Chapter 1.
Sums of Independent Random Variables

In one way or another, most probabilistic analysis entails the study of large families of random variables. The key to such analysis is an understanding of the relations among the family members; and of all the possible ways in which members of a family can be related, by far the simplest is when the relationship does not exist at all! For this reason, we will begin by looking at families of *independent* random variables.

§1.1 Independence

In this section we will introduce Kolmogorov's way of describing independence and prove a few of its consequences.

§1.1.1. Independent Sigma Algebras. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a **probability space** (i.e., Ω is a nonempty set, $\mathcal{F}$ is a σ-algebra over Ω, and $\mathbb{P}$ is a measure on the measurable space $(\Omega, \mathcal{F})$ having total mass 1); and, for each i from the (nonempty) index set $\mathcal{I}$, let $\mathcal{F}_i$ be a sub σ-algebra of $\mathcal{F}$. We say that the σ-algebras $\mathcal{F}_i$, $i \in \mathcal{I}$, are **mutually $\mathbb{P}$-independent** or, less precisely, **$\mathbb{P}$-independent**, if, for every finite subset $\{i_1, \dots, i_n\}$ of distinct elements of $\mathcal{I}$ and every choice of $A_{i_m} \in \mathcal{F}_{i_m}$, $1 \le m \le n$,

$$\mathbb{P}\big(A_{i_1} \cap \cdots \cap A_{i_n}\big) = \mathbb{P}\big(A_{i_1}\big) \cdots \mathbb{P}\big(A_{i_n}\big). \tag{1.1.1}$$

In particular, if $\{A_i : i \in \mathcal{I}\}$ is a family of sets from $\mathcal{F}$, we say that A_i, $i \in \mathcal{I}$, are **$\mathbb{P}$-independent** if the associated σ-algebras $\mathcal{F}_i = \{\emptyset, A_i, A_i\complement, \Omega\}$, $i \in \mathcal{I}$, are. To gain an appreciation for the intuition on which this definition is based, it is important to notice that independence of the pair A_1 and A_2 in the present sense is equivalent to

$$\mathbb{P}\big(A_1 \cap A_2\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big),$$

the classical definition which one encounters in elementary treatments. Thus, the notion of independence just introduced is no more than a simple generalization of the classical notion of *independent pairs of sets* encountered in non-measure theoretic presentations; and therefore, the intuition which underlies the elementary notion applies equally well to the definition given here. (See Exercise 1.1.8 below for more information about the connection between the present definition and the classical one.)

As will become increasing evident as we proceed, infinite families of independent objects possess surprising and beautiful properties. In particular, mutually independent σ-algebras tend to *fill up space* in a sense which is made precise by the following beautiful thought experiment designed by A.N. Kolmogorov. Let $\mathcal{I}$ be any index set, take $\mathcal{F}_\emptyset = \{\emptyset, \Omega\}$, and for each nonempty subset $\Lambda \subseteq \mathfrak{I}$, let

$$\mathcal{F}_\Lambda = \bigvee_{i \in \Lambda} \mathcal{F}_i$$

be the σ-algebra generated by $\bigcup_{i\in\Lambda} \mathcal{F}_i$ (i.e., the smallest σ-algebra containing all of the $\mathcal{F}_i$'s). Next, define the **tail σ-algebra** $\mathcal{T}$ to be the intersection over all finite $\Lambda \subseteq \mathcal{I}$ of the σ-algebras $\mathcal{F}_{\Lambda\complement}$. When $\mathcal{I}$ itself is finite, $\mathcal{T} = \{\emptyset, \Omega\}$ and is therefore $\mathbb{P}$-**trivial** in the sense that $\mathbb{P}(A) \in \{0,1\}$ for every $A \in \mathcal{T}$. The interesting remark made by Kolmogorov is that even when $\mathcal{I}$ is infinite, $\mathcal{T}$ is $\mathbb{P}$-trivial whenever the original $\mathcal{F}_i$'s are $\mathbb{P}$-independent. To see this, first note that, by assumption, $\mathcal{F}_{F_1}$ is $\mathbb{P}$-independent of $\mathcal{F}_{F_2}$ whenever F_1 and F_2 are finite, disjoint subsets of $\mathcal{I}$. Since for any (finite or not) $\Lambda \subseteq \mathcal{I}$, $\mathcal{F}_\Lambda$ is generated by the algebra

$$\bigcup \bigl\{\mathcal{F}_F : F \text{ is a finite subset of } \Lambda\bigr\},$$

it follows (cf. Exercise 1.1.8) first that $\mathcal{F}_\Lambda$ is $\mathbb{P}$-independent of $\mathcal{F}_{\Lambda\complement}$ for every $\Lambda \subseteq \mathcal{I}$ and then that $\mathcal{T}$ is $\mathbb{P}$-independent of $\mathcal{F}_\mathcal{I}$. But $\mathcal{T} \subseteq \mathcal{F}_\mathcal{I}$, which means that $\mathcal{T}$ is *independent of itself*; that is, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all A, $B \in \mathcal{T}$. Hence, for every $A \in \mathcal{T}$, $\mathbb{P}(A) = \mathbb{P}(A)^2$, or, equivalently, $\mathbb{P}(A) \in \{0,1\}$; and so we have now proved the following famous result.

