The Central Limit Theorem

Experimentalists think that it is a mathematical theorem, while the mathematicians believe it to be an experimental fact.

–Gabriel Lippman, in a discussion with J. H. Poincaré about the CLT

Let S_n denote the total number of successes in n independent Bernoulli trials, where the probability of success per trial is some fixed number $p \in (0\,,1)$. The De Moivre–Laplace central limit theorem (p. 19) asserts that for all real numbers $a < b$,

$$\lim_{n\to\infty} \mathrm{P}\left\{ a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b \right\} = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx. \tag{7.1}$$

We will soon see that (7.1) implies that the distribution of S_n is close to that of $N(np\,,\sqrt{np(1-p)})$; see Example 7.3 below. In this chapter we discuss the definitive formulation of this theorem. Its statement involves the notion of weak convergence which we discuss next.

1. Weak Convergence

Definition 7.1. Let $\mathbf{X}$ denote a topological space, and suppose μ, μ_1, $\mu_2, \ldots$ are probability (or more generally, finite) measures on $(\mathbf{X}\,,\mathscr{B}(\mathbf{X}))$. We say that μ_n *converges weakly* to μ, and write $\mu_n \Rightarrow \mu$, if

$$\lim_{n\to\infty} \int f\,d\mu_n = \int f\,d\mu, \tag{7.2}$$

for all bounded continuous functions $f : \mathbf{X} \to \mathbf{R}$. If the respective distributions of X_n and X are μ_n and μ, and if $\mu_n \Rightarrow \mu$, then we also say that X_n *converges weakly* to X and write $X_n \Rightarrow X$. This is equivalent to saying that

$$\lim_{n\to\infty} \mathrm{E}f(X_n) = \mathrm{E}f(X), \tag{7.3}$$

for all bounded continuous functions $f : \mathbf{X} \to \mathbf{R}$.

The following result of Lévy (1937) characterizes weak convergence on $\mathbf{R}$.

Theorem 7.2. *Let $\mu, \mu_1, \mu_2, \ldots$ denote probability measures on $(\mathbf{R}\,, \mathscr{B}(\mathbf{R}))$ with respective distibution functions $F, F_1, F_2, \ldots$. Then, $\mu_n \Rightarrow \mu$ if and only if*

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{7.4}$$

for all $x \in \mathbf{R}$ at which F is continuous.

Equivalently, $X_n \Rightarrow X$ if and only if $\mathrm{P}\{X_n \le x\} \to \mathrm{P}\{X \le x\}$ for all x such that $\mathrm{P}\{X = x\} = 0$.

Example 7.3. Consider the De Moivre–Laplace central limit theorem, and define

$$X_n := \frac{S_n - np}{\sqrt{np(1-p)}}. \tag{7.5}$$

Let F_n denote the distribution function of X_n, and F the distribution function of $N(0\,,1)$. Observe that: (i) F is continuous; and (ii) (7.1) asserts that $\lim_{n\to\infty}(F_n(b) - F_n(a)) = F(b) - F(a)$. By the preceding theorem, (7.1) is saying that $X_n \Rightarrow N(0\,,1)$.

Theorem 7.2 cannot be improved. Indeed, it can happen that $X_n \Rightarrow X$ but F_n fails to converge to F pointwise. Next is an example of this phenomenon.

Example 7.4. First let $X = \pm 1$ with probability $\frac{1}{2}$ each. Then define

$$X_n(\omega) := \begin{cases} -1 & \text{if } X(\omega) = -1, \\ 1 + \frac{1}{n} & \text{if } X(\omega) = 1. \end{cases} \tag{7.6}$$

Then, $\lim_{n\to\infty} f(X_n) = f(X)$ for all bounded continuous functions f, whence $\mathrm{E}f(X_n) \to \mathrm{E}f(X)$. However, $F_n(1) = \mathrm{P}\{X_n \le 1\} = \frac{1}{2}$ does not converge to $F(1) = \mathrm{P}\{X \le 1\} = 1$.

In order to prove Theorem 7.2 we will need the following.

Lemma 7.5. *The set $J := \{x \in \mathbf{R} : \mathrm{P}\{X = x\} > 0\}$ is denumerable.*

Proof. Define

$$J_n := \left\{ x \in \mathbf{R} : \; \mathrm{P}\{X = x\} \geq \frac{1}{n} \right\}. \tag{7.7}$$

Since $J = \cup_{n=1}^{\infty} J_n$, it suffices to prove that J_n is finite. Indeed, if J_n were infinite, then we could select a countable set $K_n \subset J_n$, and observe that

$$1 \geq \sum_{x \in K_n} \mathrm{P}\{X = x\} \geq \frac{|K_n|}{n}, \tag{7.8}$$

where $|\cdots|$ denotes cardinality. This contradicts the assumption that K_n is infinite. □

Proof of Theorem 7.2. Throughout, we let X_n denote a random variable whose distribution is μ_n ($n = 1, 2, \ldots$), and X a random variable with distribution μ.

Suppose first that $X_n \Rightarrow X$. For all fixed $x \in \mathbf{R}$ and $\epsilon > 0$, we can find a bounded continuous function $f : \mathbf{R} \to \mathbf{R}$ such that

$$f(y) \leq \mathbf{1}_{(-\infty, x]}(y) \leq f(y - \epsilon) \qquad {}^{\forall} y \in \mathbf{R}. \tag{7.9}$$

[Try a piecewise-linear function f.] It follows that

$$\mathrm{E} f(X_n) \leq F_n(x) \leq \mathrm{E} f(X_n - \epsilon). \tag{7.10}$$

Let $n \to \infty$ to obtain

$$\mathrm{E} f(X) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq \mathrm{E} f(X - \epsilon). \tag{7.11}$$

Equation (7.9) is equivalent to the following:

$$\mathbf{1}_{(-\infty, x-\epsilon]}(y) \leq f(y) \quad \text{and} \quad f(y - \epsilon) \leq \mathbf{1}_{(-\infty, x+\epsilon]}(y). \tag{7.12}$$

We apply this with $y := X$ and take expectations to see that

$$F(x - \epsilon) \leq \mathrm{E} f(X) \quad \text{and} \quad \mathrm{E} f(X - \epsilon) \leq F(x + \epsilon). \tag{7.13}$$

This and (7.11) together imply that

$$F(x - \epsilon) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x + \epsilon). \tag{7.14}$$

Let $\epsilon \downarrow 0$ to deduce that $F_n(x) \to F(x)$ whenever F is continuous at x.

For the converse we suppose that $F_n(x) \to F(x)$ for all continuity points x of F. Our goal is to prove that $\lim_{n \to \infty} \mathrm{E} f(X_n) = \mathrm{E} f(X)$ for all bounded continuous functions $f : \mathbf{R} \to \mathbf{R}$.

In accord with Lemma 7.5, for any $\delta, N > 0$, we can find real numbers $\cdots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \cdots$ (depending only on δ and N) such that: (i) $\max_{|i| \leq N} \sup_{y \in (x_i, x_{i+1}]} |f(y) - f(x_i)| \leq \delta$; (ii) F is continuous at x_i for all $i \in \mathbf{Z}$; and (iii) $F(x_{N+1}) \geq 1-\delta$ and $F(x_{-N}) \leq \delta$. Let $\Lambda_N := (x_{-N}, x_{N+1}]$. By (i),

$$\begin{aligned}
&\left| \mathrm{E}\left[f(X_n);\ X_n \in \Lambda_N\right] - \sum_{j=-N}^{N} f(x_j)\left[F_n(x_{j+1}) - F_n(x_j)\right] \right| \\
&\qquad = \left| \sum_{j=-N}^{N} \mathrm{E}\left\{f(X_n) - f(x_j);\ X_n \in (x_j\,, x_{j+1}]\right\} \right| \\
&\qquad \leq \sum_{j=-N}^{N} \mathrm{E}\left\{\left|f(X_n) - f(x_j)\right|;\ X_n \in (x_j\,, x_{j+1}]\right\} \\
&\qquad \leq \delta.
\end{aligned} \tag{7.15}$$

This remains valid if we replace X_n and F_n respectively by X and F. Note that N is held fixed, and F_n converges to F at all continuity-points of F. Therefore, as $n \to \infty$,

$$\sum_{|j|\leq N} f(x_j)\left[F_n(x_{j+1}) - F_n(x_j)\right] \to \sum_{|j|\leq N} f(x_j)\left[F(x_{j+1}) - F(x_j)\right]. \tag{7.16}$$

By the triangle inequality,

$$\limsup_{n\to\infty} \left|\mathrm{E}\left\{f(X_n);\ X_n \in \Lambda_N\right\} - \mathrm{E}\left\{f(X);\ X \in \Lambda_N\right\}\right| \leq 2\delta. \tag{7.17}$$

For the remainder terms, first note that

$$\mathrm{P}\left\{X_n \notin \Lambda_N\right\} = 1 - F_n\left(x_{N+1}\right) + F_n\left(x_N\right) \leq 2\delta. \tag{7.18}$$

Let $N \to \infty$ to find that the same quantity bounds $\mathrm{P}\{X \notin \Lambda_N\}$. Therefore, if we let $K := \sup_{y\in\mathbf{R}} |f(y)|$, then

$$\limsup_{n\to\infty} \mathrm{E}\left\{|f(X_n)|;\ X_n \notin \Lambda_N\right\} + \mathrm{E}\left\{|f(X)|;\ X \notin \Lambda_N\right\} \leq 4K\delta. \tag{7.19}$$

In conjunction with (7.17), this proves that

$$\limsup_{n\to\infty} \left|\mathrm{E}f(X_n) - \mathrm{E}f(X)\right| \leq 2\delta + 4K\delta. \tag{7.20}$$

Let δ tend to zero to finish. □

2. Weak Convergence and Compact-Support Functions

Definition 7.6. If $\mathbf{X}$ is a metric space, then $C_c(\mathbf{X})$ denotes the collection of all continuous functions $f : \mathbf{X} \to \mathbf{R}$ such that f has *compact support*; i.e., there exists a compact set K such that $f(x) = 0$ for all $x \notin K$. In addition, $C_b(\mathbf{X})$ denotes the collection of all bounded continuous functions $f : \mathbf{X} \to \mathbf{R}$.

Recall that in order to prove that $\mu_n \Rightarrow \mu$, we need to verify that $\int f\,d\mu_n \to \int f\,d\mu$ for all $f \in C_b(\mathbf{X})$. Since $C_c(\mathbf{R}^k) \subseteq C_b(\mathbf{R}^k)$, the next result simplifies our task in the case that $\mathbf{X} = \mathbf{R}^k$.

Theorem 7.7. *If $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k))$, then $\mu_n \Rightarrow \mu$ if and only if*

$$\lim_{n\to\infty} \int f\,d\mu_n = \int f\,d\mu \qquad \forall f \in C_c(\mathbf{R}^k). \tag{7.21}$$

Proof. We plan to prove that if $\int g\,d\mu_n \to \int g\,d\mu$ for all $g \in C_c(\mathbf{R}^k)$, then $\int f\,d\mu_n \to \int f\,d\mu$ for all $f \in C_b(\mathbf{R}^k)$. With this goal in mind, let us choose and fix such an $f \in C_b(\mathbf{R}^k)$. By considering f^+ and f^- separately, we can—and will—assume without loss of generality that $f(x) \geq 0$ for all x.

