathbb{E}\left[e^{i\theta X_k(p_n)}\right] \\
&= \left(1 - p_n + p_n e^{i\theta}\right)^n \\
&= \left(1 - p_n\left(1 - e^{i\theta}\right)\right)^n . \qquad (4.35)
\end{aligned}$$

Under (4.31) we get that

$$\lim_{n\to\infty} np_n\left(1 - e^{i\theta}\right) = \lambda\left(1 - e^{i\theta}\right).$$

Thus,

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\theta S_n}\right] = e^{-\lambda(1-e^{i\theta})}, \quad \theta \in \mathbb{R}$$

and the conclusion (4.32) follows since

$$\begin{aligned}
\mathbb{E}\left[e^{i\theta\Pi(\lambda)}\right] &= \sum_{k=0}^\infty \frac{\lambda^k}{k!} e^{-\lambda} \cdot e^{ik\theta} \\
&= \left(\sum_{k=0}^\infty \frac{1}{k!}\left(\lambda e^{i\theta}\right)^k\right) e^{-\lambda} = e^{-\lambda(1-e^{i\theta})}, \quad \theta \in \mathbb{R} \qquad (4.36)
\end{aligned}$$

as we use Theorem 3.9.2. ■

Chapter 5

Gaussian Random Variables

This chapter is devoted to a brief discussion of the class of Gaussian rvs. In particular, for easy reference we have collected various facts and properties to be used repeatedly.

5.1 Scalar Gaussian rvs

With

$$\mu \in \mathbb{R} \quad \text{and} \quad \sigma \geq 0,$$

an $\mathbb{R}$-valued rv X is said to be a *Gaussian* (or normally distributed) rv with mean μ and variance σ^2 if either it is degenerate to a constant with $X = \mu$ a.s. (in which case $\sigma = 0$) or the probability distribution of X is of the form

$$\mathbb{P}\left[X \leq x\right] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R}$$

(in which case $\sigma^2 > 0$). Under either circumstance, it can be shown that

$$\mathbb{E}\left[e^{i\theta X}\right] = e^{i\theta\mu - \frac{\sigma^2}{2}\cdot\theta^2}, \quad \theta \in \mathbb{R}. \tag{5.1}$$

This fact is established in Section 5.11. The following equivalent definition captures both cases.

An $\mathbb{R}$-valued rv X is said to be a Gaussian rv with mean μ (in $\mathbb{R}$) and variance $\sigma^2 > 0$ if its characteristic function is given by

$$\mathbb{E}\left[e^{i\theta X}\right] = e^{i\theta\mu - \frac{\sigma^2}{2}\cdot\theta^2}, \quad \theta \in \mathbb{R}. \tag{5.2}$$

It is then follows by differentiation that

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \mathbb{E}\left[X^2\right] = \mu^2 + \sigma^2 \tag{5.3}$$

so that $\text{Var}[X] = \sigma^2$. This confirms the meaning ascribed to the parameters μ and σ^2 as mean and variance, respectively.

It is a simple matter to check that if X is normally distributed with mean μ and variance σ^2, then for scalars a and b, the rv $aX + b$ is also normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$. In particular, with $\sigma > 0$, the rv $\sigma^{-1}(X - \mu)$ is a Gaussian rv with mean zero and unit variance.

5.2 The standard Gaussian rv

The Gaussian rv with mean zero and unit variance occupies a very special place among Gaussian rvs, and is often referred to as the *standard* Gaussian rv. Throughout, we denote by U the Gaussian rv with zero mean and unit variance. Its probability distribution function is given by

$$\mathbb{P}[U \le x] = \Phi(x) := \int_{-\infty}^{x} \phi(t)dt, \quad x \in \mathbb{R} \tag{5.4}$$

with density function ϕ given by

$$\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}. \tag{5.5}$$

As should be clear from earlier comments, the importance of this standard rv U stems from the fact that for any Gaussian rv X with mean μ and variance σ^2, it holds that $X =_{st} \mu + \sigma U$, so that

$$\begin{aligned}
\mathbb{P}[X \le x] &= \mathbb{P}\left[\sigma^{-1}(X - \mu) \le \sigma^{-1}(x - \mu)\right] \\
&= \mathbb{P}\left[U \le \sigma^{-1}(x - \mu)\right] \\
&= \Phi(\sigma^{-1}(x - \mu)), \quad x \in \mathbb{R}.
\end{aligned}$$

The evaluation of probabilities involving Gaussian rvs thus reduces to the evaluation of related probabilities for the standard Gaussian rv.