Theorem 1.1.2 (**Kolmogorov's 0–1 Law**). *Let $\{\mathcal{F}_i : i \in \mathcal{I}\}$ be a family of $\mathbb{P}$-independent sub-σ-algebras of $(\Omega, \mathcal{F}, \mathbb{P})$, and define the tail σ-algebra $\mathcal{T}$ as above. Then, for every $A \in \mathcal{T}$, $\mathbb{P}(A)$ is either 0 or 1.*

To get a feeling for the kind of conclusions which can be drawn from Kolmogorov's 0–1 Law (cf. Exercises 1.1.16 and 1.1.17 below as well), let $\{A_n\}_1^\infty$ be a sequence of subsets of Ω, and recall the notation

$$\begin{aligned}\varlimsup_{n\to\infty} A_n &\equiv \bigcap_{m=1}^{\infty} \bigcup_{n \ge m} A_n \\ &= \bigl\{\omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{Z}^+\bigr\}.\end{aligned}$$

Obviously, $\varlimsup_{n\to\infty} A_n$ is measurable with respect to the tail field determined by the sequence of σ-algebras $\{\emptyset, A_n, A_n\complement, \Omega\}$, $n \in \mathbb{Z}^+$; and therefore, if the A_n's are $\mathbb{P}$-independent elements of $\mathcal{F}$, then

$$\mathbb{P}\left(\varlimsup_{n\to\infty} A_n\right) \in \{0,1\}.$$

In words, this conclusion can be summarized as the statement that: *for any sequence of $\mathbb{P}$-independent events A_n, $n \in \mathbb{Z}^+$, either $\mathbb{P}$-almost every $\omega \in \Omega$ is in infinitely many A_n's or $\mathbb{P}$-almost every $\omega \in \Omega$ is in at most finitely many A_n's.* A more quantitative statement of this same fact is contained in the second part of the following useful result.

LEMMA 1.1.3 (**Borel–Cantelli Lemma**). *Let $\{A_n : n \in \mathbb{Z}^+\} \subseteq \mathcal{F}$ be given. Then*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\left(\overline{\lim_{n\to\infty}}\, A_n\right) = 0. \tag{1.1.4}$$

Conversely, if the A_n's are $\mathbb{P}$-independent sets, then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \implies \mathbb{P}\left(\overline{\lim_{n\to\infty}}\, A_n\right) = 1. \tag{1.1.5}$$

*(See part (**iii**) of Exercise (5.2.34?) and Lemma for variations on this theme.)*

PROOF: The first assertion is an easy application of countable additivity. Namely, by countable additivity,

$$\mathbb{P}\left(\overline{\lim_{n\to\infty}}\, A_n\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} A_n\right) \leq \lim_{m\to\infty} \sum_{n\geq m} \mathbb{P}(A_n) = 0$$

if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.

To prove (1.1.5), note that, by countable additivity, $\mathbb{P}\left(\overline{\lim}_{n\to\infty} A_n\right) = 1$ if and only if

$$\lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n\geq m} A_n\complement\right) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n\geq m} A_n\complement\right) = \mathbb{P}\left(\left(\overline{\lim_{n\to\infty}}\, A_n\right)\complement\right) = 0.$$

But, again by countable additivity and independence, for given $m \geq 1$ we have that:

$$\mathbb{P}\left(\bigcap_{n=m}^{\infty} A_n\complement\right) = \lim_{N\to\infty} \prod_{n=m}^{N} \bigl(1 - \mathbb{P}(A_n)\bigr) \leq \lim_{N\to\infty} \exp\left[-\sum_{n=m}^{N} \mathbb{P}\bigl(A_n\bigr)\right] = 0$$

if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. (In the preceding, we have used the trivial inequality $1 - t \leq e^{-t}$, $t \in [0, \infty)$.) $\square$

Another, and perhaps more dramatic, statement of the conclusion drawn in the second part of the preceding is the following. Let $\mathbf{N}(\omega) \in \mathbb{Z}^+ \cup \{\infty\}$ be the number of $n \in \mathbb{Z}^+$ such that $\omega \in A_n$. If the A_n's are independent, then Tonelli's Theorem implies that (1.1.5) is equivalent to*

$$\mathbb{P}(\mathbf{N} < \infty) > 0 \implies \mathbb{E}^{\mathbb{P}}[\mathbf{N}] < \infty.$$

§1.1.2. Independent Functions. Having described what it means for the σ-algebras to be $\mathbb{P}$-independent, we can now transfer the notion to random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Namely, for each $i \in \mathcal{I}$, let X_i be a **random variable** (i.e., a measurable function on $(\Omega, \mathcal{F})$) with values in the measurable space $(E_i, \mathcal{B}_i)$. We will say that the random variables X_i, $i \in \mathcal{I}$, are (**mutually**) **$\mathbb{P}$-independent** if the σ-algebras

$$\sigma(X_i) = X_i^{-1}\big(\mathcal{B}_i\big) \equiv \big\{X_i^{-1}(B_i) : B_i \in \mathcal{B}_i\big\}, \ i \in \mathcal{I},$$

are $\mathbb{P}$-independent. Using

$$B(E;\mathbb{R}) = B\big((E,\mathcal{B});\mathbb{R}\big)$$

to denote the space of bounded measurable $\mathbb{R}$-valued functions on the measurable space $(E, \mathcal{B})$, notice that $\mathbb{P}$-independence of $\{X_i : i \in \mathcal{I}\}$ is equivalent to the statement that

$$\mathbb{E}^{\mathbb{P}}\big[f_{i_1} \circ X_{i_1} \cdots f_{i_n} \circ X_{i_n}\big] = \mathbb{E}^{\mathbb{P}}\big[f_{i_1} \circ X_{i_1}\big] \cdots \mathbb{E}^{\mathbb{P}}\big[f_{i_n} \circ X_{i_n}\big]$$

for all finite subsets $\{i_1, \dots, i_n\}$ of distinct elements of $\mathcal{I}$ and all choices of $f_{i_1} \in B\big(E_{i_1};\mathbb{R}\big), \dots,$ and $f_{i_n} \in B\big(E_{i_n};\mathbb{R}\big)$. Finally, if we use $\mathbf{1}_A$ given by

$$\mathbf{1}_A(\omega) \equiv \begin{cases} 1 & \text{if} \quad \omega \in A \\ 0 & \text{if} \quad \omega \notin A \end{cases}$$

to denote the **indicator function** of the set $A \subseteq \Omega$, notice that the family of sets $\{A_i : i \in \mathcal{I}\} \subseteq \mathcal{F}$ is $\mathbb{P}$-independent if and only if the random variables $\mathbf{1}_{A_i}$, $i \in \mathcal{I}$, are $\mathbb{P}$-independent.