Step 1. The Lower Bound. For any $p > 0$ choose and fix a function $f_p \in C_c(\mathbf{R}^k)$ such that:

(1) For all $x \in [-p\,,p]^k$, $f_p(x) = f(x)$.
(2) For all $x \notin [-p-1\,,p+1]^k$, $f_p(x) = 0$.
(3) For all $x \in \mathbf{R}^k$, $0 \leq f_p(x) \leq f(x)$, and $f_p(x) \uparrow f(x)$ as $p \uparrow \infty$.

It follows that

$$\liminf_{n\to\infty} \int f\,d\mu_n \geq \lim_{n\to\infty} \int f_p\,d\mu_n = \int f_p\,d\mu. \tag{7.22}$$

Let $p \uparrow \infty$ and apply the dominated convergence theorem to deduce that

$$\liminf_{n\to\infty} \int f\,d\mu_n \geq \int f\,d\mu. \tag{7.23}$$

This proves half of the theorem.

Step 2. A Variant. In this step we prove that, in (7.23), f can be replaced by the indicator function of an *open* k-dimensional hypercube. More precisely, given any real numbers $a_1 < b_1, \ldots, a_k < b_k$,

$$\liminf_{n\to\infty} \mu_n\left((a_1\,,b_1) \times \cdots \times (a_k\,,b_k)\right) \geq \mu\left((a_1\,,b_1) \times \cdots \times (a_k\,,b_k)\right). \tag{7.24}$$

To prove this, we first find continuous functions $\psi_m \uparrow \mathbf{1}_{(a_1,b_1)\times\cdots\times(a_k,b_k)}$, pointwise. By definition, $\psi_m \in C_c(\mathbf{R}^k)$ for all $m \geq 1$, and

$$\liminf_{n\to\infty} \mu_n\left((a_1\,,b_1) \times \cdots \times (a_k\,,b_k)\right) \geq \lim_{n\to\infty} \int \psi_m\,d\mu_n = \int \psi_m\,d\mu. \tag{7.25}$$

Let $m \uparrow \infty$ to deduce (7.24) from the dominated convergence theorem.

Step 3. The Upper Bound. Recall f_p from Step 1 and write

$$\begin{aligned}\int f\,d\mu_n &= \int_{[-p,p]^k} f\,d\mu_n + \int_{\mathbf{R}^k\setminus[-p,p]^k} f\,d\mu_n \\ &\le \int f_p\,d\mu_n + \sup_{z\in\mathbf{R}^k}|f(z)|\cdot\left[1-\mu_n\left([-p\,,p]^k\right)\right].\end{aligned} \tag{7.26}$$

Now let $n\to\infty$ and appeal to (7.24) to find that

$$\limsup_{n\to\infty}\int f\,d\mu_n \le \int f_p\,d\mu + \sup_{z\in\mathbf{R}^k}|f(z)|\cdot\left[1-\mu\left((-p\,,p)^k\right)\right]. \tag{7.27}$$

Let $p\uparrow\infty$ and use the monotone convergence theorem to deduce that

$$\limsup_{n\to\infty}\int f\,d\mu_n \le \int f\,d\mu. \tag{7.28}$$

This finishes the proof. □

3. Harmonic Analysis in Dimension One

Definition 7.8. The *Fourier transform* of a probability measure μ on $\mathbf{R}$ is

$$\widehat{\mu}(t) := \int_{-\infty}^{\infty} e^{itx}\,\mu(dx) \qquad {}^\forall t\in\mathbf{R}, \tag{7.29}$$

where $i := \sqrt{-1}$. This definition continues to makes sense if μ is a finite measure. It also makes sense if μ is replaced by a Lebesgue-integrable function $f:\mathbf{R}\to\mathbf{R}$. In that case, we set

$$\widehat{f}(t) := \int_{-\infty}^{\infty} e^{ixt} f(x)\,dx \qquad {}^\forall t\in\mathbf{R}. \tag{7.30}$$

[We identify the Fourier transform of the function $f=(d\mu/dx)$ with that of the measure μ.] If X is a real-valued random variable whose distribution is some probability measure μ, then $\widehat{\mu}$ is also called the *characteristic function* of X and/or μ, and $\widehat{\mu}(t)$ is equal to $\mathrm{E}\exp(itX)=\mathrm{E}\cos(tX)+i\mathrm{E}\sin(tX)$.

Here are some of the elementary properties of characteristic functions.

Lemma 7.9. *If μ is a finite measure on $(\mathbf{R}\,,\mathscr{B}(\mathbf{R}))$, then $\widehat{\mu}$ exists, is uniformly continuous on $\mathbf{R}$, and satisfies the following:*

(1) $\sup_{t\in\mathbf{R}}|\widehat{\mu}(t)| = \widehat{\mu}(0)=\mu(\mathbf{R})$ *and* $\widehat{\mu}(-t)=\overline{\widehat{\mu}(t)}$.

(2) $\widehat{\mu}$ *is nonnegative definite. That is,* $\sum_{j=1}^n\sum_{k=1}^n \widehat{\mu}(t_j-t_k)z_j\overline{z_k}\ge 0$ *for all* $z_1,\ldots,z_n\in\mathbf{C}$ *and* $t_1,\ldots,t_n\in\mathbf{R}$.

Proof. Without loss of generality, we may assume that μ is a probability measure. Otherwise we can prove the theorem for the probability measure $\nu(\cdots)=\mu(\cdots)/\mu(\mathbf{R})$, and then multiply through by $\mu(\mathbf{R})$.

Let X be a random variable whose distribution is μ; $\widehat{\mu}(t) = \mathrm{E}e^{itX}$ is always defined and bounded since $|e^{itX}| \le 1$. To prove uniform continuity, we note that for all $a, b \in \mathbf{R}$,

$$\left|e^{ia} - e^{ib}\right| = \left|1 - e^{i(a-b)}\right| = \left|\int_0^{a-b} e^{ix}\,dx\right| \le |a-b|. \tag{7.31}$$

Consequently,

$$\left|e^{ia} - e^{ib}\right| \le |a-b| \wedge 2. \tag{7.32}$$

It follows from this that

$$\sup_{|s-t|\le\delta} |\widehat{\mu}(t) - \widehat{\mu}(s)| \le \sup_{|s-t|\le\delta} \mathrm{E}\left|e^{itX} - e^{isX}\right| \le \mathrm{E}\left(\delta|X| \wedge 2\right). \tag{7.33}$$

Thanks to the dominated convergence theorem, the preceding tends to 0 as δ converges down to 0. The uniform continuity of $\widehat{\mu}$ follows.

Part (1) is elementary. To prove (2) we first observe that

$$\sum\sum_{1\le j,k\le n} \widehat{\mu}(t_j - t_k) z_j \overline{z_k} = \sum\sum_{1\le j,k\le n} \mathrm{E}e^{i(t_j - t_k)X} z_j \overline{z_k}. \tag{7.34}$$

This is the expectation of $\left|\sum_{j=1}^n e^{it_jX} z_j\right|^2$, and hence is real as well as non-negative. □

Example 7.10 (§5.1, p. 11)**.** If $X = \mathrm{Unif}(a\,, b)$ for some $a < b$, then $\mathrm{E}e^{itX} = (e^{itb} - e^{ita})/it(b-a)$ for all $t \in \mathbf{R}$.

Example 7.11 (Problem 1.11, p. 13)**.** If X has the exponential distribution with some parameter $\lambda > 0$, then $\mathrm{E}e^{itX} = \lambda/(\lambda - it)$ for all $t \in \mathbf{R}$.

Example 7.12 (§5.2, p. 11)**.** If $X = N(\mu\,, \sigma^2)$ for some $\mu \in \mathbf{R}$ and $\sigma \ge 0$, then $\mathrm{E}e^{itX} = \exp\left(it\mu - \frac{1}{2}t^2\sigma^2\right)$ for all $t \in \mathbf{R}$.

Example 7.13 (§4.1, p. 8)**.** If $X = \mathrm{Bin}(n, p)$ for an integer $n \ge 1$ and some $p \in [0\,, 1]$, then $\mathrm{E}e^{itX} = (pe^{it} + 1 - p)^n$ for all $t \in \mathbf{R}$.

Example 7.14 (Problem 1.9, p. 13)**.** If $X = \mathrm{Poiss}(\lambda)$ for some $\lambda > 0$, then $\mathrm{E}e^{itX} = \exp(-\lambda + \lambda e^{it})$ for all $t \in \mathbf{R}$.

4. The Plancherel Theorem

In this section we state and prove a variant of a result of Plancherel (1910, 1933). Roughly speaking, Plancherel's theorem shows us how to reconstruct a distribution from its characteristic function. In order to state things more precisely we need some notation.

Definition 7.15. Suppose $f, g : \mathbf{R} \to \mathbf{R}$ are measurable. Then, when defined, the *convolution* $f * g$ is the function,

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)\,dy. \tag{7.35}$$

Convolution is a symmetric operation; i.e., $f*g = g*f$ for all measurable $f, g : \mathbf{R} \to \mathbf{R}$. This tacitly implies that one side of the stated identity converges if and only if the other side does. Next are two less obvious properties of convolutions. Henceforth, let ϕ_ϵ denote the density function of $N(0\,, \epsilon^2)$; i.e.,

$$\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\epsilon^2}\right) \qquad {}^\forall x \in \mathbf{R}. \tag{7.36}$$

The first important property of convolutions is that they provide us with smooth approximations to nice functions.

Fejér's Theorem. *If $f \in C_c(\mathbf{R})$, then $f * \phi_\epsilon$ is infinitely differentiable for all $\epsilon > 0$, and the kth derivative is $f * \phi_\epsilon^{(k)}$ for all $k \geq 1$. Moreover,*

$$\lim_{\epsilon\to 0} \sup_{x\in\mathbf{R}} |(f * \phi_\epsilon)(x) - f(x)| = 0. \tag{7.37}$$

Proof. Let $\phi_\epsilon^{(0)} := \phi_\epsilon$. Then for all $k \geq 0$ and all $\epsilon > 0$ fixed,

$$\begin{aligned} &\frac{\left(f * \phi_\epsilon^{(k)}\right)(x+h) - \left(f * \phi_\epsilon^{(k)}\right)(x)}{h} \\ &\qquad = \int_{-\infty}^{\infty} f(y)\, \frac{\phi_\epsilon^{(k)}(x+h-y) - \phi_\epsilon^{(k)}(x-y)}{h}\, dy. \end{aligned} \tag{7.38}$$

Because $\phi_\epsilon^{(k+1)}$ is bounded and f has compact support, the bounded convergence theorem implies that $f * \phi_\epsilon^{(k)}$ is differentiable, and the derivative is $f * \phi_\epsilon^{(k+1)}$. Now we apply induction to find that the kth derivative of $f * \phi_\epsilon$ exists and is equal to $f * \phi_\epsilon^{(k)}$ for all $k \geq 1$.

Let Z denote a standard normal random variable, and note that ϕ_ϵ is the density function of ϵZ; thus, $(f * \phi_\epsilon)(x) = \mathrm{E} f(x - \epsilon Z)$. By the uniform continuity of f, $\lim_{\epsilon\to 0} \sup_{x\in\mathbf{R}} |f(x-\epsilon Z) - f(x)| = 0$ a.s. Because f is bounded, this and the bounded convergence theorem together imply the result. □

The second property of convolutions, alluded to earlier, is the Plancherel theorem.