For each x in $\mathbb{R}$, we note by symmetry that $\mathbb{P}[U \leq -x] = \mathbb{P}[U > x]$, so that $\Phi(-x) = 1 - \Phi(x)$, and Φ is therefore fully determined by the complementary probability distribution function of U on $[0, \infty)$, namely

$$Q(x) := 1 - \Phi(x) = \mathbb{P}[U > x], \quad x \geq 0. \tag{5.6}$$

5.3 Evaluating $Q(x)$

The complementary distribution function (5.6) repeatedly enters the computation of various probabilities of error. Given its importance, we need to develop good approximations to $Q(x)$ over the entire range $x \geq 0$.

The error function In the literature on digital communications, probabilities of error are often expressed in terms of the so-called *error function* $\mathrm{Erf} : \mathbb{R}_+ \to \mathbb{R}$ and of its complement $\mathrm{Erfc} : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad x \geq 0 \tag{5.7}$$

and

$$\mathrm{Erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt, \quad x \geq 0. \tag{5.8}$$

A simple change of variables ($t = \frac{u}{\sqrt{2}}$) in these integrals leads to the relationships

$$\mathrm{Erf}(x) = 2\left(\Phi(x\sqrt{2}) - \frac{1}{2}\right) \quad \text{and} \quad \mathrm{Erfc}(x) = 2Q(x\sqrt{2}),$$

so that

$$\mathrm{Erf}(x) = 1 - \mathrm{Erfc}(x), \quad x \geq 0.$$

Conversely, we also have

$$\Phi(x) = \frac{1}{2}\left(1 + \mathrm{Erf}\left(\frac{x}{\sqrt{2}}\right)\right) \quad \text{and} \quad Q(x) = \frac{1}{2}\mathrm{Erfc}\left(\frac{x}{\sqrt{2}}\right).$$

Thus, knowledge of any one of the quantities Φ, Q, Erf or Erfc is equivalent to that of the other three quantities. Although the last two quantities do not have a probabilistic interpretation, evaluating Erf is computationally more efficient. Indeed, $\mathrm{Erf}(x)$ is an integral of a positive function over the *finite* interval $[0, x]$ (and not over an infinite interval as in the other cases).

Chernoff bounds To approximate $Q(x)$ we begin with a crude bound which takes advantage of (**??**): Fix $x > 0$. For each $\theta > 0$, the usual Chernoff bound argument gives

$$\begin{aligned} \mathbb{P}\left[U > x\right] &\leq \mathbb{E}\left[e^{\theta U}\right] e^{-\theta x} \\ &= e^{-\theta x + \frac{\theta^2}{2}} \\ &= e^{-\frac{x^2}{2}} e^{\frac{(\theta - x)^2}{2}} \end{aligned} \tag{5.9}$$

where in the last equality we made use of a completion-of-square argument. The best lower bound

$$Q(x) \leq e^{-\frac{x^2}{2}}, \quad x \geq 0 \tag{5.10}$$

is achieved upon selecting $\theta = x$ in (5.9). We refer to the bound (5.10) as a Chernoff bound; it is not very accurate for small $x > 0$ since $\lim_{x\to 0} Q(x) = \frac{1}{2}$ while $\lim_{x\to 0} e^{-\frac{x^2}{2}} = 1$.

Approximating $Q(x)$ ($x \to \infty$) The Chernoff bound shows that $Q(x)$ decays to zero for large x at least as fast as $e^{-\frac{x^2}{2}}$. However, sometimes more precise information is needed regarding the rate of decay of $Q(x)$. This issue is addressed as follows:

For each $x \geq 0$, a straigthforward change of variable yields

$$\begin{aligned} Q(x) &= \int_x^\infty \phi(t)dt \\ &= \int_0^\infty \phi(x+t)dt \\ &= \phi(x) \int_0^\infty e^{-xt} e^{-\frac{t^2}{2}} dt. \end{aligned} \tag{5.11}$$

With the Taylor series expansion of $e^{-\frac{t^2}{2}}$ in mind, approximations for $Q(x)$ of increased accuracy thus suggest themselves by simply approximating the second exponential factor (namely e^{-xt}) in the integral at (5.11) by terms of the form

$$\sum_{k=0}^{n} \frac{(-1)^k}{2^k k!} t^{2k}, \quad n = 0, 1, \ldots \tag{5.12}$$

To formulate the resulting approximation contained in Proposition 5.3.1 given next, we set

$$Q_n(x) = \phi(x) \int_0^\infty \left(\sum_{k=0}^n \frac{(-1)^k}{2^k k!} t^{2k} \right) e^{-xt} dt, \quad x \geq 0$$

for each $n = 0, 1, \ldots$.