Thus far we have discussed only the abstract notion of independence and have yet to show that the concept is not vacuous. In the modern literature, the standard way to construct lots of independent quantities is to take products of probability spaces. Namely, if $\big(E_i, \mathcal{B}_i, \mu_i\big)$ is a probability space for each $i \in \mathcal{I}$, one sets $\Omega = \prod_{i\in\mathcal{I}} E_i$, defines $\pi_i : \Omega \longrightarrow E_i$ to be the natural projection map for each $i \in \mathcal{I}$, takes $\mathcal{F}_i = \pi_i^{-1}(\mathcal{B}_i)$, $i \in \mathcal{I}$, and $\mathcal{F} = \bigvee_{i\in\mathcal{I}} \mathcal{F}_i$, and shows that there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ with the properties that

$$\mathbb{P}\big(\pi_i^{-1}\Gamma_i\big) = \mu_i\big(\Gamma_i\big) \quad \text{for all} \quad i \in \mathcal{I} \text{ and } \Gamma_i \in \mathcal{B}_i$$

* Throughout this book, we use $\mathbb{E}^{\mathbb{P}}[X,\, A]$ to denote the expected value under $\mathbb{P}$ of X over the set A. That is, $\mathbb{E}^{\mathbb{P}}[X,\, A] = \int_A X\, d\mathbb{P}$. Finally, when $A = \Omega$ we will write $\mathbb{E}^{\mathbb{P}}[X]$.

and the σ-algebras $\mathcal{F}_i$, $i \in \mathcal{I}$, are $\mathbb{P}$-independent. Although this procedure is extremely powerful, it is rather mechanical. For this reason, we have chosen to defer the details of the product construction to Exercise 1.1.12 below and to, instead, spend the rest of this section developing a more hands-on approach to constructing independent sequences of real-valued random variables. Indeed, although the product method is more ubiquitous and has become the construction of choice, the one which we are about to present has the advantage that it shows independent random variables can arise "naturally" and even in a familiar context.

§1.1.3. The Rademacher Functions. Until further notice, we take $(\Omega, \mathcal{F}) = \big([0,1), \mathcal{B}_{[0,1)}\big)$ (when E is a metric space, we use $\mathcal{B}_E$ to denote the Borel field over E) and $\mathbb{P}$ to be the restriction $\lambda_{[0,1)}$ of Lebesgue's measure $\lambda_{\mathbb{R}}$ to $[0,1)$. We next define the **Rademacher functions** R_n, $n \in \mathbb{Z}^+$, on Ω as follows. Define the **integer part** $[t]$ of $t \in \mathbb{R}$ to be the largest integer dominated by t and consider the function $R : \mathbb{R} \longrightarrow \{-1, 1\}$ given by

$$R(t) = \begin{cases} -1 & \text{if} \quad t - [t] \in \big[0, \frac{1}{2}\big) \\ \ \ 1 & \text{if} \quad t - [t] \in \big[\frac{1}{2}, 1\big) \end{cases}.$$

The function R_n is then defined on $[0,1)$ by

$$R_n(\omega) = R\big(2^{n-1}\omega\big), \qquad n \in \mathbb{Z}^+ \text{ and } \omega \in [0,1).$$

We will now show that the Rademacher functions are $\mathbb{P}$-independent. To this end, first note that every real-valued function f on $\{-1,1\}$ is of the form $\alpha + \beta x$, $x \in \{-1,1\}$, for some pair of real numbers α and β. Thus, all that we have to show is that

$$\mathbb{E}^{\mathbb{P}}\big[(\alpha_1 + \beta_1 R_1) \cdots (\alpha_n + \beta_n R_n)\big] = \alpha_1 \cdots \alpha_n$$

for any $n \in \mathbb{Z}^+$ and $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n) \in \mathbb{R}^2$. Since this is obvious when $n = 1$, we will assume that it holds for n and will deduce that it must also hold for $n+1$; and clearly this comes down to checking that

$$\mathbb{E}^{\mathbb{P}}\big[F(R_1, \dots, R_n)\, R_{n+1}\big] = 0$$

for any $F : \{-1,1\}^n \longrightarrow \mathbb{R}$. But $(R_1, \dots, R_n)$ is constant on each interval

$$I_{m,n} \equiv \left[\frac{m}{2^n}, \frac{m+1}{2^n}\right), \quad 0 \le m < 2^n$$

whereas R_{n+1} integrates to 0 on each $I_{m,n}$. Hence, by writing the integral over Ω as the sum of integrals over the $I_{m,n}$'s, we get the desired result.

At this point we have produced a countably infinite sequence of independent **Bernoulli random variables** (i.e., two-valued random variables whose range is usually either $\{-1,1\}$ or $\{0,1\}$) with mean-value 0. In order to get more general random variables, we combine our Bernoulli random variables together in a clever way.

Recall that a random variable U is said to be **uniformly distributed** on the finite interval $[a,b]$ if

$$\mathbb{P}(U \le t) = \frac{t-a}{b-a} \quad \text{for } t \in [a,b].$$

LEMMA 1.1.6. *Let $\{Y_\ell : \ell \in \mathbb{Z}^+\}$ be a sequence of $\mathbb{P}$-independent $\{0,1\}$-valued Bernoulli random variables with mean-value $\frac{1}{2}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and set*

$$U = \sum_{\ell=1}^{\infty} \frac{X_\ell}{2^\ell}.$$

Then U is uniformly distributed on $[0,1]$.