Plancherel's Theorem. *If μ is a finite measure on $\mathbf{R}$ and $f : \mathbf{R} \to \mathbf{R}$ is Lebesgue-integrable, then*

$$\int_{-\infty}^{\infty} (f * \phi_\epsilon)(x)\, \mu(dx) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2} \widehat{f}(t)\, \overline{\widehat{\mu}(t)}\, dt \qquad \forall \epsilon > 0. \tag{7.39}$$

Consequently, if $f \in C_c(\mathbf{R})$ and $\widehat{f} \in L^1(\mathbf{R})$, then

$$\int_{-\infty}^{\infty} f\, d\mu = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(t)\, \overline{\widehat{\mu}(t)}\, dt. \tag{7.40}$$

Proof. By the Fubini–Tonelli theorem,

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2} \widehat{f}(t)\, \overline{\widehat{\mu}(t)}\, dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2} \left(\int_{-\infty}^{\infty} f(x) e^{itx}\, dx\right)\left(\int_{-\infty}^{\infty} e^{-ity}\, \mu(dy)\right) dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2} e^{it(x-y)}\, dt\right) \mu(dy)\, f(x)\, dx. \end{aligned} \tag{7.41}$$

A direct calculation reveals that

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2} e^{it(x-y)}\, dt &= \frac{\sqrt{2\pi}}{\epsilon}\exp\left(-\frac{(x-y)^2}{2\epsilon^2}\right) \\ &= 2\pi\phi_\epsilon(x-y). \end{aligned} \tag{7.42}$$

See Example 7.12. Since f is integrable, all of the integrals in the right-hand side of (7.41) converge absolutely. Therefore, (7.39) follows from the Fubini–Tonelli theorem; (7.40) follows from (7.39) and the Fejér theorem. □

The Plancherel theorem is a deep result, and has a number of profound consequences. We state two of them.

The Uniqueness Theorem. *If μ and ν are two finite measures on $\mathbf{R}$ and $\widehat{\mu} = \widehat{\nu}$, then $\mu = \nu$.*

Proof. By the theorems of Plancherel and Fejér, $\int f\, d\mu = \int f\, d\nu$ for all $f \in C_c(\mathbf{R})$. Choose $f_k \in C_c(\mathbf{R})$ such that $f_k \downarrow \mathbf{1}_{[a,b]}$. The monotone convergence theorem then implies that $\mu([a\,,b]) = \nu([a\,,b])$. Thus, μ and ν agree on all finite unions of disjoint closed intervals of the form $[a\,,b]$. Because the said collection generates $\mathscr{B}(\mathbf{R})$, $\mu = \nu$ on $\mathscr{B}(\mathbf{R})$. □

The following convergence theorem of P. Lévy is another significant consequence of the Plancherel theorem.

The Convergence Theorem. *Suppose μ, μ_1, $\mu_2, \ldots$ are probability measures on $(\mathbf{R}\,, \mathscr{B}(\mathbf{R}))$. If $\lim_{n\to\infty} \widehat{\mu}_n = \widehat{\mu}$ pointwise, then $\mu_n \Rightarrow \mu$.*

Proof. In accord with Theorem 7.7 it suffices to prove that $\lim_{n\to\infty}\int f\,d\mu_n = \int f\,d\mu$ for all $f\in C_c(\mathbf{R})$. Thanks to the Fejér theorem, for all $\delta>0$ we can choose $\epsilon>0$ such that

$$\sup_{x\in\mathbf{R}} |(f*\phi_\epsilon)(x)-f(x)|\le\delta. \tag{7.43}$$

Apply the triangle inequality twice to see that for all $\delta>0$,

$$\begin{aligned}\left|\int f\,d\mu_n-\int f\,d\mu\right| &\le 2\delta+\left|\int (f*\phi_\epsilon)\,d\mu_n-\int (f*\phi_\epsilon)\,d\mu\right|\\ &=2\delta+\left|\int_{-\infty}^{\infty}\widehat{f}(t)e^{-\epsilon^2t^2/2}\left(\frac{\overline{\widehat{\mu_n}(t)}-\overline{\widehat{\mu}(t)}}{2\pi}\right)dt\right|.\end{aligned} \tag{7.44}$$

The last line holds by the Plancherel theorem. Since $f\in C_c(\mathbf{R})$, $\widehat{f}$ is uniformly bounded by $\int_{-\infty}^{\infty}|f(x)|\,dx<\infty$ (Lemma 7.9). Therefore, by the dominated convergence theorem,

$$\limsup_{n\to\infty}\left|\int f\,d\mu_n-\int f\,d\mu\right|\le 2\delta. \tag{7.45}$$

The theorem follows because $\delta>0$ is arbitrary. □

5. The 1-D Central Limit Theorem

We are ready to state and prove the main result of this chapter: The one-dimensional central limit theorem (CLT). The CLT is generally considered to be a cornerstone of classical probability theory.

The Central Limit Theorem. *Suppose $\{X_i\}_{i=1}^{\infty}$ are i.i.d., real-valued, and have two finite moments. If $S_n:=X_1+\cdots+X_n$ and $\mathrm{Var}X_1\in(0\,,\infty)$, then*

$$\frac{S_n-n\mathrm{E}X_1}{\sqrt{n}}\Rightarrow N(0\,,\mathrm{Var}X_1). \tag{7.46}$$

Because $n\mathrm{E}X_1+\sqrt{n}N(0\,,\mathrm{Var}X_1)$ and $N(n\mathrm{E}X_1\,,n\mathrm{Var}X_1)$ have the same distribution, the central limit theorem states that the distribution of S_n is close to that of $N(n\mathrm{E}X_1\,,n\mathrm{Var}X_1)$.

Proof. By considering instead $X_j^*:=(X_j-\mathrm{E}X_1)/\mathrm{SD}(X_1)$ and $S_n^*:=\sum_{j=1}^n X_j^*$ we can assume without loss of generality that the X_j's have mean zero and variance one.

We apply the Taylor expansion with remainder to deduce that for all $x\in\mathbf{R}$,

$$e^{ix}=1+ix-\frac{1}{2}x^2+R(x), \tag{7.47}$$

where $|R(x)| \le \frac{1}{6}|x|^3 \le |x|^3$. If $|x| \le 4$, then this is a good estimate, but when $|x| > 4$, we can use $|R(x)| \le |e^{ix}|+1+|x|+\frac{1}{2}x^2 \le x^2$ instead. Combine terms to obtain the bound:

$$|R(x)| \le |x|^3 \wedge x^2. \tag{7.48}$$

Because the X_j's are i.i.d., Lemma 6.12 on page 68 implies that

$$\mathrm{E}e^{itS_n/\sqrt{n}} = \prod_{j=1}^{n} \mathrm{E}e^{itX_j/\sqrt{n}}. \tag{7.49}$$

This and (7.47) together imply that

$$\begin{aligned} \mathrm{E}e^{itS_n/\sqrt{n}} &= \left(1 + i\mathrm{E}\left[\frac{tX_1}{\sqrt{n}}\right] - \frac{1}{2}\mathrm{E}\left[\frac{(tX_1)^2}{n}\right] + \mathrm{E}\left[R\left(\frac{tX_1}{\sqrt{n}}\right)\right]\right)^n \\ &= \left(1 - \frac{t^2}{2n} + \mathrm{E}\left[R\left(\frac{tX_1}{\sqrt{n}}\right)\right]\right)^n. \end{aligned} \tag{7.50}$$

By (7.48) and the dominated convergence theorem,

$$n\left|\mathrm{E}\left[R\left(tX_1/\sqrt{n}\right)\right]\right| \le \mathrm{E}\left[\frac{|tX_1|^3}{\sqrt{n}} \wedge (tX_1)^2\right] = o(1) \qquad (n \to \infty). \tag{7.51}$$

By the Taylor expansion $\ln(1-z) = -z + o(|z|)$ as $|z| \to 0$, where "ln" denotes the principal branch of the logarithm. It follows that

$$\lim_{n\to\infty} \mathrm{E}e^{itS_n/\sqrt{n}} = \lim_{n\to\infty}\left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = e^{-t^2/2}. \tag{7.52}$$

The CLT follows from the convergence theorem (p. 99) and Example 7.12 (p. 97). □

6. Complements to the CLT

6.1. The Multidimensional CLT. Now we turn our attention to the study of random variables in $\mathbf{R}^d$. Throughout, $X, X^1, X^2, \ldots$ are i.i.d. random variables that take values in $\mathbf{R}^d$, and $S_n := \sum_{i=1}^n X^i$. Our discussion is a little sketchy. But this should not cause too much confusion, since we encountered most of the key ideas earlier on in this chapter. Throughout this section, $\|x\|$ denotes the usual Euclidean norm of a variable $x \in \mathbf{R}^d$. That is,

$$\|x\| := \sqrt{x_1^2 + \cdots + x_d^2} \qquad {}^\forall x \in \mathbf{R}^d. \tag{7.53}$$

Definition 7.16. The *characteristic function* of X is the function $f(t) = \mathrm{E}e^{it\cdot X}$ where $t \cdot x = \sum_{i=1}^d t_i x_i$ for $t \in \mathbf{R}^d$. If μ denotes the distribution of X, then f is also written as $\widehat{\mu}$.

The following is the simplest analogue of the uniqueness theorem; it is an immediate consequence of the convergence theorem (p. 99).

Theorem 7.17. *If $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $(\mathbf{R}^d, \mathscr{B}(\mathbf{R}^d))$ and $\widehat{\mu_n} \to \widehat{\mu}$ pointwise, then $\mu_n \Rightarrow \mu$.*

This leads us to our next result.

Theorem 7.18. *Suppose $\{X^i\}_{i=1}^{\infty}$ are i.i.d. random variables in $\mathbf{R}^d$ with $\mathrm{E}X_1^i = \mu_i$, and $\mathrm{Cov}(X_1^i, X_1^j) = Q_{i,j}$ for an invertible $(d \times d)$ matrix $Q := (Q_{i,j})$. Then for all d-dimensional hypercubes $G = (a_1, b_1] \times \cdots \times (a_d, b_d]$,*

$$\lim_{n\to\infty} \mathrm{P}\left\{ \frac{S_n - n\mu}{\sqrt{n}} \in G \right\} = \int_G \frac{e^{-\frac{1}{2}y'Q^{-1}y}}{(2\pi)^{d/2}\sqrt{\det(Q)}}\, dy. \tag{7.54}$$

That is, $(S_n - n\mu)/\sqrt{n}$ converges weakly to a multidimensional Gaussian distribution with mean vector 0 and covariance matrix Q.

The preceding theorems are the natural d-dimensional extensions of their 1-D counterparts. On the other hand, the following is inherently multidimensional.

The Cramér–Wold Device. *$X_n \Rightarrow X$ if and only if $(t \cdot X_n) \Rightarrow (t \cdot X)$ for all $t \in \mathbf{R}^d$.*

If we were to prove that X_n converges weakly, then the Crámer–Wold device boils our task down to proving the weak convergence of the one-dimensional $(t \cdot X_n)$. But this needs to be proved for all $t \in \mathbf{R}^d$.

Proof. Suppose $X_n \Rightarrow X$, and choose and fix $f \in C_b(\mathbf{R}^d)$. Because $g_t(x) = t \cdot x$ is continuous, $\mathrm{E}f(g_t(X_n))$ converges to $\mathrm{E}f(g_t(X))$ as $n \to \infty$. This is half of the theorem.

Conversely, let μ_n and μ denote the distributions of X_n and X, respectively. Then $(t \cdot X_n) \Rightarrow (t \cdot X)$ for all $t \in \mathbf{R}^d$ iff $\widehat{\mu_n}(t) \to \widehat{\mu}(t)$. The converse now follows from Theorem 7.17. □

6.2. The Projective Central Limit Theorem. The *projective CLT* describes another natural way of arriving at the standard normal distribution. In kinetic theory this CLT implies that, for an ideal gas, all normalized *Gibbs states follow the standard normal distribution.* We are concerned only with the mathematical formulation of this CLT.