Proposition 5.3.1 *Fix $n = 0, 1, \ldots$. For each $x > 0$ it holds that*

$$Q_{2n+1}(x) \leq Q(x) \leq Q_{2n}(x), \tag{5.13}$$

with

$$| Q(x) - Q_n(x) | \leq \frac{(2n)!}{2^n n!} x^{-(2n+1)} \phi(x). \tag{5.14}$$

where

$$Q_n(x) = \phi(x) \sum_{k=0}^n \frac{(-1)^k (2k)!}{2^k k!} x^{-(2k+1)}. \tag{5.15}$$

A proof of Proposition 5.3.1 can be found in Section **??**. Upon specializing (5.13) to $n = 0$ we get

$$\frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} \left(1 - \frac{1}{x^2} \right) \leq Q(x) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}, \quad x > 0 \tag{5.16}$$

and the asymptotics

$$Q(x) \sim \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} \quad (x \to \infty) \tag{5.17}$$

follow. Note that the lower bound in (5.16) is meaningful only when $x \geq 1$.

5.4 Gaussian random vectors

Let $\boldsymbol{\mu}$ denote a vector in $\mathbb{R}^d$ and let $\boldsymbol{\Sigma}$ be a symmetric and non-negative definite $d \times d$ matrix, i.e., $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$ and $\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta} \geq 0$ for all $\boldsymbol{\theta}$ in $\mathbb{R}^d$.

An $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ is said to be a Gaussian rv with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if there exist a $d \times p$ matrix $\boldsymbol{T}$ for some positive integer p and i.i.d. zero mean unit variance Gaussian rvs $U_1, \ldots, U_p$ such that

$$\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma} \tag{5.18}$$

and

$$\boldsymbol{X} =_{st} \boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{U}_p \tag{5.19}$$

where $\boldsymbol{U}_p$ is the $\mathbb{R}^p$-valued rv $(U_1, \ldots, U_p)'$.

From (5.18) and (5.19) it is plain that

$$\mathbb{E}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{U}_p] = \boldsymbol{\mu} + \boldsymbol{T}\mathbb{E}[\boldsymbol{U}_p] = \boldsymbol{\mu}$$

and

$$\begin{aligned} \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right] &= \mathbb{E}\left[\boldsymbol{T}\boldsymbol{U}_p(\boldsymbol{T}\boldsymbol{U}_p)'\right] \\ &= \boldsymbol{T}\mathbb{E}\left[\boldsymbol{U}_p\boldsymbol{U}_p'\right]\boldsymbol{T}' \\ &= \boldsymbol{T}\boldsymbol{I}_p\boldsymbol{T}' = \boldsymbol{\Sigma}, \end{aligned} \tag{5.20}$$

whence

$$\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{Cov}[\boldsymbol{X}] = \boldsymbol{\Sigma}.$$

Again this confirms the terminology used for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as mean vector and covariance matrix, respectively.

It is a well-known fact from Linear Algebra [, , p.] that for any symmetric and non-negative definite $d \times d$ matrix $\boldsymbol{\Sigma}$, there exists a $d \times d$ matrix $\boldsymbol{T}$ such that (5.18) holds with $p = d$. This matrix $\boldsymbol{T}$ can be selected to be *symmetric* and *non-negative definite*, and is called the *square root* of $\boldsymbol{\Sigma}$. Consequently, for any vector $\boldsymbol{\mu}$ in $\mathbb{R}^d$ and any symmetric non-negative definite $d \times d$ matrix $\boldsymbol{\Sigma}$, there always exists an $\mathbb{R}^d$-valued Gaussian rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ – Simply take

$$\boldsymbol{X} =_{st} \boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{U}_d$$

where $\boldsymbol{T}$ is the square root of $\boldsymbol{\Sigma}$.

5.5 Characteristic functions

The characteristic function of Gaussian rvs has an especially simple form which is now developed.

Lemma 5.5.1 *The characteristic function of a Gaussian $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by*

$$\mathbb{E}\left[e^{i\boldsymbol{\theta}'\boldsymbol{X}}\right] = e^{i\boldsymbol{\theta}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \in \mathbb{R}^d. \tag{5.21}$$

Conversely, any $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ whose characteristic function is given by (5.21) for some vector $\boldsymbol{\mu}$ in $\mathbb{R}^d$ and symmetric non-negative definite $d \times d$ matrix $\boldsymbol{\Sigma}$ is a Gaussian $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Proof. Consider an $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ which is a Gaussian rv with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. By definition, there exist a $d \times p$ matrix $\boldsymbol{T}$ for some positive integer p and i.i.d. zero mean unit variance Gaussian rvs $U_1, \ldots, U_p$ such that (5.18) and (5.19) hold.