PROOF: Because the assertion only involves properties of distributions, it will be proved in general as soon as we prove it for a particular realization of independent, mean-value $\frac{1}{2}$, $\{0,1\}$-valued Bernoulli random variables. In particular, by the preceding discussion, we need only consider the random variables

$$\epsilon_n(\omega) \equiv \frac{1+R_n(\omega)}{2}, \quad n \in \mathbb{Z}^+ \text{ and } \omega \in [0,1),$$

on $\big([0,1), \mathcal{B}_{[0,1)}, \lambda_{[0,1)}\big)$. But, as is easily checked, for each $\omega \in [0,1]$, $\omega = \sum_{n=1}^{\infty} 2^{-n}\epsilon_n(\omega)$. Hence, the desired conclusion is trivial in this case. $\square$

Now let $(k,\ell) \in \mathbb{Z}^+ \times \mathbb{Z}^+ \longmapsto n(k,\ell) \in \mathbb{Z}^+$ be any one-to-one mapping of $\mathbb{Z}^+ \times \mathbb{Z}^+$ onto $\mathbb{Z}^+$, and set

$$Y_{k,\ell} = \frac{1+R_{n(k,\ell)}}{2}, \qquad (k,\ell) \in \big(\mathbb{Z}^+\big)^2.$$

Clearly, each $Y_{k,\ell}$ is a $\{0,1\}$-valued Bernoulli random variable with mean-value $\frac{1}{2}$, and the family $\big\{Y_{k,\ell} : (k,\ell) \in \big(\mathbb{Z}^+\big)^2\big\}$ is $\mathbb{P}$-independent. Hence, by Lemma 1.1.6, each of the random variables

$$U_k \equiv \sum_{\ell=1}^{\infty} \frac{Y_{k,\ell}}{2^\ell}, \qquad k \in \mathbb{Z}^+,$$

is uniformly distributed on $[0,1)$. In addition, the U_k's are obviously mutually independent. Hence, we have now produced a sequence of mutually independent random variables, each of which is uniformly distributed on $[0, 1)$. To complete our program, we use the time-honored transformation which takes a uniform random variable into an arbitrary one. Namely, given a **distribution function** F on $\mathbb{R}$ (i.e., F is a right-continuous, nondecreasing function which tends to 0 at $-\infty$ and 1 at $+\infty$), define F^{-1} on $[0, 1]$ to be the left-continuous inverse of F. That is,

$$F^{-1}(t) = \inf\{s \in \mathbb{R} : F(s) \geq t\}, \qquad t \in [0, 1].$$

(Throughout, the infimum over the empty set is taken to be $+\infty$.) It is then an easy matter to check that when U is uniformly distributed on $[0, 1)$ the random variable $X = F^{-1} \circ U$ has distribution function F:

$$\mathbb{P}(X \leq t) = F(t), \qquad t \in \mathbb{R}.$$

Hence, after combining this with what we already know, we have now completed the proof of the following theorem.

THEOREM 1.1.7. *Let $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, and $\mathbb{P} = \lambda_{[0,1)}$. Then for any sequence $\{F_k : k \in \mathbb{Z}^+\}$ of distribution functions on $\mathbb{R}$ there exists a sequence $\{X_k : k \in \mathbb{Z}^+\}$ of $\mathbb{P}$-independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with the property that $\mathbb{P}\big(X_k \leq t\big) = F_k(t)$, $t \in \mathbb{R}$, for each $k \in \mathbb{Z}^+$.*

Exercises for § 1.1

EXERCISE 1.1.8. As we pointed out, $\mathbb{P}\big(A_1 \cap A_2\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big)$ if and only if the σ-algebra generated by A_1 is $\mathbb{P}$-independent of the one generated by A_2. Construct an example to show that the analogous statement is false when dealing with three, instead of two, sets. That is, just because $\mathbb{P}\big(A_1 \cap A_2 \cap A_3\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big)$, it is not necessarily true that the three σ-algebras generated by A_1, A_2, and A_3 are $\mathbb{P}$-independent.

Next, for any $A \in \mathcal{F}$, show that $\{B \in \mathcal{F} : \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$ is a σ-algebra. Use this to conclude that if, for each $i \in \mathcal{I}$, $\mathcal{F}_i$ is the smallest σ-algebra $\sigma(\mathcal{C}_i)$ containing $\mathcal{C}_i \subseteq \mathcal{F}$, then $\{\mathcal{F}_i : i \in \mathcal{I}\}$ are mutually independent if (1.1.1) holds for $A_{i_m} \in \mathcal{C}_{i_m}$, $1 \leq m \leq n$.

EXERCISE 1.1.9. In this exercise we point out two elementary, but important, properties of independent random variables. Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space.

(**i**) Let X_1 and X_2 be a pair of $\mathbb{P}$-independent random variables with values in the measurable spaces $(E_1, \mathcal{B}_1)$ and $(E_2, \mathcal{B}_2)$, respectively. Given a $\mathcal{B}_1 \times \mathcal{B}_2$-measurable function $F : E_1 \times E_2 \longrightarrow \mathbb{R}$ which is either nonnegative or bounded, use Tonelli's or Fubini's Theorem to show that

$$x_2 \in E_2 \longmapsto f(x_2) \equiv \mathbb{E}^{\mathbb{P}}\Big[F\big(X_1, x_2\big)\Big] \in \mathbb{R}$$

is $\mathcal{B}_2$-measurable and that

$$\mathbb{E}^{\mathbb{P}}\Big[F\big(X_1, X_2\big)\Big] = \mathbb{E}^{\mathbb{P}}\Big[f\big(X_2\big)\Big].$$

(**ii**) Suppose that $X_1, \dots, X_n$ are $\mathbb{P}$-independent, real-valued random variables. If each of the X_m's is P-integrable, show that $X_1 \cdots X_n$ is also P-integrable and that

$$\mathbb{E}^{\mathbb{P}}\big[X_1 \cdots X_n\big] = \mathbb{E}^{\mathbb{P}}\big[X_1\big] \cdots \mathbb{E}^{\mathbb{P}}\big[X_n\big].$$

EXERCISE 1.1.10. Given a nonempty set Ω, recall* that a collection $\mathcal{C}$ of subsets of Ω is called a **$\boldsymbol{\pi}$-system** if $\mathcal{C}$ is closed under finite intersections. At the same time, recall that a collection $\mathcal{L}$ is called a **$\boldsymbol{\lambda}$-system** if $\Omega \in \mathcal{L}$, $A \cup B \in \mathcal{L}$ whenever A and B are disjoint members of $\mathcal{L}$, $B \setminus A \in \mathcal{L}$ whenever A and B are members of $\mathcal{L}$ with $A \subseteq B$, and $\bigcup_1^\infty A_n \in \mathcal{L}$ whenever $\{A_n\}_1^\infty$ is a nondecreasing sequence of members of $\mathcal{L}$. Finally, recall (cf. Lemma 3.1.3 in *ibid.*) that if $\mathcal{C}$ is a π-system, then the σ-algebra $\sigma(\mathcal{C})$ is the smallest $\mathcal{L}$-system $\mathcal{L} \supseteq \mathcal{C}$.