Definition 7.19. Define $\mathbf{S}^{n-1} := \{x \in \mathbf{R}^n : \|x\| = 1\}$ to be the *unit sphere* in $\mathbf{R}^n$. This is topologized by the relative topology in $\mathbf{R}^n$. That is, $U \subset \mathbf{S}^{n-1}$ is *open* in $\mathbf{S}^{n-1}$ iff U is an open subset of $\mathbf{R}^n$.

Recall that an $(n \times n)$ matrix M is a *rotation* if $M'M$ is the identity.

Definition 7.20. A measure μ on $\mathscr{B}(\mathbf{S}^{n-1})$ is called the *uniform distribution* on $\mathbf{S}^{n-1}$ if: (i) $\mu(\mathbf{S}^{n-1}) = 1$; and (ii) $\mu(A) = \mu(MA)$ for all $A \in \mathscr{B}(\mathbf{S}^{n-1})$ and all $(n \times n)$ rotation matrices M. If X is a random variable whose distribution is μ, then we say that X is *distributed uniformly* on $\mathbf{S}^{n-1}$. Item (ii) states that μ is *rotation invariant.*

Theorem 7.21. *If $X^{(n)}$ is distributed uniformly on $\mathbf{S}^{n-1}$, then*

$$\sqrt{n}\, X_1^{(n)} \Rightarrow N(0\,,1). \tag{7.55}$$

Remark 7.22. Without worrying too much about what this really means let X denote the first coordinate of a random variable that is distributed uniformly on the centered ball of radius $\sqrt{\infty}$ in $\mathbf{R}^\infty$. The projective CLT asserts that X is standard normal.

Before we prove Theorem 7.21 we need to demonstrate that there are, in fact, rotation-invariant probability measures on $\mathbf{S}^{n-1}$. The following is a special case of a more general result in abstract harmonic analysis.

Theorem 7.23. *For all $n \geq 1$ there exists a unique rotation-invariant probability measure on $\mathbf{S}^{n-1}$.*

Proof. Let $\{Z_i\}_{i=1}^\infty$ denote a sequence of i.i.d. standard normal random variables, and define $Z^{(n)} = (Z_1, \ldots, Z_n)$. We normalize the latter as follows:

$$X^{(n)} := \frac{Z^{(n)}}{\|Z^{(n)}\|} \qquad {}^\forall n \geq 1. \tag{7.56}$$

By independence, the characteristic function of $Z^{(n)}$ is $f(t) := \exp(-\|t\|^2/2)$. Because f is rotation-invariant, $Z^{(n)}$ and $MZ^{(n)}$ have the same characteristic function as long as M is an $(n \times n)$ rotation matrix. Consequently, $Z^{(n)}$ and $MZ^{(n)}$ have the same distribution for all rotations M; confer with the uniqueness theorem on page 99. It follows that the distribution of $X^{(n)}$ is rotation invariant, and hence the existence of a uniform distribution on $\mathbf{S}^{n-1}$ follows. Next we prove the more interesting uniqueness portion.

For all $\epsilon > 0$ and all sets $A \subseteq \mathbf{S}^{n-1}$ define $K_A(\epsilon)$ to be the largest number of disjoint open balls of radius ϵ that can fit inside A. By compactness, if A is closed then $K_A(\epsilon)$ is finite. The function K_A is known as *Kolmogorov ϵ-entropy*, *Kolmogorov complexity*, as well as the *packing number* of A.

Let μ and ν be two uniform probability measures on $\mathscr{B}(\mathbf{S}^{n-1})$. By the maximality condition in the definition of K_A, and by the rotational invariance of μ and ν, for all closed sets $A \subset \mathbf{S}^{n-1}$,

$$K_A(\epsilon)\mu(B_\epsilon) \leq \mu(A) \leq (K_A(\epsilon)+1)\mu(B_\epsilon), \tag{7.57}$$

where $B_\epsilon := \{x \in \mathbf{S}^{n-1} : \|x\| < \epsilon\}$. The preceding display remains valid if we replace μ by ν everywhere. Therefore, for all closed sets A that have positive ν-measure,

$$\left(\frac{K_A(\epsilon)}{K_A(\epsilon)+1}\right)\frac{\mu(A)}{\nu(A)} \le \frac{\mu(B_\epsilon)}{\nu(B_\epsilon)} \le \left(\frac{K_A(\epsilon)+1}{K_A(\epsilon)}\right)\frac{\mu(A)}{\nu(A)}. \tag{7.58}$$

Consequently,

$$\left|\frac{\mu(A)}{\nu(A)} - \frac{\mu(B_\epsilon)}{\nu(B_\epsilon)}\right| \le \frac{1}{K_A(\epsilon)}\frac{\mu(A)}{\nu(A)}. \tag{7.59}$$

We apply this with $A := \mathbf{S}^{n-1}$ to find that

$$\left|1 - \frac{\mu(B_\epsilon)}{\nu(B_\epsilon)}\right| \le \frac{1}{K_{\mathbf{S}^{n-1}}(\epsilon)}. \tag{7.60}$$

We plug this back in (7.59) to conclude that for all closed sets A with positive ν-measure,

$$\left|\frac{\mu(A)}{\nu(A)} - 1\right| \le \frac{1}{K_A(\epsilon)}\frac{\mu(A)}{\nu(A)} + \frac{1}{K_{\mathbf{S}^{n-1}}(\epsilon)} \qquad \forall \epsilon > 0. \tag{7.61}$$

As ϵ tends to zero, the right-hand side converges to zero. This implies that $\mu(A) = \nu(A)$ for all closed sets $A \in \mathscr{B}(\mathbf{S}^{n-1})$ that have positive ν-measure. Next, we reverse the roles of μ and ν to find that $\mu(A) = \nu(A)$ for all closed sets $A \in \mathscr{B}(\mathbf{S}^{n-1})$. Because closed sets generate all of $\mathscr{B}(\mathbf{S}^{n-1})$, the monotone class theorem (p. 30) implies that $\mu = \nu$. □

Proof of Theorem 7.21. We follow the proof of Theorem 7.23 closely, and observe that by the strong law of large numbers (p. 73), $\|Z^{(n)}\|/\sqrt{n} \to 1$ a.s. Therefore, $\sqrt{n}\, X_1^{(n)} \to Z_1$ a.s. The latter is standard normal. Since a.s.-convergence implies weak convergence, the theorem follows. □

6.3. The Replacement Method of Liapounov. There are other approaches to the CLT than the harmonic-analytic ones of the previous sections. In this section we present an alternative probabilistic method of Lindeberg (1922) who, in turn, used an ingenious "replacement method" of Liapounov (1900, pp. 362–364). This method makes clear the fact that the CLT is a *local phenomenon*. By this we mean that the structure of the CLT does not depend on the behavior of any fixed number of the increments.

In words, the method proceeds as follows: We estimate the distribution of S_n by replacing the increments, one at a time, by independent normal random variables. Then we use an idea of Lindeberg, and appeal to Taylor's theorem of calculus to keep track of the errors incurred by the replacement method.

As a nice by-product we obtain quantitative bounds on the error-rate in the CLT without further effort. To be concrete, we derive the following using the Liapounov method; the heart of the matter lies in its derivation.

Theorem 7.24. *Fix an integer $n \geq 1$, and suppose $\{X_i\}_{i=1}^n$ are independent mean-zero random variables in $L^3(\mathrm{P})$. Define $S_n := \sum_{i=1}^n X_i$ and $s_n^2 := \mathrm{Var} S_n$. Then for any three times continuously differentiable function f,*

$$\left|\mathrm{E}f(S_n) - \mathrm{E}f\big(N(0\,,s_n^2)\big)\right| \leq \frac{2M_f}{3\sqrt{\pi/2}} \sum_{i=1}^n \|X_i\|_3^3, \tag{7.62}$$

provided that $M_f := \sup_z |f'''(z)|$ is finite.

Proof. Let σ_i^2 denote the variance of X_i for all $i = 1, \ldots, n$, so that $s_n^2 = \sum_{i=1}^n \sigma_i^2$. By Taylor expansion,

$$\left| f(S_n) - f(S_{n-1}) - X_n f'(S_{n-1}) - \frac{X_n^2}{2} f''(S_{n-1}) \right| \leq \frac{M_f}{6} |X_n|^3. \tag{7.63}$$

Because $\mathrm{E}X_n = 0$ and $\mathrm{E}[X_n^2] = \sigma_n^2$, the independence of the X's implies that

$$\left| \mathrm{E}f(S_n) - \mathrm{E}f(S_{n-1}) - \frac{\sigma_n^2}{2} \mathrm{E}f''(S_{n-1}) \right| \leq \frac{M_f}{6} \|X_n\|_3^3. \tag{7.64}$$

Next consider a normal random variable Z_n that has the same mean and variance as X_n, and is independent of $X_1, \ldots, X_n$. If we apply (7.64), but replace X_n by Z_n, then we obtain

$$\left| \mathrm{E}f(S_{n-1} + Z_n) - \mathrm{E}f(S_{n-1}) - \frac{\sigma_n^2}{2} \mathrm{E}f''(S_{n-1}) \right| \leq \frac{M_f}{6} \|Z_n\|_3^3. \tag{7.65}$$

This and (7.64) together yield

$$|\mathrm{E}f(S_n) - \mathrm{E}f(S_{n-1} + Z_n)| \leq \frac{M_f}{6} \left(\|Z_n\|_3^3 + \|X_n\|_3^3 \right). \tag{7.66}$$

A routine computation reveals that $\|Z_n\|_3^3 = A\sigma_n^3$, where $A := \mathrm{E}\{|N(0\,,1)|^3\} = 2/\sqrt{\pi/2} > 1$. Since $\sigma_n^3 \leq \|X_n\|_3^3$ (Proposition 4.16, p. 42), we find that

$$|\mathrm{E}f(S_n) - \mathrm{E}f(S_{n-1} + Z_n)| \leq \frac{2M_f}{3\sqrt{\pi/2}} \|X_n\|_3^3. \tag{7.67}$$

Now we iterate this procedure: Bring in an independent normal Z_{n-1} with the same mean and variance as X_{n-1}. Replace X_{n-1} by Z_{n-1} in (7.67) to find that

$$|\mathrm{E}f(S_n) - \mathrm{E}f(S_{n-2} + Z_{n-1} + Z_n)| \leq \frac{2M_f}{3\sqrt{\pi/2}} \left(\|X_{n-1}\|_3^3 + \|X_n\|_3^3 \right). \tag{7.68}$$

Next replace X_{n-2} by another independent normal Z_{n-2}, etc. After n steps, we arrive at

$$\left|\mathrm{E}f(S_n) - \mathrm{E}f\left(\sum_{i=1}^n Z_i\right)\right| \leq \frac{2M_f}{3\sqrt{\pi/2}} \sum_{i=1}^n \|X_i\|_3^3. \tag{7.69}$$

The theorem follows because $\sum_{i=1}^n Z_i = N(0\,,s_n^2)$; see Problem 7.18. □

To understand how this can be used suppose $\{X_i\}_{i=1}^n$ are i.i.d., with mean zero and variance σ^2. We can then apply Theorem 7.24 with $f(x) := g(x\sqrt{n})$ to deduce the following.