For each $\boldsymbol{\theta}$ in $\mathbb{R}^d$, we get

$$\begin{aligned}
\mathbb{E}\left[e^{i\boldsymbol{\theta}'\boldsymbol{X}}\right] &= e^{i\boldsymbol{\theta}'\boldsymbol{\mu}} \cdot \mathbb{E}\left[e^{i\boldsymbol{\theta}'\boldsymbol{T}\boldsymbol{U}_p}\right] \\
&= e^{i\boldsymbol{\theta}'\boldsymbol{\mu}} \cdot \mathbb{E}\left[e^{i(\boldsymbol{T}'\boldsymbol{\theta})'\boldsymbol{U}_p}\right] \\
&= e^{i\boldsymbol{\theta}'\boldsymbol{\mu}} \cdot \mathbb{E}\left[e^{i\sum_{k=1}^p (\boldsymbol{T}'\boldsymbol{\theta})_k U_k}\right] \\
&= e^{i\boldsymbol{\theta}'\boldsymbol{\mu}} \cdot \prod_{k=1}^p \mathbb{E}\left[e^{i(\boldsymbol{T}'\boldsymbol{\theta})_k U_k}\right] \qquad (5.22) \\
&= e^{i\boldsymbol{\theta}'\boldsymbol{\mu}} \cdot \prod_{k=1}^p e^{-\frac{1}{2}|(\boldsymbol{T}'\boldsymbol{\theta})_k|^2} \qquad (5.23)
\end{aligned}$$

The equality (5.22) is a consequence of the independence of the rvs $U_1, \ldots, U_p$, while (5.23) follows from their Gaussian character (and (**??**)).

Next, we note that

$$\begin{aligned}
\sum_{k=1}^p |(\boldsymbol{T}'\boldsymbol{\theta})_k|^2 &= (\boldsymbol{T}'\boldsymbol{\theta})'(\boldsymbol{T}'\boldsymbol{\theta}) \\
&= \boldsymbol{\theta}'(\boldsymbol{T}\boldsymbol{T}')\boldsymbol{\theta} = \boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta} \qquad (5.24)
\end{aligned}$$

upon invoking (5.18). It is now plain from (5.23) that the characteristic function of the Gaussian $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ is given by (5.21).

Conversely, consider an $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ with characteristic function of the form (5.21) for some vector $\boldsymbol{\mu}$ in $\mathbb{R}^d$ and some symmetric non-negative definite $d \times d$ matrix $\boldsymbol{\Sigma}$. By comments made earlier, there exists a $d \times d$ matrix $\boldsymbol{T}$ such that (5.18) holds. By the first part of the proof, the $\mathbb{R}^d$-valued rv $\widetilde{\boldsymbol{X}}$ given by $\widetilde{\boldsymbol{X}} := \boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{U}_d$ has characteristic function given by (5.21). Since a probability distribution is completely determined by its characteristic function, it follows that the rvs $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$ obey the same distribution. The rv $\widetilde{\boldsymbol{X}}$ being Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the rv $\boldsymbol{X}$ is necessarily Gaussian as well with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. ■

5.6 Existence of a density

In general, an $\mathbb{R}^d$-valued Gaussian rv as defined above may not admit a density function. To see why, consider the null space of its covariance matrix $\boldsymbol{\Sigma}$,[1] namely

$$N(\boldsymbol{\Sigma}) := \{\boldsymbol{x} \in \mathbb{R}^d : \ \boldsymbol{\Sigma}\boldsymbol{x} = \boldsymbol{0}_d\}.$$

Observe that $\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta} = 0$ if and only if $\boldsymbol{\theta}$ belongs to $N(\boldsymbol{\Sigma})$, in which case (5.21) yields

$$\mathbb{E}\left[e^{i\boldsymbol{\theta}'(\boldsymbol{X}-\boldsymbol{\mu})}\right] = 1$$

and we conclude that

$$\boldsymbol{\theta}'(\boldsymbol{X}-\boldsymbol{\mu}) = 0 \quad \text{a.s.}$$

In other words, with probability one, the rv $\boldsymbol{X}-\boldsymbol{\mu}$ is orthogonal to the linear space $N(\boldsymbol{\Sigma})$.

To proceed, we assume that the covariance matrix $\boldsymbol{\Sigma}$ is not trivial (in that it has some non-zero entries) for otherwise $\boldsymbol{X} = \boldsymbol{\mu}$ a.s. In the non-trivial case, there are now two possibilities depending on the $d \times d$ matrix $\boldsymbol{\Sigma}$ being positive definite or not. Note that the positive definiteness of $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta} = 0$ necessarily implies $\boldsymbol{\theta} = \boldsymbol{0}_d$, is equivalent to the condition $N(\boldsymbol{\Sigma}) = \{\boldsymbol{0}_d\}$.

If the $d \times d$ matrix $\boldsymbol{\Sigma}$ is not positive definite, hence only positive semi-definite, then the mass of the rv $\boldsymbol{X}-\boldsymbol{\mu}$ is concentrated on the orthogonal space $N(\boldsymbol{\Sigma})^\perp$ of $N(\boldsymbol{\Sigma})$, whence the distribution of $\boldsymbol{X}$ has its support on the linear manifold $\boldsymbol{\mu} + N(\boldsymbol{\Sigma})^\perp$ and is singular with respect to Lebesgue measure.