Show that if $\mathcal{C}$ is a π-system and $\mathcal{F} = \sigma(\mathcal{C})$, then two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equal on $\mathcal{F}$ if they are equal on $\mathcal{C}$.

EXERCISE 1.1.11. In this exercise we discuss two criteria for determining when random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are independent.

(**i**) Let $X_1, \dots,$ and X_n be bounded, real-valued random variables. Using Weierstrass's approximation theorem, show that the X_m's are $\mathbb{P}$-independent if and only if

$$\mathbb{E}^{\mathbb{P}}\big[X_1^{m_1} \cdots X_n^{m_n}\big] = \mathbb{E}^{\mathbb{P}}\big[X_1^{m_1}\big] \cdots \mathbb{E}^{\mathbb{P}}\big[X_n^{m_n}\big]$$

for all $m_1, \dots, m_n \in \mathbb{N}$.

(**ii**) Let $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^m$ and $\mathbf{Y} : \Omega \longrightarrow \mathbb{R}^n$ be random variables. Show that $\mathbf{X}$ and $\mathbf{Y}$ are $\mathbb{P}$-independent if and only if

$$\begin{aligned} &\mathbb{E}^{\mathbb{P}}\Big[\exp\Big[\sqrt{-1}\,\Big(\big(\boldsymbol{\alpha}, \mathbf{X}\big)_{\mathbb{R}^m} + \big(\boldsymbol{\beta}, \mathbf{Y}\big)_{\mathbb{R}^n}\Big)\Big]\Big] \\ &\qquad = \mathbb{E}^{\mathbb{P}}\Big[\exp\Big[\sqrt{-1}\,\big(\boldsymbol{\alpha}, \mathbf{X}\big)_{\mathbb{R}^m}\Big]\Big]\,\mathbb{E}^{\mathbb{P}}\Big[\exp\Big[\sqrt{-1}\,\big(\boldsymbol{\beta}, \mathbf{Y}\big)_{\mathbb{R}^n}\Big]\Big] \end{aligned}$$

for all $\boldsymbol{\alpha} \in \mathbb{R}^m$ and $\boldsymbol{\beta} \in \mathbb{R}^n$.

Hint: The *only if* assertion is obvious. To prove the *if* assertion, first check that $\mathbf{X}$ and $\mathbf{Y}$ are independent if

$$\mathbb{E}^{\mathbb{P}}\big[f(\mathbf{X})\,g(\mathbf{Y})\big] = \mathbb{E}^{\mathbb{P}}\big[f(\mathbf{X})\big]\,\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{Y})\big]$$

* See, for example, §3.1 in the author's *A Concise Introduction to the Theory of Integration*, Third Edition publ. by Birkhäuser (1998).

for all $f \in C_c^\infty\big(\mathbb{R}^m;\mathbb{C}\big)$ and $g \in C_c^\infty\big(\mathbb{R}^n;\mathbb{C}\big)$. Second, given such f and g, apply elementary Fourier analysis to write

$$f(\mathbf{x}) = \int_{\mathbb{R}^m} e^{\sqrt{-1}\,(\boldsymbol{\alpha},\mathbf{x})_{\mathbb{R}^m}}\,\varphi(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \quad \text{and} \quad g(\mathbf{y}) = \int_{\mathbb{R}^n} e^{\sqrt{-1}\,(\boldsymbol{\beta},\mathbf{y})_{\mathbb{R}^n}}\,\psi(\boldsymbol{\beta})\,d\boldsymbol{\beta},$$

where φ and ψ are smooth functions with **rapidly decreasing** (i.e., tending to 0 as $|\mathbf{x}| \to \infty$ faster than any power of $(1+|\mathbf{x}|)^{-1}$) derivatives of all orders. Finally, apply Fubini's Theorem.

EXERCISE 1.1.12. Given a pair of measurable spaces $(E_1,\mathcal{B}_1)$ and $(E_2,\mathcal{B}_2)$, recall that their product is the measurable space $\big(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\big)$, where $\mathcal{B}_1 \times \mathcal{B}_2$ is the σ-algebra over the Cartesian product space $E_1 \times E_2$ generated by the sets $\Gamma_1 \times \Gamma_2$, $\Gamma_i \in \mathcal{B}_i$. Further, recall that, for any probability measures μ_i on $(E_i,\mathcal{B}_i)$, there is a unique probability measure $\mu_1 \times \mu_2$ on $\big(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\big)$ such that

$$(\mu_1 \times \mu_2)\big(\Gamma_1 \times \Gamma_2\big) = \mu_1(\Gamma_1)\mu_2(\Gamma_2) \quad \text{for } \Gamma_i \in \mathcal{B}_i.$$

More generally, for any $n \geq 2$ and measurable spaces $\big\{(E_i,\mathcal{B}_i)\big\}_1^n$, one takes $\prod_1^n \mathcal{B}_i$ to be the σ-algebra over $\prod_1^n E_i$ generated by the sets $\prod_1^n \Gamma_i$, $\Gamma_i \in \mathcal{B}_i$. In particular, since $\prod_1^{n+1} E_i$ and $\prod_1^{n+1} \mathcal{B}_i$ can be identified with $(\prod_1^n E_i)\times E_{n+1}$ and $(\prod_1^n \mathcal{B}_i)\times\mathcal{B}_{n+1}$, respectively, one can use induction to show that, for every choice of probability measures μ_i on $(E_i,\mathcal{B}_i)$, there is a unique probability measure $\prod_1^n \mu_i$ on $(\prod_1^n E_i, \prod_1^n \mathcal{B}_i)$ such that

$$\left(\prod_1^n \mu_i\right)\left(\prod_1^n \Gamma_i\right) = \prod_1^n \mu_i(\Gamma_i), \quad \Gamma_i \in \mathcal{B}_i.$$