Corollary 7.25. *If $\{X_i\}_{i=1}^n$ are i.i.d. with mean zero, variance σ^2, and three bounded moments, then for all three times continuously differentiable functions g,*

$$\left|\mathrm{E}g(S_n/\sqrt{n}) - \mathrm{E}g(N(0,\sigma^2))\right| \le \frac{A}{\sqrt{n}}, \tag{7.70}$$

where $A := 2\sup_z |g'''(z)| \cdot \|X_1\|_3^3/(3\sqrt{\pi/2})$.

We let $g(x) := e^{itx}$, and extend the preceding to complex-valued functions in the obvious way to obtain the central limit theorem (p. 100) under the extra condition that $X_1 \in L^3(\mathrm{P})$. Moreover, when $X_1 \in L^3(\mathrm{P})$ we find that the rate of convergence to the CLT is of the order $n^{-1/2}$.

Theorem 7.24 is not restricted to increments that are in $L^3(\mathrm{P})$. For the case where $X_1 \in L^{2+\rho}(\mathrm{P})$ for some $\rho \in (0\,,1)$ see Problem 7.44. Even when $X_1 \in L^2(\mathrm{P})$ only, Theorem 7.24 can be used to prove the CLT, viz.,

Lindeberg's Proof of the CLT. Without loss of generality, we may assume that $\mu = 0$ and $\sigma = 1$. Choose and fix $\epsilon > 0$, and define $X_i' := X_i \mathbf{1}_{\{|X_i| \le \epsilon\sqrt{n}\}}$, $S_n' := \sum_{i=1}^n X_i'$, $\mu_n := \mathrm{E}S_n'$, and $s_n^2 := \mathrm{Var}S_n'$.

Choose and fix a function $g : \mathbf{R} \to \mathbf{R}$ such that g and its first three derivatives are bounded and continuous. According to Theorem 7.24,

$$\begin{aligned}\left|\mathrm{E}g\left(\frac{S_n' - \mu_n}{\sqrt{n}}\right) - \mathrm{E}g\left(N\left(0\,, \frac{s_n^2}{n}\right)\right)\right| &\le \frac{2M_g}{3\sqrt{\pi n/2}}\mathrm{E}\left(\left|X_1' - \mathrm{E}X_1'\right|^3\right) \\ &\le \frac{32M_g}{3\sqrt{\pi n/2}}\|X_1'\|_3^3.\end{aligned} \tag{7.71}$$

The last line follows from the inequality $|a+b|^3 \le 8(|a|^3 + |b|^3)$ and the fact that $\|X_1'\|_1 \le \|X_1'\|_3$ (Proposition 4.16, p. 42). Because $|X_1'|$ is bounded above by $\epsilon\sqrt{n}$,

$$\|X_1'\|_3^3 \le \epsilon\sqrt{n}\,\mathrm{E}\left(|X_1'|^2\right) \le \epsilon\sqrt{n}\,\mathrm{E}\left(X_1^2\right) = \epsilon\sqrt{n}. \tag{7.72}$$

Consequently,

$$\left|\mathrm{E}g\left(\frac{S_n' - \mu_n}{\sqrt{n}}\right) - \mathrm{E}g\left(N\left(0\,, \frac{s_n^2}{n}\right)\right)\right| \le \frac{32M_g}{3\sqrt{\pi/2}}\epsilon := A\epsilon. \tag{7.73}$$

A one-term Taylor expansion simplifies the first term as follows:

$$\begin{aligned}\left|\mathrm{E}g\left(\frac{S_n'-\mu_n}{\sqrt{n}}\right)-\mathrm{E}g\left(\frac{S_n}{\sqrt{n}}\right)\right| &\le \sup_z|g'(z)|\ \mathrm{E}\left|\frac{S_n'-S_n-\mu_n}{\sqrt{n}}\right|\\ &\le \sup_z|g'(z)|\ \frac{\mathrm{SD}(S_n-S_n')}{\sqrt{n}}.\end{aligned} \tag{7.74}$$

Since $S_n - S_n' = \sum_{i=1}^n X_i \mathbf{1}_{\{|X_i|\ge\epsilon\sqrt{n}\}}$ is a sum of n i.i.d. random variables,

$$\mathrm{Var}(S_n-S_n') = n\mathrm{Var}\left(X_1\mathbf{1}_{\{|X_1|>\epsilon\sqrt{n}\}}\right) \le n\mathrm{E}\left[X_1^2;\ |X_1|>\epsilon\sqrt{n}\right]. \tag{7.75}$$

Therefore,

$$\left|\mathrm{E}g\left(\frac{S_n}{\sqrt{n}}\right)-\mathrm{E}g\left(N\left(0\,,\frac{s_n^2}{n}\right)\right)\right| \le A\epsilon+\sqrt{\mathrm{E}\left[X_1^2;\ |X_1|>\epsilon\sqrt{n}\right]}. \tag{7.76}$$

Now, $s_n^2/n = \mathrm{Var}(S_n')/n = \mathrm{Var}(X_1\mathbf{1}_{\{|X_1|>\epsilon\sqrt{n}\}})$. By the dominated convergence theorem, this converges to $\mathrm{Var}X_1 = 1$ as $n\to\infty$. Therefore by scaling (Problem 1.14, p. 14),

$$\mathrm{E}g\left(N\left(0\,,\frac{s_n^2}{n}\right)\right) = \mathrm{E}g\left(N(0\,,1)\,\frac{s_n}{\sqrt{n}}\right)\to \mathrm{E}g(N(0\,,1)), \tag{7.77}$$

as $n\to\infty$. This, the continuity of g, and (7.76), together yield

$$\limsup_{n\to\infty}\left|\mathrm{E}g\left(S_n/\sqrt{n}\right)-\mathrm{E}g(N(0\,,1))\right|\le A\epsilon. \tag{7.78}$$

Because the left-hand side is independent of ϵ, it must therefore be equal to zero. It follows that $\mathrm{E}g(S_n/\sqrt{n})\to\mathrm{E}g(N(0\,,1))$ if g and its first three derivatives are continuous and bounded.

Now suppose $\psi\in C_c(\mathbf{R})$ is fixed. By Fejér's theorem (p. 98), for all $\delta>0$ we can find g such that g and its first three derivatives are bounded and continuous, and $\sup_z|g(z)-\psi(z)|\le\delta$. Because δ is arbitrary, the triangle inequality and what we have proved so far together prove that $\mathrm{E}\psi(S_n/\sqrt{n})\to\mathrm{E}\psi(N(0\,,1))$. This is the desired result. □

6.4. Cramér's Theorem. In this section we use characteristic function methods to prove the following striking theorem of Cramér (1936). This section requires only a rudimentary knowledge of complex analysis.

Theorem 7.26. *Suppose X_1 and X_2 are independent real-valued random variables such that X_1+X_2 is a possibly degenerate normal random variable. Then X_1 and X_2 are possibly degenerate normal random variables too.*

Remark 7.27. Cramér's theorem states that if μ_1 and μ_2 are probability measures such that $\widehat{\mu_1}(t)\widehat{\mu_2}(t) = e^{i\mu t-\sigma^2t^2}$ $(\mu\in\mathbf{R},\sigma\ge 0)$, then μ_1 and μ_2 are Gaussian probability measures.

Remark 7.28. Cramér's theorem does not rule out the possibility that one or both of the X_i's are constants. It might help to recall our convention that $N(\mu\,,0)=\mu$.

We prove Cramér's theorem by first deriving three elementary lemmas from complex analysis, and one from probability. Recall that a function $f:\mathbf{C}\to\mathbf{C}$ is *entire* function if is is analytic on $\mathbf{C}$.

Lemma 7.29 (The Liouville Theorem)**.** *Suppose $f:\mathbf{C}\to\mathbf{C}$ is an entire function, and there exists an integer $n\ge 0$ such that*

$$|f(z)|=O\left(|z|^n\right)\qquad \text{as } |z|\to\infty. \tag{7.79}$$

Then there exist $a_0,\ldots,a_n\in\mathbf{C}$ such that $f(z)=\sum_{j=0}^n a_jz^j$ on $\mathbf{C}$.

Remark 7.30. When $n=0$, Lemma 7.29 asserts that *bounded entire functions are constants.* This is the more usual form of the Liouville theorem.

Proof. For any $z_0\in\mathbf{C}$ and $\rho>0$, define $\gamma(\theta):=z_0+\rho e^{i\theta}$ for all $\theta\in(0\,,2\pi]$. By the Cauchy integral formula on circles, for any $n\ge 0$, the nth derivative $f^{(n)}$ is analytic and satisfies

$$\begin{aligned} f^{(n+1)}(z_0) &= \frac{(n+1)!}{2\pi i}\int_\gamma \frac{f(z)}{(z-z_0)^{n+2}}\,dz \\ &= \frac{(n+1)!}{2\pi i\rho^{n+1}}\int_0^{2\pi}\frac{f\left(z_0+\rho e^{i\theta}\right)}{e^{i(n+2)\theta}}\,d\theta. \end{aligned} \tag{7.80}$$

Since f is continuous, (7.79) tells us that there exists a constant $A>0$ such that $|f(z_0+\rho e^{i\theta})|\le A\rho^n$ for all $\rho>0$ sufficiently large and all $\theta\in[0\,,2\pi)$. In particular, $|f^{(n+1)}(z_0)|\le(n+1)!A\rho^{-1}$. Because this holds for all large $\rho>0$, $f^{(n+1)}(z_0)=0$ for all $z_0\in\mathbf{C}$, whence follows the result. □

Lemma 7.31 (Schwarz)**.** *Choose and fix $A,\rho>0$. Suppose f is analytic on $B_\rho:=\{w\in\mathbf{C}:|w|<\rho\}$, $f(0)=0$, and $\sup_{z\in B_\rho}|f(z)|\le A$. Then,*

$$|f(z)|\le\frac{A|z|}{\rho}\qquad \text{on } B_\rho. \tag{7.81}$$

Proof. Define

$$F(z):=\begin{cases} f(z)/z & \text{if } z\neq 0,\\ f'(0) & \text{if } z=0.\end{cases} \tag{7.82}$$

Evidently, F is analytic on B_ρ. According to the *maximum principle*, an analytic function in a given domain attains its maximum on the boundary of the domain. Therefore, whenever $r\in(0\,,\rho)$, it follows that

$$|F(z)|\le\sup_{|w|=r}|F(w)|\le\frac{A}{r}\qquad \forall|z|<r. \tag{7.83}$$

Let r converge upward to ρ to finish. □

The following is our final requirement from complex analysis.

Lemma 7.32 (Borel and Carathéodory)**.** *If $f : \mathbf{C} \to \mathbf{C}$ is entire, then*

$$\sup_{|z|\le r/2} |f(z)| \le 4 \sup_{|z|\le r} |\mathrm{Re}\, f(z)| + 5|f(0)| \qquad \forall r > 0. \tag{7.84}$$

Proof. Let $g(z) := f(z) - f(0)$, so that g is entire and $g(0) = 0$. Define $R(r) := \sup_{|z|\le r} |\mathrm{Re}\, g(z)|$ for all $r > 0$, and consider the function

$$T(w) := \frac{w}{2R(r) - w} \qquad \forall |w| \le R(r). \tag{7.85}$$

Evidently,

$$g(z) = 2R(r)\, \frac{T(g(z))}{1 + T(g(z))}. \tag{7.86}$$

One can check directly that $|T(f(z))| \le 1$ for all $z \in B_r$, and hence $T \circ g$ is analytic on B_r. Because $T(g(0)) = 0$, Lemma 7.31 implies that $|T(g(z))| \le |z|/r$ for all $z \in B_r$. It follows that for all $z \in B_r$,

$$|g(z)| \le 2R(r) \frac{|z|/r}{1 - (|z|/r)}. \tag{7.87}$$

This proves that, $|g(z)| \le 4R(r)$, uniformly for $|z| \le r/2$, and hence,

$$\sup_{|z|\le r/2} |f(z) - f(0)| \le 4 \sup_{|z|\le r} |\mathrm{Re}\, f(z) - \mathrm{Re}\, f(0)|. \tag{7.88}$$

The lemma follows from this and the triangle inequality. □

Finally, we need a preparatory lemma from probability.