On the other hand, if the $d \times d$ matrix $\boldsymbol{\Sigma}$ is positive definite, then the matrix $\boldsymbol{\Sigma}$ is invertible, $\det(\boldsymbol{\Sigma}) \neq 0$ and the Gaussian rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ admits a density function given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \quad \boldsymbol{x} \in \mathbb{R}^d.$$

[1] This linear space is sometimes called the kernel of $\boldsymbol{\Sigma}$.

5.7 Linear transformations

The following result is very useful in many contexts, and shows that linear transformations preserve the Gaussian character:

Lemma 5.7.1 *let $\boldsymbol{\nu}$ be an element of $\mathbb{R}^q$ and let $\boldsymbol{A}$ be an $q \times d$ matrix. Then, for any Gaussian rv $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the $\mathbb{R}^q$-valued rv $\boldsymbol{Y}$ given by*

$$\boldsymbol{Y} = \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{X}$$

is also a Gaussian rv with mean vector $\boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$.

Proof. First, by linearity we note that

$$\mathbb{E}\left[\boldsymbol{Y}\right] = \mathbb{E}\left[\boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{X}\right] = \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{\mu}$$

so that

$$\begin{aligned} \text{Cov}[\boldsymbol{Y}] &= \mathbb{E}\left[\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})\left(\boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\mu})\right)'\right] \\ &= \boldsymbol{A}\mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\right]\boldsymbol{A}' \\ &= \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'. \end{aligned} \tag{5.25}$$

Consequently, the $\mathbb{R}^q$-valued rv $\boldsymbol{Y}$ has mean vector $\boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$.

Next, by the Gaussian character of $\boldsymbol{X}$, there exist a $d \times p$ matrix $\boldsymbol{T}$ for some positive integer p and i.i.d. zero mean unit variance Gaussian rvs $U_1, \ldots, U_p$ such that (5.18) and (5.19) hold. Thus,

$$\begin{aligned} \boldsymbol{Y} &=_{st} \boldsymbol{\nu} + \boldsymbol{A}\left(\boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{U}_p\right) \\ &= \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{T}\boldsymbol{U}_p \\ &= \widetilde{\boldsymbol{\mu}} + \widetilde{\boldsymbol{T}}\boldsymbol{U}_p \end{aligned} \tag{5.26}$$

with

$$\widetilde{\boldsymbol{\mu}} := \boldsymbol{\nu} + \boldsymbol{A}\boldsymbol{\mu} \quad \text{and} \quad \widetilde{\boldsymbol{T}} := \boldsymbol{A}\boldsymbol{T}$$

and the Gaussian character of $\boldsymbol{Y}$ is established. ■

This result can also be established through the evaluation of the characteristic function of the rv $\boldsymbol{Y}$. As an immediate consequence of Lemma 5.7.1 we get

Corollary 5.7.1 *Consider a Gaussian rv $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For any subset I of $\{1, \ldots, d\}$ with $|I| = q \leq d$, the $\mathbb{R}^q$-valued rv $\boldsymbol{X}_I$ given by $\boldsymbol{X}_I = (X_i,\ i \in I)'$ is a Gaussian rv with mean vector $(\mu_i,\ i \in I)'$ and covariance matrix $(\Sigma_{ij},\ i, j \in I)$.*

5.8 Independence of Gaussian rvs

Characterizing the mutual independence of Gaussian rvs turns out to be quite straightforward as the following suggests: Consider the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ where for each $s = 1, \ldots, r$, the rv$\boldsymbol{X}_s$ is an $\mathbb{R}^{d_s}$-valued rv with mean vector $\boldsymbol{\mu}_s$ and covariance matrix $\boldsymbol{\Sigma}_s$. With $d = d_1 + \ldots + d_r$, let $\boldsymbol{X}$ denote the $\mathbb{R}^d$-valued rv obtained by concatenating $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$, namely

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_r \end{pmatrix}. \tag{5.27}$$

Its mean vector $\boldsymbol{\mu}$ is simply

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_r \end{pmatrix} \tag{5.28}$$

while its covariance matrix $\boldsymbol{\Sigma}$ can be written in block form as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{1,2} & \ldots & \boldsymbol{\Sigma}_{1,r} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_2 & \ldots & \boldsymbol{\Sigma}_{2,r} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{r,1} & \boldsymbol{\Sigma}_{r,2} & \ldots & \boldsymbol{\Sigma}_r \end{pmatrix} \tag{5.29}$$

with the notation

$$\boldsymbol{\Sigma}_{s,t} := \mathrm{Cov}[\boldsymbol{X}_s, \boldsymbol{X}_t] \quad s, t = 1, \ldots, r.$$