The purpose of this exercise is to generalize the preceding construction to infinite collections. Thus, let $\mathfrak{I}$ be an infinite index set, and, for each $i \in \mathfrak{I}$, let $(E_i,\mathcal{B}_i)$ be a measurable space. Given $\emptyset \neq \Lambda \subseteq \mathfrak{I}$, we will use $\mathbf{E}_\Lambda$ to denote the Cartesian product space $\prod_{i\in\Lambda} E_i$ and π_Λ to denote the natural projection map taking $\mathbf{E}_\mathfrak{I}$ onto $\mathbf{E}_\Lambda$. Further, we use $\mathcal{B}_\mathfrak{I} = \prod_{i\in\mathfrak{I}} \mathcal{B}_i$ to stand for the σ-algebra over $\mathbf{E}_\mathfrak{I}$ generated by the collection $\mathcal{C}$ of subsets

$$\pi_F^{-1}\left(\prod_{i\in F} \Gamma_i\right), \quad \Gamma_i \in \mathcal{B}_i,$$

as F varies over nonempty, finite subsets of $\mathfrak{I}$ (abbreviated by: $\emptyset \neq F \subset\subset \mathfrak{I}$). In the following steps, we will outline a proof that, for every choice of probability measures μ_i on $(E_i,\mathcal{B}_i)$, there is a unique probability measure $\prod_{i\in\mathfrak{I}} \mu_i$ on $\big(\mathbf{E}_\mathfrak{I}, \mathcal{B}_\mathfrak{I}\big)$ with the property that

$$\left(\prod_{i\in\mathfrak{I}} \mu_i\right)\left(\pi_F^{-1}\left(\prod_{i\in F}\Gamma_i\right)\right) = \prod_{i\in F} \mu_i\big(\Gamma_i\big), \quad \Gamma_i \in \mathcal{B}_i, \tag{1.1.13}$$

for every $\emptyset \neq F \subset\subset \mathfrak{I}$. Not surprisingly, the probability space

$$\left(\prod_{i\in\mathfrak{I}} E_i, \prod_{i\in\mathfrak{I}} \mathcal{B}_i, \prod_{i\in\mathfrak{I}} \mu_i\right)$$

is called the **product** over $\mathfrak{I}$ of the spaces $(E_i, \mathcal{B}_i, \mu_i)$; and when all the factors are the same space $(E, \mathcal{B}, \mu)$, it is customary to denote it by $(E^{\mathfrak{I}}, \mathcal{B}^{\mathfrak{I}}, \mu^{\mathfrak{I}})$, and if, in addition, $\mathfrak{I} = \{1, \dots, N\}$, one uses $(E^N, \mathcal{B}^N, \mu^N)$.

(**i**) After noting that two probability measures which agree on a π-system agree on the σ-algebra generated by that π-system, show that there is at most one probability measure on $(\mathbf{E}_{\mathfrak{I}}, \mathcal{B}_{\mathfrak{I}})$ which satisfies the condition in (1.1.13). Hence, the problem is purely one of existence.

(**ii**) Let $\mathcal{A}$ be the *algebra* over $\mathbf{E}_{\mathfrak{I}}$ generated by $\mathcal{C}$, and show that there is a *finitely* additive $\mu : \mathcal{A} \longrightarrow [0,1]$ with the property that

$$\mu\Big(\pi_F^{-1}(\Gamma_F)\Big) = \left(\prod_{i\in F} \mu_i\right)(\Gamma_F), \quad \Gamma_F \in \mathcal{B}_F,$$

for all $\emptyset \neq F \subset\subset \mathfrak{I}$. Hence, all that we have to do is check that μ admits a σ-additive extension to $\mathcal{B}_{\mathfrak{I}}$, and, by Carathéodory's Extension Theorem, this comes down to checking that $\mu(A_n) \searrow 0$ whenever $\{A_n\}_1^\infty \subseteq \mathcal{A}$ and $A_n \searrow \emptyset$. Thus, let $\{A_n\}_1^\infty$ be a nonincreasing sequence from $\mathcal{A}$, and assume that $\mu(A_n) \geq \epsilon$ for some $\epsilon > 0$ and all $n \in \mathbb{Z}^+$. We must show that $\bigcap_1^\infty A_n \neq \emptyset$.

(**iii**) Referring to the last part of (**ii**), show that there is no loss in generality if we assume that $A_n = \pi_{F_n}^{-1}(\Gamma_{F_n})$, where, for each $n \in \mathbb{Z}^+$, $\emptyset \neq F_n \subset\subset \mathfrak{I}$ and $\Gamma_{F_n} \in \mathcal{B}_{F_n}$. In addition, show that we may assume that $F_1 = \{i_1\}$ and that $F_n = F_{n-1} \cup \{i_n\}$, $n \geq 2$, where $\{i_n\}_1^\infty$ is a sequence of distinct elements of $\mathfrak{I}$. Now, make these assumptions and show that it suffices for us to find $a_\ell \in E_{i_\ell}$, $\ell \in \mathbb{Z}^+$, with the property, for each $m \in \mathbb{Z}^+$, $(a_1, \dots, a_m) \in \Gamma_{F_m}$.

(**iv**) Continuing (**iii**), for each m, $n \in \mathbb{Z}^+$, define $g_{m,n} : \mathbf{E}_{F_m} \longrightarrow [0,1]$ so that

$$g_{m,n}(\mathbf{x}_{F_m}) = \mathbf{1}_{\Gamma_{F_n}}(x_{i_1}, \dots, x_{i_n}) \quad \text{if } n \leq m$$

and

$$g_{m,n}(\mathbf{x}_{F_m}) = \int_{\mathbf{E}_{F_n \setminus F_m}} \mathbf{1}_{\Gamma_{F_n}}(\mathbf{x}_{F_m}, \mathbf{y}_{F_n\setminus F_m}) \left(\prod_{\ell=m+1}^{n} \mu_{i_\ell}\right)(d\mathbf{y}_{F_n\setminus F_m})$$

if $n > m$. After noting that, for each m and n, $g_{m,n+1} \leq g_{m,n}$ and

$$g_{m,n}(\mathbf{x}_{F_m}) = \int_{E_{i_{m+1}}} g_{m+1,n}(\mathbf{x}_{F_m}, y_{i_{m+1}})\, \mu_{i_{m+1}}(dy_{i_{m+1}}),$$

set $g_m = \lim_{n\to\infty} g_{m,n}$ and conclude that

$$g_m\bigl(\mathbf{x}_{F_m}\bigr) = \int_{E_{i_{m+1}}} g_{m+1}\bigl(\mathbf{x}_{F_m}, y_{i_{m+1}}\bigr)\,\mu_{i_{m+1}}\bigl(dy_{i_{m+1}}\bigr).$$