Lemma 7.33. *If $V \ge 0$ a.s., then for any $a > 0$,*

$$\mathrm{E}e^{aV} = 1 + a \int_0^\infty e^{ax} \mathrm{P}\{V \ge x\}\, dx. \tag{7.89}$$

In particular, suppose U is non-negative, and there exists $r \ge 1$ such that

$$\mathrm{P}\{V \ge x\} \le r\mathrm{P}\{U \ge x\} \qquad \forall x > 0. \tag{7.90}$$

Then, $\mathrm{E}e^{aV} \le r\mathrm{E}e^{aU}$ for all $a > 0$.

Proof. Because $e^{aV(\omega)} = 1 + a \int_0^\infty \mathbf{1}_{\{V(\omega)\ge x\}} e^{ax}\, dx$ and the integrand is non-negative, we can take expectations and use Fubini–Tonelli to deduce (7.89). Because $r \ge 1$, the second assertion is a ready corollary of the first □

Proof of Theorem 7.26. Throughout, let $Z := X_1 + X_2$; Z is normally distributed. We can assume without loss of generality that $\mathrm{E}Z = 0$; else we consider $Z - \mathrm{E}Z$ in place of Z. The proof is now carried out in two natural steps.

Step 1. Identifying the Modulus. We begin by finding the form of $\left|\mathrm{E}e^{itX_k}\right|$ for $k = 1, 2$.

Because $\mathrm{E}Z = 0$, there exists $\sigma \geq 0$ such that $\mathrm{E}\exp(zZ) = \exp(z^2\sigma^2)$ for all $z \in \mathbf{C}$. Since $|Z| \geq |X_1| - |X_2|$, if $|X_1| \geq \lambda$ and $|X_2| \leq m$ then $|Z| \geq \lambda - m$. Therefore, by independence,

$$\begin{aligned}\mathrm{P}\{|Z| \geq \lambda - m\} &\geq \mathrm{P}\{|X_1| \geq \lambda\}\,\mathrm{P}\{|X_2| \leq m\} \\ &\geq \frac{1}{4}\mathrm{P}\{|X_1| \geq \lambda\},\end{aligned} \tag{7.91}$$

provided that we choose a sufficiently large m. Choose and fix such an m.

In accord with Lemma 7.33, $\mathrm{E}e^{c|X_1|} \leq 4e^{cm}\mathrm{E}e^{c|Z|}$ for all $c > 0$. But

$$\mathrm{E}e^{c|Z|} \leq \mathrm{E}e^{cZ} + \mathrm{E}e^{-cZ} \leq 2e^{c^2\sigma^2} \qquad \forall c > 0. \tag{7.92}$$

Consequently,

$$\left|\mathrm{E}e^{zX_1}\right| \leq \mathrm{E}e^{|z|\cdot|X_1|} \leq 8\exp\left(|z|m + \sigma^2|z|^2\right) \qquad \forall z \in \mathbf{C}. \tag{7.93}$$

Because $|Z| \geq |X_2| - |X_1|$, the same bound holds if we replace X_1 by X_2 everywhere. This proves that $f_k(z) := \mathrm{E}\exp(zX_k)$ exists for all $z \in \mathbf{C}$, and defines an entire function (why?).

To summarize, $\mathbf{R} \ni t \mapsto f_k(it)$ is the characteristic function of X_k, and

$$|f_k(z)| \leq 8\exp\left(|z|m + \sigma^2|z|^2\right) \qquad \forall z \in \mathbf{C}, k = 1, 2. \tag{7.94}$$

Because $f_1(z)f_2(z) = \mathrm{E}\exp(zZ) = \exp(z^2\sigma^2)$, (7.94) implies that for all $z \in \mathbf{C}$ and $k = 1, 2$,

$$8\exp\left(|z|m + \sigma^2|z|^2\right)|f_k(z)| \geq |\exp(z^2\sigma^2)| \geq \exp\left(-|z|^2\sigma^2\right). \tag{7.95}$$

It follows from this and (7.94) that for all $z \in \mathbf{C}$ and $k = 1, 2$,

$$\frac{1}{8}\exp\left(-|z|m - 2\sigma^2|z|^2\right) \leq |f_k(z)| \leq 8\exp\left(|z|m + \sigma^2|z|^2\right). \tag{7.96}$$

Consequently, $\ln|f_k|$ is an entire function that satisfies the growth condition (7.79) of Lemma 7.29 with $n = 2$, and hence,

$$|f_1(z)| = \exp\left(a_0 + a_1z + a_2z^2\right) \qquad \forall z \in \mathbf{C}. \tag{7.97}$$

A similar expression holds for $|f_2(z)|$.

Step 2. Estimating the Imaginary Part. Because f_k is non-vanishing and entire, we can write

$$f_k(z) = \exp(g_k(z)), \tag{7.98}$$

where g_k is entire for $k = 1, 2$. To prove this we first note that f_k'/f_k is entire, and therefore so is

$$g_k(z) := \int_0^z \frac{f_k'(w)}{f_k(w)}\,dw. \tag{7.99}$$

Next we compute directly to find that $(e^{-g_k} f_k)'(z) = 0$ for all $z \in \mathbf{C}$. Because $f_k(0) = 1$ and $g_k(0) = 0$, it follows that $f_k(z) = \exp(g_k(z))$, as asserted.

It follows then that $|f_k(z)| = \exp(\operatorname{Re} g_k(z))$, and Step 1 implies that $\operatorname{Re} g_k$ is a complex quadratic polynomial for $k = 1, 2$. Thanks to this and Lemma 7.32, we can deduce that the entire function g_k satisfies (7.79) with $n = 2$. Therefore, by Liouville's theorem, $g_k(z) = \alpha_k + \beta_k z + \gamma_k z^2$ where $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2$ are complex numbers. Consequently,

$$\mathrm{E}e^{itX_k} = f_k(it) = \exp\left(\alpha_k + it\beta_k - t^2\gamma_k\right) \qquad \forall t \in \mathbf{R},\ k = 1, 2. \tag{7.100}$$

Plug in $t = 0$ to find that $\alpha_k = 0$. Also part (1) of Lemma 7.9 implies that $f_k(-it)$ is the complex conjugate of $f_k(it)$. We can write this out to find that

$$\exp(-it\beta_k - t^2\gamma_k) = \exp(-it\overline{\beta_k} - t^2\overline{\gamma_k}) \qquad \forall t \in \mathbf{R}. \tag{7.101}$$

This proves that

$$it\beta_k - t^2\gamma_k = it\overline{\beta_k} - t^2\overline{\gamma_k} + 2\pi i N(t), \tag{7.102}$$

where $N(t)$ is integer-valued for every $t \in \mathbf{R}$. All else being continuous, this proves that N is a continuous integer-valued function. Therefore, $N(t) = N(0) = 0$, and so it follows from the preceding display that β_k and γ_k are real-valued. Because $|f_k(it)| \le 1$, we have also that $\gamma_k \ge 0$. The result follows from these calculations. □

Problems

7.1. Define $C_c^\infty(\mathbf{R}^k)$ to be the collection of all infinitely differentiable functions $f : \mathbf{R}^k \to \mathbf{R}$ that have compact support. If $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k))$, then prove that $\mu_n \Rightarrow \mu$ iff $\int f\, d\mu_n \to \int f\, d\mu$ for all $f \in C_c^\infty(\mathbf{R}^k)$.

7.2. If $\mu, \mu_1, \mu_2, \ldots, \mu_n$ is a sequence of probability measures on $(\mathbf{R}^d, \mathscr{B}(\mathbf{R}^d))$, then show that the following are characteristic functions of probability measures:

(1) $\overline{\widehat{\mu}}$;
(2) $\operatorname{Re} \widehat{\mu}$,
(3) $|\widehat{\mu}|^2$;
(4) $\prod_{j=1}^n \widehat{\mu}_j$; and
(5) $\sum_{j=1}^n p_j \widehat{\mu_j}$, where $p_1, \ldots, p_n \ge 0$ and $\sum_{j=1}^n p_j = 1$.

Also prove that $\overline{\widehat{\mu}(\xi)} = \widehat{\mu}(-\xi)$. Consequently, if μ is a *symmetric measure* (i.e., $\mu(-A) = \mu(A)$ for all $A \in \mathscr{B}(\mathbf{R}^d)$) then $\widehat{\mu}$ is a real-valued function.

7.3. Use characteristic functions to derive Problem 1.17 on page 14. Apply this to prove that if $X = \mathrm{Unif}[-1\,,1]$, then we can write it as

$$X := \sum_{j=1}^\infty \frac{X_j}{2^j}, \tag{7.103}$$

where the X_j's are i.i.d., taking the values ± 1 with probability $\frac{1}{2}$ each.

7.4 (Problem 7.3, continued)**.** Prove that

$$\frac{\sin x}{x} = \prod_{k=1}^{\infty} \cos\left(\frac{x}{2^k}\right) \qquad \forall x \in \mathbf{R} \setminus \{0\}. \tag{7.104}$$

By continuity, this is true also for $x = 0$.

7.5. Let X and Y denote two random variables on the same probability space. Suppose that $X + Y$ and $X - Y$ are independent standard-normal random variables. Then prove that X and Y are independent normal random variables. You may not use Theorem 7.26 or its proof.

7.6. Suppose X_1 and X_2 are independent random variables. Use characteristic functions to prove that:

(1) If $X_i = \text{Bin}(n_i\,, p)$ for the same $p \in [0\,,1]$, then $X_1 + X_2 = \text{Bin}(n_1 + n_2\,, p)$.
(2) If $X_i = \text{Poiss}(\lambda_i)$, then $X_1 + X_2 = \text{Poiss}(\lambda_1 + \lambda_2)$.
(3) If $X_i = N(\mu_i\,, \sigma_i^2)$, then $X_1 + X_2 = N(\mu_1 + \mu_2\,, \sigma_1^2 + \sigma_2^2)$.

7.7. Let X have the gamma distribution with parameters $(\alpha\,,\lambda)$. Compute, carefully, the characteristic function of X. Use it to prove that if $X_1, X_2, \ldots$ are i.i.d. exponential random variables with parameter λ each, then $S_n := X_1 + \cdots + X_n$ has a gamma distribution. Identify the latter distribution's parameters.

7.8. Let f be a symmetric and bounded probability density function on $\mathbf{R}$. Suppose there exists $C > 0$ and $\alpha \in (0\,,1]$ such that

$$f(x) \sim C|x|^{-(1+\alpha)} \qquad \text{as } |x| \to \infty. \tag{7.105}$$

Prove that

$$\widehat{f}(t) = 1 - D|t|^{\alpha} + o(|t|^{\alpha}) \qquad \text{as } |t| \to 0, \tag{7.106}$$

and compute D. Check also that $D < \infty$. What happens if $\alpha > 1$?

7.9 (Lévy's Concentration Inequality)**.** Prove that if μ is a probability measure on the line, then

$$\mu\left(\left\{x : |x| > \frac{1}{\epsilon}\right\}\right) \leq \frac{7}{\epsilon}\int_0^{\epsilon} (1 - \text{Re}\,\widehat{\mu}(t))\,dt \qquad \forall \epsilon > 0. \tag{7.107}$$

(Hint: Start with the right-hand side.)