Lemma 5.8.1 *With the notation above, assume the $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ to be a Gaussian rv with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, for each $s = 1, \ldots, r$, the rv $\boldsymbol{X}_s$ is a Gaussian rv with mean vector $\boldsymbol{\mu}_s$ and covariance matrix $\boldsymbol{\Sigma}_s$. Moreover, the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ are mutually independent Gaussian rvs if and only they are uncorrelated, i.e.,*

$$\boldsymbol{\Sigma}_{s,t} = \delta(s,t)\boldsymbol{\Sigma}_t, \quad s, t = 1, \ldots, r. \tag{5.30}$$

The first part of Lemma 5.8.1 is a simple rewrite of Corollary 5.7.1. Sometimes we refer to the fact that the rv $\boldsymbol{X}$ is Gaussian by saying that the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ are *jointly* Gaussian. A converse to Lemma 5.8.1 is available:

Lemma 5.8.2 *Assume that for each $s = 1, \ldots, r$, the rv $\boldsymbol{X}_s$ is a Gaussian rv with mean vector $\boldsymbol{\mu}_s$ and covariance matrix Σ_s. If the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ are mutually independent, then the $\mathbb{R}^d$-valued rv $\boldsymbol{X}$ is an $\mathbb{R}^d$-valued Gaussian rv with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ as given by (5.29) with (5.30).*

It might be tempting to conclude that the Gaussian character of *each* of the rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ *alone* suffices to imply the Gaussian character of the combined rv $\boldsymbol{X}$. However, it can be shown through simple counterexamples that this is not so. In other words, the joint Gaussian character of $\boldsymbol{X}$ does not follow merely from that of its components $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ *without* further assumptions.

5.9 Convergence and limits of Gaussian rvs

In later chapters we will need to define integrals with respect to Gaussian processes. As in the deterministic case, these *stochastic* integrals will be defined as limits of partial sums of the form

$$X_n := \sum_{i=1}^{k_n} a_j^{(n)} Y_j^{(n)}, \quad n = 1, 2, \ldots \tag{5.31}$$

where for each $n = 1, 2, \ldots$, the integer k_n and the coefficients $a_j^{(n)}$, $j = 1, \ldots, k_n$, are non-random while the rvs $\{Y_j^{(n)},\ j = 1, \ldots, k_n\}$ are *jointly* Gaussian rvs. Typically, as n goes to infinity so does k_n. Note that under the foregoing assumptions for each $n = 1, 2, \ldots$, the rv X_n is Gaussian with

$$\mathbb{E}\left[X_n\right] = \sum_{i=1}^{k_n} a_j^{(n)} \mathbb{E}\left[Y_j^{(n)}\right] \tag{5.32}$$

and

$$\text{Var}[X_n] = \sum_{i=1}^{k_n} \sum_{j=1}^{k_n} a_i^{(n)} a_j^{(n)} \text{Cov}[Y_i^{(n)}, Y_j^{(n)}]. \tag{5.33}$$

Therefore, the study of such integrals is expected to pass through the convergence of sequence of rvs $\{X_n,\ n = 1, 2, \ldots\}$ of the form (5.31). Such considerations lead naturally to the need for the following result [, Thm. , p.]:

Lemma 5.9.1 *Let $\{\boldsymbol{X}_k,\ k = 1, 2, \ldots\}$ denote a collection of $\mathbb{R}^d$-valued Gaussian rvs. For each $k = 1, 2, \ldots$, let $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ denotes the mean vector and covariance matrix of the rv $\boldsymbol{X}_k$. The rvs $\{\boldsymbol{X}_k,\ k = 1, \ldots\}$ converge in distribution (in law) if and only there exist an element $\boldsymbol{\mu}$ in $\mathbb{R}^d$ and a $d \times d$ matrix $\boldsymbol{\Sigma}$ such that*

$$\lim_{k \to \infty} \boldsymbol{\mu}_k = \boldsymbol{\mu} \quad \text{and} \quad \lim_{k \to \infty} \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}. \tag{5.34}$$

In that case,

$$\boldsymbol{X}_k \Longrightarrow_k \boldsymbol{X}$$

where $\boldsymbol{X}$ is an $\mathbb{R}^d$-valued Gaussian rv with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

The second half of condition (5.34) ensures that the matrix $\boldsymbol{\Sigma}$ is symmetric and non-negative definite, hence a covariance matrix.