In addition, note that

$$\begin{aligned}\int_{E_{i_1}} g_1\bigl(x_{i_1}\bigr)\,\mu_{i_1}\bigl(dx_{i_1}\bigr) &= \lim_{n\to\infty}\int_{E_{i_1}} g_{1,n}\bigl(x_{i_1}\bigr)\,\mu_{i_1}\bigl(dx_{i_1}\bigr)\\ &= \lim_{n\to\infty}\mu(A_n) \ge \epsilon,\end{aligned}$$

and proceed by induction to produce $a_\ell \in E_{i_\ell}$, $\ell \in \mathbb{Z}^+$, so that

$$g_m\bigl((a_1,\dots,a_m)\bigr) \ge \epsilon \quad \text{for all } m \in \mathbb{Z}^+.$$

Finally, check that $\{a_m\}_1^\infty$ is a sequence of the sort for which we were looking at the end of part (**iii**).

Exercise 1.1.14. Recall that if Φ is a measurable map from one measurable space $(E,\mathcal{B})$ into a second one $(E',\mathcal{B}')$, then the **distribution** of Φ under a measure μ on $(E,\mathcal{B})$ is the **pushforward** measure $\Phi_*\mu$ (also denoted by $\mu\circ\Phi^{-1}$) defined on $(E',\mathcal{B}')$ by

$$\Phi_*\mu(\Gamma) = \mu\bigl(\Phi^{-1}(\Gamma)\bigr) \quad \text{for} \quad \Gamma \in \mathcal{B}'.$$

Given a nonempty index set $\mathfrak{I}$ and, for each $i \in \mathfrak{I}$, a measurable space $(E_i,\mathcal{B}_i)$ and an E_i-valued random variable X_i on the probability space $(\Omega,\mathcal{F},\mathbb{P})$, define $\mathbf{X} : \Omega \longrightarrow \prod_{i\in\mathfrak{I}} E_i$ so that $\mathbf{X}(\omega)_i = X_i(\omega)$ for each $i \in \mathfrak{I}$ and $\omega \in \Omega$. Show that $\bigl\{X_i : i \in \mathfrak{I}\bigr\}$ is a family of $\mathbb{P}$-independent random variables if and only if $\mathbf{X}_*P = \prod_{i\in\mathfrak{I}}(X_i)_*P$. In particular, given probability measures μ_i on $(E_i,\mathcal{B}_i)$, set

$$\Omega = \prod_{i\in\mathfrak{I}} E_i, \quad \mathcal{F} = \prod_{i\in\mathfrak{I}} \mathcal{B}_i, \quad P = \prod_{i\in\mathfrak{I}} \mu_i,$$

let $X_i : \Omega \longrightarrow E_i$ be the natural projection map from Ω onto E_i, and show that $\{X_i : i \in \mathfrak{I}\}$ is a family of mutually $\mathbb{P}$-independent random variables such that, for each $i \in \mathfrak{I}$, X_i has distribution μ_i.

Exercise 1.1.15. Although it does not entail infinite product spaces, an interesting example of the way in which the preceding type of construction can be effectively applied is provided by the following elementary version of a *coupling* argument.

(**i**) Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and X and Y a pair of square $\mathbb{P}$-integrable $\mathbb{R}$-valued random variables with the property that

$$\big(X(\omega) - X(\omega')\big)\,\big(Y(\omega) - Y(\omega')\big) \geq 0 \quad \text{for all } (\omega, \omega') \in \Omega^2.$$

Show that

$$\mathbb{E}^{\mathbb{P}}\big[X\,Y\big] \geq \mathbb{E}^{\mathbb{P}}[X]\,\mathbb{E}^{\mathbb{P}}[Y].$$

Hint: Define X_i and Y_i on Ω^2 for $i \in \{1, 2\}$ so that $X_i(\boldsymbol{\omega}) = X(\omega_i)$ and $Y_i(\boldsymbol{\omega}) = Y(\omega_i)$ when $\boldsymbol{\omega} = (\omega_1, \omega_2)$, and integrate the inequality

$$0 \leq \big(X(\omega_1) - X(\omega_2)\big)\,\big(Y(\omega_1) - Y(\omega_2)\big) = \big(X_1(\boldsymbol{\omega}) - X_2(\boldsymbol{\omega})\big)\,\big(Y_1(\boldsymbol{\omega}) - Y_2(\boldsymbol{\omega})\big)$$

with respect to $\mathbb{P}^2$.

(**ii**) Suppose that $n \in \mathbb{Z}^+$ and that f and g are $\mathbb{R}$-valued, Borel measurable functions on $\mathbb{R}^n$ which are nondecreasing with respect to each coordinate (separately). Show that if $\mathbf{X} = \big(X_1, \dots, X_n\big)$ is an $\mathbb{R}^n$-valued random variable on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ whose coordinates are mutually $\mathbb{P}$-independent, then

$$\mathbb{E}^{\mathbb{P}}\big[f(\mathbf{X})\,g(\mathbf{X})\big] \geq \mathbb{E}^{\mathbb{P}}\big[f(\mathbf{X})\big]\,\mathbb{E}^{\mathbb{P}}\big[g(\mathbf{X})\big]$$

so long as $f(\mathbf{X})$ and $g(\mathbf{X})$ are both square $\mathbb{P}$-integrable.

Hint: First check that the case when $n = 1$ reduces to an application of (**i**). Next, describe the general case in terms of a multiple integral, apply Fubini's Theorem, and make repeated use of the case when $n = 1$.