7.10 (Fourier Series)**.** Suppose X is a random variable that takes values in $\mathbf{Z}^d$ and has mass function $p(x) = \text{P}\{X = x\}$. Define $\widehat{p}(t) = \text{E}e^{it\cdot X}$, and derive the following *inversion formula*:

$$p(x) = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} \exp(-it\cdot x)\,\widehat{p}(t)\,dt \qquad \forall x \in \mathbf{Z}^d. \tag{7.108}$$

Is the latter identity valid for all $x \in \mathbf{R}^d$?

7.11. Derive the following variant of Plancherel's theorem (p. 99): For any $a < b$ and all probability measures μ on $(\mathbf{R}\,, \mathscr{B}(\mathbf{R}))$,

$$\lim_{\epsilon\downarrow 0}\frac{1}{2\pi i}\int_{-\infty}^{\infty} e^{-\epsilon^2 t^2/2}\left(\frac{e^{-ita} - e^{-itb}}{t}\right)\widehat{\mu}(t)\,dt = \mu\left((a\,,b)\right) + \frac{\mu(\{a\}) + \mu(\{b\})}{2}. \tag{7.109}$$

7.12 (Inversion Theorem)**.** Derive the *inversion theorem*: If μ is a probability measure on $\mathscr{B}(\mathbf{R}^k)$ such that $\widehat{\mu}$ is integrable $[dx]$, then μ is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^k$. Moreover, then μ has a uniformly continuous density function f, and

$$f(x) = \frac{1}{(2\pi)^k}\int_{\mathbf{R}^k} e^{-it\cdot x}\widehat{f}(t)\,dt \qquad \forall x \in \mathbf{R}^k. \tag{7.110}$$

7.13 (The Triangular Distribution)**.** Consider the density function $f(x) := (1 - |x|)^+$ for $x \in \mathbf{R}$. If the density function of X is f, then compute the characteristic function of X. Prove that f itself is the characteristic function of a probability measure. (Hint: Problem 7.12.)

7.14. Suppose f is a probability density function on $\mathbf{R}$; i.e., $f \geq 0$ a.e. and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

(1) We say that f is of *positive type* if $\widehat{f}$ is non-negative and integrable. Prove that if f is of positive type, then $f(x) \le f(0)$ for all $x \in \mathbf{R}$.

(2) Prove that if f is of positive type, then $g(x) := \widehat{f}(x)/(2\pi f(0))$ is a density function, and $\widehat{g}(t) = f(t)/f(0)$. (HINT: Problem 7.12.)

(3) Compute the characteristic function of $g(x) = \frac{1}{2}\exp(-|x|)$. Use this to conclude that $f(x) := \pi^{-1}(1+x^2)^{-1}$ is a probability density function whose characteristic function is $\widehat{f}(t) = \exp(-|t|)$. The function f defines the so-called *Cauchy density function.* [Alternatively, you may use contour integration to arrive at the end result.]

7.15 (Riemann–Lebesgue lemma)**.** Prove that $\lim_{|t|\to\infty} \mathrm{E}e^{it\cdot X} = 0$ for all k-dimensional absolutely continuous random variables X. Can the absolute-continuity condition be removed altogether? (HINT: Consider first a nice X.)

7.16. Suppose X and Y are two independent random variables; X is absolutely continuous with density function f, and the distribution of Y is μ. Prove that $X+Y$ is absolutely continuous with density function

$$(f * \mu)(x) := \int f(x-y)\,\mu(dy). \tag{7.111}$$

Prove also that if Y is absolutely continuous with density function g, then the density function of $X+Y$ is $f * g$.

7.17. Prove that the CLT (p. 100) continues to hold when $\sigma = 0$.

7.18. A probability measure μ on $(\mathbf{R}, \mathscr{B}(\mathbf{R}))$ is said to be *infinitely divisible* if for any $n \ge 1$ there exists a probability measure ν such that $\widehat{\mu} = (\widehat{\nu})^n$. Prove that the normal and the Poisson distributions are infinitely divisible. So is the probability density

$$f(x) := \frac{1}{\pi(1+x^2)} \qquad \forall x \in \mathbf{R}. \tag{7.112}$$

This is called the *Cauchy distribution.* (HINT: Problem 7.14.)

7.19. Prove that if $\{X_i\}_{i=1}^\infty$ are i.i.d. uniform-$[0,1]$ random variables, then

$$\frac{4\sum_{i=1}^n iX_i - n^2}{n^{3/2}} \quad \text{converges weakly.} \tag{7.113}$$

Identify the limiting distribution.

7.20 (Extreme Values)**.** If $\{X_i\}_{i=1}^\infty$ are i.i.d. standard normal random variables, then find non-random sequences $a_n, b_n \to \infty$ such that $a_n \max_{1\le i\le n} X_i - b_n$ converges weakly. Identify the limiting distribution. Replace "standard normal" by "mean-λ exponential," where $\lambda > 0$ is a fixed number, and repeat the exercise.

7.21. Let $\{X_i\}_{i=1}^\infty$ denote independent random variables such that

$$X_j = \begin{cases} \pm j & \text{each with probability } (4j^2)^{-1}, \\ \pm 1 & \text{with probability } \frac{1}{2} - (4j^2)^{-1}. \end{cases} \tag{7.114}$$

Prove that

$$\frac{S_n}{\mathrm{SD}(S_n)} \Rightarrow N(0, \sigma^2), \tag{7.115}$$

and compute σ.

7.22 (An abelian CLT)**.** Suppose that $\{X_i\}_{i=1}^\infty$ are i.i.d. with $\mathrm{E}X_1 = 0$ and $\mathrm{E}[X_1^2] = \sigma^2 < \infty$. First establish that $\sum_{i=1}^\infty r^i X_i$ converges almost surely for all $r \in (0,1)$. Then, prove that

$$\sqrt{1-r}\,\sum_{i=0}^\infty r^i X_i \Rightarrow N(0, \gamma^2) \qquad \text{as } n \to \infty, \tag{7.116}$$

and compute γ (Bovier and Picco, 1996).

7.23. State and prove a variant of Theorem 7.18 that does not assume Q to be non-singular.

7.24 (Liapounov Condition)**.** In the notation of Problem 7.38 below assume there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} \mathrm{E}\left[|X_j - \mu_j|^{2+\delta}\right] = 0. \tag{7.117}$$

Prove the theorem of Liapounov (1900, 1922):

$$\frac{S_n - \sum_{j=1}^{n} \mu_j}{s_n} \Rightarrow N(0\,,1). \tag{7.118}$$

Check that the variables of Problem 7.21 do not satisfy (7.129).

7.25. Compute

$$\lim_{n\to\infty} e^{-n}\left(1 + n + \frac{n^2}{2} + \frac{n^3}{3!} + \frac{n^4}{4!} + \cdots + \frac{n^n}{n!}\right). \tag{7.119}$$

7.26 (The Simple Walk)**.** Let $\mathbf{e}_1, \ldots, \mathbf{e}_d$ denote the usual basis vectors of $\mathbf{R}^d$; i.e., $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, etc. Consider i.i.d. random variables $\{X_i\}_{i=1}^{\infty}$ such that

$$\mathrm{P}\{X_1 = \pm\mathbf{e}_j\} = \frac{1}{2d}. \tag{7.120}$$

Then the random process $S_n = X_1 + \cdots + X_n$ with $S_0 = 0$ is the *simple walk* on $\mathbf{Z}^d$. It starts at zero and moves to each of the neighboring sites in $\mathbf{Z}^d$ with equal probability, and the process continues in this way ad infinitum. Find vectors a_n and constants b_n such that $(S_n - a_n)/b_n$ converges weakly to a nontrivial limit distribution. Compute the latter distribution.

7.27 (Problem 7.26, continued)**.** Consider a collection of points $\Pi = \{\pi_i\}_{i=0}^{n}$ in $\mathbf{Z}^d$. We say that Π is a *lattice path of length* n if $\pi_0 = 0$, and for all $i = 2, \ldots, n-1$ the distance between π_i and π_{i+1} is one. Prove that all lattice paths Π of length n are equally likely for the first n steps in a simple walk.

7.28 (Problem 7.27, continued)**.** Let $N_n(d)$ denote the number of length-n lattice-paths $\{\pi_i\}_{i=0}^{n}$ such that $\pi_n = 0$. Then prove that

$$N_n(d) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \left[2\sum_{j=1}^{d} \cos t_j\right]^n dt \tag{7.121}$$

if $n \geq 2$ is even; else $N_n(d) = 0$. Conclude the 1655 *Wallis formula*:

$$\int_{-\pi}^{\pi} (\cos t)^n \, dt = \binom{n}{n/2} \frac{\pi}{2^{n-1}}, \tag{7.122}$$

valid for all even $n \geq 2$. (HINT: Problem 7.10.)

7.29. Suppose $\{X_i\}_{i=1}^{\infty}$ are i.i.d., mean-zero, and in $L^2(\mathrm{P})$. Prove that there exists a positive constant c such that

$$\mathrm{E}\left(\max_{1\leq j\leq n} |S_j|\right) \geq c\, \mathrm{SD}(X_1)\sqrt{n} \qquad \forall n \geq 1. \tag{7.123}$$

Compare to Problem 6.27 on page 87.

7.30. Suppose that $X_n \Rightarrow X$ and $Y_n \Rightarrow Y$ as $n \to \infty$, where X_n, Y_n, X, and Y are real-valued.

(1) Prove that if Y is non-random, then $Y_n \to Y$ in probability. Conclude from this that $(X_n, Y_n) \Rightarrow (X, Y)$.

(2) Prove that if $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ are independent from one another, then (X_n, Y_n) converges weakly to (X, Y).

(3) Find an example where $X_n \Rightarrow X$, $Y_n \Rightarrow Y$, and $(X_n, Y_n) \not\Rightarrow (X, Y)$.

7.31 (Variance-Stabilizing Transformations). Suppose $g : \mathbf{R} \to \mathbf{R}$ has at least three bounded continuous derivatives, and let $X_1, X_2, \ldots$ be i.i.d. and in $L^2(\mathrm{P})$. Prove that

$$\sqrt{n}\,[g(\bar{X}_n) - g(\mu)] \Rightarrow N(0\,,\sigma^2), \tag{7.124}$$

where $\bar{X}_n := n^{-1}\sum_{i=1}^n X_i$, $\mu := \mathrm{E}X_1$, and $\sigma := \mathrm{SD}(X_1)g'(\mu)$. Also prove that

$$\mathrm{E}g(\bar{X}_n) - g(\mu) = \frac{\sigma^2 g''(\mu)}{2n} + o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty. \tag{7.125}$$

7.32 (Microcanonical Distributions). Prove that if $X^{(n)}$ is distributed uniformly on $\mathbf{S}^{n-1}$, then $(X_1^{(n)}, \ldots, X_k^{(n)}) \Rightarrow Z$ for any fixed $k \geq 1$, where $Z = (Z_1, \ldots, Z_k)$ and the Z_i's are i.i.d. standard normals.

7.33. Choose and fix an integer $n \geq 1$ and let $X_1, X_2, \ldots$ be i.i.d. with common distribution given by $\mathrm{P}\{X_1 = k\} = 1/n$ for $k = 1, \ldots, n$. Let T_n denote the smallest integer $l \geq 1$ such that $X_1 + \cdots + X_l > n$, and compute $\lim_{n\to\infty} \mathrm{P}\{T_n = k\}$ for all k.