Returning to the partial sums (5.31) we see that Lemma 5.9.1 (applied with $d = 1$) requires identifying the limits $\mu = \lim_{n\to\infty} \mathbb{E}[X_n]$ and $\sigma^2 = \lim_{n\to\infty} \mathrm{Var}[X_n]$, in which case $X_n \Longrightarrow_n X$ where X is an $\mathbb{R}$-valued Gaussian rv with mean μ and variance Σ. In Section **??** we discuss a situation where this can be done quite easily.

5.10 Rvs derived from Gaussian rvs

Rayleigh rvs A rv X is said to be a *Rayleigh* rv with parameter σ ($\sigma > 0$) if

$$X =_{st} \sqrt{Y^2 + Z^2} \tag{5.35}$$

with Y and Z independent zero mean Gaussian rvs with variance σ^2. It is easy to check that

$$\mathbb{P}[X > x] = e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0 \tag{5.36}$$

with corresponding density function

$$\frac{d}{dx}\mathbb{P}[X \leq x] = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0. \tag{5.37}$$

It is also well known that the rv Θ given by

$$\Theta := \arctan\left(\frac{Z}{Y}\right) \tag{5.38}$$

is uniformly distributed over $[0, 2\pi)$ and independent of the Rayleigh rv X, i.e.,

$$\mathbb{P}\left[X \leq x, \Theta \leq \theta\right] = \frac{\theta}{2\pi}\left(1 - e^{-\frac{x^2}{2\sigma^2}}\right), \quad \theta \in [0, 2\pi),\ x \geq 0. \tag{5.39}$$

Rice rvs A rv X is said to be a *Rice* rv with parameters α (in $\mathbb{R}$) and σ ($\sigma > 0$) if

$$X =_{st} \sqrt{(\alpha + Y)^2 + Z^2} \tag{5.40}$$

with Y and Z independent zero mean Gaussian rvs with variance σ^2. It is easy to check that X admits a probability density function given by

$$\frac{d}{dx}\mathbb{P}\left[X \leq x\right] = \frac{x}{\sigma^2}e^{-\frac{x^2+\alpha^2}{2\sigma^2}} \cdot I_0\left(\frac{\alpha x}{\sigma^2}\right), \quad x \geq 0. \tag{5.41}$$

Here,

$$I_0(x) := \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos t}dt, \quad x \in \mathbb{R} \tag{5.42}$$

is the modified Bessel function of the first kind of order zero.

Chi-square rvs For each $n = 1, 2, \ldots$, the Chi-square rv with n degrees of freedom is the rv defined by

$$\chi_n^2 =_{st} U_1^2 + \ldots + U_n^2$$

where $U_1, \ldots, U_n$ are n i.i.d. standard Gaussian rvs.

5.11 A proof of (5.1)

Assume $\mu = 0$ and $\sigma^2 = 1$. Fix θ in $\mathbb{R}$. We need to evaluate

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{i\theta x}e^{-\frac{x^2}{2}}dx$$

Our starting point is the Taylor series expansion

$$e^{i\theta x} = \sum_{k=0}^{\infty} \frac{(i\theta x)^k}{k!}, \quad x \in \mathbb{R}.$$

Assuming a valid interchange of integration and summation (to be justified below), we get

$$\begin{aligned} \int_{\mathbb{R}} e^{i\theta x} e^{-\frac{x^2}{2}} dx &= \int_{\mathbb{R}} \left(\sum_{k=0}^{\infty} \frac{(i\theta x)^k}{k!} \right) e^{-\frac{x^2}{2}} dx \\ &= \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} \int_{\mathbb{R}} x^k e^{-\frac{x^2}{2}} dx \\ &= \sum_{k=0}^{\infty} \frac{(i\theta)^k}{k!} m_k \end{aligned} \tag{5.43}$$

where we have set

$$m_k = \int_{\mathbb{R}} x^k e^{-\frac{x^2}{2}} dx, \quad k = 0, 1, \ldots$$

Note that

$$m_k = 0, \quad k = 1, 3, 5, \ldots$$

by symmetry, so that

$$\begin{aligned} \int_{\mathbb{R}} e^{i\theta x} e^{-\frac{x^2}{2}} dx &= \sum_{\ell=0}^{\infty} \frac{(i\theta)^{2\ell}}{(2\ell)!} m_{2\ell} \\ &= \sum_{\ell=0}^{\infty} \frac{(-\theta^2)^{\ell}}{(2\ell)!} m_{2\ell} \end{aligned} \tag{5.44}$$

Therefore, it remains to compute $m_{2\ell}$, $\ell = 0, 1, \ldots$.

To that end, fix $\ell = 0, 1, \ldots$. By integration by parts yields

$$\begin{aligned} m_{2(\ell+1)} &= \int_{\mathbb{R}} x^{2(\ell+1)} e^{-\frac{x^2}{2}} dx \\ &= 2 \int_0^{\infty} x^{2(\ell+1)} e^{-\frac{x^2}{2}} dx \end{aligned}$$

$$
\begin{aligned}
&= 2\int_0^\infty x^{2\ell+1}\left(xe^{-\frac{x^2}{2}}\right)dx \\
&= 2\int_0^\infty x^{2\ell+1}\left(-e^{-\frac{x^2}{2}}\right)' dx \\
&= 2\left(\left[-x^{2\ell+1}e^{-\frac{x^2}{2}}\right]_0^\infty + \int_0^\infty (2\ell+1)x^{2\ell}e^{-\frac{x^2}{2}}\,dx\right) \\
&= 2(2\ell+1)\int_0^\infty x^{2\ell}e^{-\frac{x^2}{2}}\,dx. \qquad (5.45)
\end{aligned}
$$

In other words,

$$m_{2(\ell+1)} = (2\ell+1)m_{2\ell}, \quad \ell = 0, 1, \ldots$$

Iterating we get

$$
\begin{aligned}
m_{2\ell} &= (2\ell-1)m_{2(\ell-1)} \\
&= (2\ell-1)(2\ell-3)m_{2(\ell-2)} \\
&\;\vdots \\
&= (2\ell-1)(2\ell-3)(2\ell-5)\cdot\ldots\cdot 5\cdot 3\cdot 1\cdot m_0. \qquad (5.46)
\end{aligned}
$$

It follows that

$$m_{2\ell} = \frac{(2\ell)!}{(2\ell)(2(\ell-1))(2(\ell-2))\cdots(2\cdot 3)(2\cdot 2)(2\cdot 1)}\cdot m_0 = \frac{(2\ell)!}{2^\ell \ell!}\cdot m_0$$

for each $\ell = 1, 2, \ldots$. Collecting we conclude that

$$
\begin{aligned}
\int_{\mathbb{R}} e^{i\theta x}e^{-\frac{x^2}{2}}\,dx &= \sum_{\ell=0}^\infty \frac{\left(-\theta^2\right)^\ell}{(2\ell)!}\cdot\frac{(2\ell)!}{2^\ell \ell!}\cdot m_0 \\
&= \sum_{\ell=0}^\infty \frac{1}{\ell!}\left(-\frac{\theta^2}{2}\right)^\ell\cdot m_0 \\
&= e^{-\frac{\theta^2}{2}}\cdot m_0. \qquad (5.47)
\end{aligned}
$$

The desired conclusion now follows from the fact that $m_0 = \sqrt{2\pi}$ since

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,dx = 1.$$

5.12 Exercises

Ex. 5.1 Derive the relationships between the quantities Φ, Q, Erf or Erfc which are given in Section 5.3.

Ex. 5.2 Given the covariance matrix $\boldsymbol{\Sigma}$, explain why the representation (5.18)–(5.19) may not be unique. Give a counterexample.

Ex. 5.3 Give a proof for Lemma 5.8.1 and of Lemma 5.8.2.

Ex. 5.4 Construct an $\mathbb{R}^2$-valued rv $\boldsymbol{X} = (X_1, X_2)$ such that the $\mathbb{R}$-valued rvs X_1 and X_2 are each Gaussian but the $\mathbb{R}^2$-valued rv $\boldsymbol{X}$ is not (jointly) Gaussian.

Ex. 5.5 Derive the probability distribution function (5.36) of a Rayleigh rv with parameter σ ($\sigma > 0$).

Ex. 5.6 Show by direct arguments that if X is a Rayleigh distribution with parameter σ, then X^2 is exponentially distributed with parameter $(2\sigma^2)^{-1}$ [Hint: Compute $\mathbb{E}\left[e^{-\theta X^2}\right]$ for a Rayleigh rv X for $\theta \geq 0$.]

Ex. 5.7 Derive the probability distribution function (5.41) of a Rice rv with parameters α (in $\mathbb{R}$) and σ ($\sigma > 0$).

Ex. 5.8 Write a program to evaluate $Q_n(x)$.

Ex. 5.9 Let $X_1, \ldots, X_n$ be i.i.d. Gaussian rvs with zero mean and unit variance and write $S_n = X_1 + \ldots + X_n$. For each $a > 0$ show that

$$\mathbb{P}\left[S_n > na\right] \sim \frac{e^{-\frac{na^2}{2}}}{a\sqrt{2\pi n}} \quad (n \to \infty). \tag{5.48}$$

This asymptotic is known as the Bahadur-Rao correction to the large deviations asymptotics of S_n.

Ex. 5.10 Find all the moments $\mathbb{E}\left[U^p\right]$ ($p = 1, \ldots$) where U is a zero-mean unit variance Gaussian rv.

Ex. 5.11 Find all the moments $\mathbb{E}\left[U^p\right]$ ($p = 1, \ldots$) where X is a χ^2_n-rv with n degrees of freedom.