7.34 (Uniform Integrability). Suppose $X, X_1, X_2, \ldots$ are real-valued random variables such that: (i) $X_n \Rightarrow X$; and (ii) $\sup_n \|X_n\|_p < \infty$ for some $p > 1$. Then prove that $\lim_{n\to\infty} \mathrm{E}X_n = \mathrm{E}X$. (Hint: See Problem 4.28 on page 51.) Use this to prove the following: Fix some $p_0 \in (0\,,1)$, and define $f(t) = |t - p_0|$ $(t \in [0\,,1])$. Then prove that there exists a constant $c > 0$ such that the Bernstein polynomial $\mathcal{B}_n f$ satisfies

$$|(\mathcal{B}_n f)(p_0) - f(p_0)| \geq \frac{c}{\sqrt{n}} \qquad \forall n \geq 1. \tag{7.126}$$

Thus, (6.50) on page 78 is sharp (Kac, 1937).

7.35 (Hard). Define the *Fourier* map $\mathcal{F}f = \widehat{f}$ for $f \in L^1(\mathbf{R}^k)$. Prove that

$$\|f\|_{L^2(\mathbf{R}^k)} = \frac{1}{(2\pi)^{k/2}}\|\mathcal{F}f\|_{L^2(\mathbf{R}^k)} \qquad \forall f \in L^1(\mathbf{R}^k) \cap L^2(\mathbf{R}^k). \tag{7.127}$$

This is sometimes known as the *Plancherel theorem*. Use it to extend $\mathcal{F}$ to a homeomorphism from $L^2(\mathbf{R}^k)$ onto itself. Conclude from this that if μ is a finite measure on $\mathscr{B}(\mathbf{R}^k)$ such that $\int_{\mathbf{R}^k} |\widehat{\mu}(t)|^2\,dt < \infty$, then μ is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^k$. Warning: The formula $(\mathcal{F}f)(t) = \int_{\mathbf{R}^k} f(x)e^{it\cdot x}\,dx$ is valid only when $f \in L^1(\mathbf{R}^k)$.

7.36 (An Uncertainty Principle; Hard). Prove that if $f : \mathbf{R} \to \mathbf{R}$ is a probability density function that is zero outside $[-\pi\,,\pi]$, then there exists $t \notin [-1/2\,,1/2]$ such that $\hat{f}(t) \neq 0$ (Donoho and Stark, 1989). (Hint: View f as a function on $[-\pi\,,\pi]$, and develop it as a Fourier series. Then study the Fourier coefficients.)

7.37 (Hard). Choose and fix $\lambda_1, \ldots, \lambda_m > 0$ and $a_1, \ldots, a_m \in \mathbf{R}$. Then prove that if $m < \infty$, then f_m defines the characteristic function of a probability measure, where

$$f_m(t) := \exp\left(-\sum_{j=1}^m \lambda_j\,(1 - \cos(a_j t))\right) \qquad \forall t \in \mathbf{R},\ 1 \leq m \leq \infty. \tag{7.128}$$

Prove that f_∞ is a characteristic function provided that $\sum_j (a_j^2 \wedge |a_j|)\lambda_j < \infty$. (Hint: Consult Example 7.14 on page 97.)

7.38 (Lindeberg CLT; Hard). Let $\{X_i\}_{i=1}^\infty$ be independent $L^2(\mathrm{P})$-random variables in $\mathbf{R}$, and for all n define $s_n^2 = \sum_{j=1}^n \mathrm{Var}X_j$ and $\mu_n = \mathrm{E}X_n$. In addition, suppose that $s_n \to \infty$, and

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{j=1}^n \mathrm{E}\left[(X_j - \mu_j)^2;\ |X_j - \mu_j| > \epsilon s_n\right] = 0 \qquad \forall \epsilon > 0. \tag{7.129}$$

Prove the *Lindeberg CLT* (1922):

$$\frac{S_n - \sum_{j=1}^n \mu_j}{s_n} \Rightarrow N(0\,,1). \tag{7.130}$$

Check that the variables of Problem 7.21 do not satisfy (7.129).

7.39 (Hard)**.** Let (X, Y) be a random vector in $\mathbf{R}^2$ and for all $\theta \in (0\,, 2\pi]$ define

$$X_\theta := \cos(\theta)X + \sin(\theta)Y \quad \text{and} \quad Y_\theta := \sin(\theta)X - \cos(\theta)Y. \tag{7.131}$$

Prove that if X_θ and Y_θ are independent for all $\theta \in (0\,, 2\pi]$, then X and Y are independent normal variables. (Hint: Use Cramér's theorem to reduce the problem to the case that X and Y are symmetric; or you can consult the original paper of Kac (1939).)

7.40 (Skorohod's Theorem; Hard)**.** Weak convergence does not imply a.s. convergence. To wit, $X_n \Rightarrow X$ does not even imply that any of the random variables $\{X_n\}_{n=1}^\infty$ and/or X live on the same probability space. The converse, however, is always true; check that $X_n \Rightarrow X$ whenever $X_n \to X$ almost surely. On the other hand, if you are willing to work on *some* probability space, then weak convergence is equivalent to a.s. convergence as we now work to prove.

(1) If F is a distribution function on $\mathbf{R}$ that has a continuous inverse, and if U is uniformly distributed on $(0\,, 1)$, then find the distribution function of $F^{-1}(U)$.

(2) Suppose $F_n \Rightarrow F$: All are distribution functions; each has a continuous inverse. Then prove that $\lim_{n\to\infty} F_n^{-1}(U) = F^{-1}(U)$ a.s.

(3) Use this to prove that whenever $X_n \Rightarrow X_\infty$, we can find, on a suitable probability space, random variables X_n' and X' such that: (i) For every $1 \le n \le \infty$, X_n' has the same distribution as X_n; and (ii) $\lim_n X_n' = X'$ almost surely Skorohod (1961, 1965).

(Hint: Problem 6.9.)

7.41 (Ville's CLT; Hard)**.** Let Ω denote the collection of all permutations of $1, \ldots, n$, and let P be the probability measure that puts mass $(n!)^{-1}$ on each of the $n!$ elements of Ω. For each $\omega \in \Omega$ define $X_1(\omega) = 0$, and for all $k = 2, \ldots, n$ let $X_k(\omega)$ denote the number of *inversions* of k in the permutation ω; i.e., the number of times $1, \ldots, k-1$ precede k in the permutation ω. [For instance, suppose $n = 4$. If $\omega = \{3, 1, 4, 2\}$, then $X_2(\omega) = 1$, $X_3(\omega) = 0$, and $X_4(\omega) = 2$.]

Prove that $\{X_i\}_{i=1}^n$ are independent. Compute their distribution, and prove that the total number of inversions $S_n := \sum_{i=1}^n X_i$ in a *random permutation* satisfies

$$\frac{S_n - (n^2/4)}{n^{3/2}} \Rightarrow N(0\,, 1/36). \tag{7.132}$$

(Hint: Problem 7.38.)

7.42 (A Poincaré Inequality; Hard)**.** Suppose X and Y are independent standard normal random variables.

(1) Prove that for all twice continuously differentiable functions $f, g : \mathbf{R} \to \mathbf{R}$ that have bounded derivatives,

$$\mathrm{Cov}(f(X)\,, g(X)) = \int_0^1 \mathrm{E}\left[f'(X) g'\left(sX + \sqrt{1-s^2}\, Y \right) \right] ds.$$

(Hint: Check it first for $f(x) := \exp(itx)$ and $g(x) := \exp(i\tau x)$.)

(2) Conclude the "Poincaré inequality" of Nash (1958):

$$\mathrm{Var} f(X) \le \|f'(X)\|_2^2.$$

7.43 (Problem 7.18, continued; Harder)**.** Prove that the uniform distribution on $(0\,, 1)$ is not infinitely divisible. (Hint: $\widehat{\mu} \neq (\widehat{\nu})^3$. Simpler derivations exist, but depend on more advanced Fourier-analytic methods.)

7.44 (Harder)**.** Suppose $\{X_i\}_{i=1}^n$ are i.i.d. mean-zero variance-σ^2 random variables such that $\mathrm{E}\{|X_1|^{2+\rho}\} < \infty$ for some $\rho \in (0\,, 1)$. Then prove that there exists a constant A, independent of n, such that

$$|\mathrm{E} g(S_n/\sqrt{n}) - \mathrm{E} g(N(0\,, \sigma^2))| \le \frac{A}{n^{\rho/2}}, \tag{7.133}$$

provided that g has three bounded and continuous derivatives.

Notes

(1) The term "central limit theorem" seems to be due to Pólya (1920). Our treatment covers only the beginning of a rich and well-developed theory (Lévy, 1937; Feller, 1966; Gnedenko and Kolmogorov, 1968).

(2) The present form of the CLT is due to Lindeberg (1922). See also Problem 7.38 on page 115. Zabell (1995) discusses the independent discovery of the Lindeberg CLT (1922) by the nineteen-year-old Turing (1934). See also Note (8) below.

(3) Fejér's Theorem (p. 98) appeared in 1900. Tandori (1983) discusses the fascinating history of the problem, as well as the life of Fejér.

(4) Equation (7.40) is sometimes referred to as the *Parseval identity*, named after M.-A. Parseval des Chénes for his 1801 discovery of a discrete version of (7.40) in the context of Fourier series.

(5) For an amusing consequence of Problem 7.4 plug in $x = \pi/2$ and solve to obtain the 1593 *Viéte formula* for computing π:

$$\pi = 2\left[\frac{\sqrt{2}}{2}\,\frac{\sqrt{2+\sqrt{2}}}{2}\,\frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2}\,\frac{\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}}{2}\cdots\right]^{-1}.$$

(6) Lévy (1925, p. 195) has found the following stronger version of the convergence theorem: "*If $L(t) = \lim_n \widehat{\mu_n}(t)$ exists and is continuous in a neighborhood of $t = 0$, then there exists a probability measure μ such that $L = \widehat{\mu}$ and $\mu_n \Rightarrow \mu$.*" Lévy's argument was simplified by Glivenko (1936).

(7) The term "projective CLT" is non-standard. Kac (1956, p. 182, fn. 7) states that this result "is due to Maxwell but is often ascribed to Borel." See also Kac (1939, p. 728), as well as Problem 7.39 above. The mentioned attribution of Kac seems to agree with that of Borel (1925, p. 92). For a historical survey see the final section of Diaconis and Freedman (1987), as well as Stroock and Zeitouni (1991, Introduction).

(8) The term "Liapounov replacement method" is non-standard. Many authors ascribe this method incorrectly to Lindeberg (1922). Lindeberg used the replacement method in order to deduce the modern-day statement of the CLT.

Trotter (1959) devised a fixed-point proof of the Lindeberg CLT. His proof can be viewed as a translation—into the langauge of analysis—of the replacement method of Liapounov. In this regard see also Hamedani and Walter (1984).

(9) Cramér's theorem (p. 107) is intimately connected to general central limit theory (Gnedenko and Kolmogorov, 1968; Lévy, 1937). The original proof of Cramér's theorem uses hard analytic-function theory. The ascription in Lemma 7.32 comes from Veech (1967, Lemma 7.1, p. 183).

(10) Problem 7.5 goes at least as far back as 1941; see the collected works of Bernšteĭn (1964, pp. 314–315).

(11) Problem 7.41 is borrowed from Ville (1943).

(12) Problem 7.42 is due to Nash (1958), and plays a key role in his estimate for the solution to the Dirichlet problem. The elegant method outlined here is due to Houdré, Pérez-Abreu, and Surgailis (1998).