Exercise 1.1.16. A σ-algebra is said to be **countably generated** if it contains a countable collection of sets which generate it. In this exercise, we will show that just because a σ-algebra is itself countably generated does not mean that all its sub-σ-algebras are.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measurable space and $\big\{\mathcal{F}_n : n \in \mathbb{Z}^+\big\}$ be a sequence of $\mathbb{P}$-independent sub-σ-algebras of $\mathcal{F}$. Further, assume that, for each $n \in \mathbb{Z}^+$, there is an $A_n \in \mathcal{F}_n$ which satisfies $\alpha \leq P\big(A_n\big) \leq 1 - \alpha$ for some fixed $\alpha \in (0, \frac{1}{2})$. Show that the tail σ-algebra $\mathcal{T}$ determined by $\big\{\mathcal{F}_n : n \in \mathbb{Z}^+\big\}$ cannot be countably generated.

Hint: First, reduce to the case when each $\mathcal{F}_n$ is generated by the set A_n. After making this reduction, show that C is an **atom** in $\mathcal{T}$ (i.e., $B = C$ whenever $B \in \mathcal{T} \setminus \{\emptyset\}$ is contained in C) only if one can write

$$C = \varliminf_{n\to\infty} C_n \equiv \bigcup_{m=1}^{\infty} \bigcap_{n \geq m} C_n$$

where, for each $n \in \mathbb{Z}^+$, C_n equals either A_n or $A_n\complement$. Conclude that every atom in $\mathcal{T}$ must have $\mathbb{P}$-measure 0. Now suppose that $\mathcal{T}$ were generated by $\{B_\ell : \ell \in \mathbb{N}\}$. By Kolmogorov's 0–1 Law (cf. Theorem 1.1.2), $\mathbb{P}(B_\ell) \in \{0,1\}$ for every $\ell \in \mathbb{N}$. Take

$$\hat{B}_\ell = \begin{cases} B_\ell & \text{if} \quad P(B_\ell) = 1 \\ B_\ell\complement & \text{if} \quad P(B_\ell) = 0 \end{cases} \quad \text{and set} \quad C = \bigcap_{\ell\in\mathbb{N}} \hat{B}_\ell.$$

Note that, on the one hand, $\mathbb{P}(C) = 1$, while, on the other hand, C is an atom in $\mathcal{T}$ and therefore has probability 0.

Exercise 1.1.17. Here is an interesting application of Kolmogorov's 0–1 Law to a property of the real numbers.

(**i**) Referring to the discussion preceding Lemma 1.1.6, define the transformations $T_n : [0,1) \longrightarrow [0,1)$ for $n \in \mathbb{Z}^+$ so that

$$T_n(\omega) = \omega - \frac{1 + R_n(\omega)}{2^{n+1}}, \quad \omega \in [0,1),$$

and notice (cf. the proof of Lemma 1.1.6) that $T_n(\omega)$ simply *flips* the nth coefficient in the binary expansion ω. Next, let $\Gamma \in \mathcal{B}_{[0,1)}$, and show that Γ is measurable with respect of the σ-algebra $\sigma(R_n : n > m)$ generated by $\{R_n : n > m\}$ if and only if $T_n(\Gamma) = \Gamma$ for each $1 \le n \le m$. In particular, conclude that $\lambda_{[0,1)}(\Gamma) \in \{0,1\}$ if $T_n\Gamma = \Gamma$ for every $n \in \mathbb{Z}^+$.

(**ii**) Let $\mathfrak{F}$ denote the set of all finite subsets of $\mathbb{Z}^+$, and for each $F \in \mathfrak{F}$, define $T^F : [0,1) \longrightarrow [0,1)$ so that $T^\emptyset$ is the identity mapping and

$$T^{F\cup\{m\}} = T^F \circ T_m \quad \text{for each } F \in \mathfrak{F} \text{ and } m \in \mathbb{Z}^+ \setminus F.$$

As an application of (**i**), show that for every $\Gamma \in \mathcal{B}_{[0,1)}$ with $\lambda_{[0,1)}(\Gamma) > 0$,

$$\lambda_{[0,1)}\left(\bigcup_{F\in\mathfrak{F}} T^F(\Gamma)\right) = 1.$$

In particular, this means that if Γ has positive measure, then almost every $\omega \in [0,1)$ can be moved to Γ by *flipping* a finite number of the coefficients in the binary expansion of ω.

§1.2 The Weak Law of Large Numbers

Starting with this section, and for the rest of this chapter, we will be studying what happens when one averages $\mathbb{P}$-independent, real-valued random variables. The remarkable fact, which will be confirmed repeatedly, is that the limiting behavior of such averages depends hardly at all on the variables involved. Intuitively, one can explain this phenomenon by pretending that the random variables are building blocks which, in the averaging process, first get homothetically shrunk and then reassembled according to a regular pattern. Hence, by the time that one passes to the limit, the peculiarities of the original blocks get lost.

Throughout our discussion, $(\Omega, \mathcal{F}, \mathbb{P})$ will be a probability space on which we have a sequence $\{X_n\}_1^\infty$ of real-valued random variables. Given $n \in \mathbb{Z}^+$, we will use S_n to denote the partial sum $X_1 + \cdots + X_n$ and $\overline{S}_n$ to denote the average

$$\frac{S_n}{n} = \frac{1}{n} \sum_{\ell=1}^{n} X_\ell .$$

§1.2.1. Orthogonal Random Variables. Our first result is a very general one; in fact, it even applies to random variables which are not necessarily independent and do not necessarily have mean 0.

LEMMA 1.2.1. *Assume that*

$$\mathbb{E}^{\mathbb{P}}\big[X_n^2\big] < \infty \text{ for } n \in \mathbb{Z}^+ \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}\big[X_k\, X_\ell\big] = 0 \text{ if } \; k \neq \ell.$$

Then, for each $\epsilon > 0$,

$$\epsilon^2\, P\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) \leq \mathbb{E}^{\mathbb{P}}\big[\,\overline{S}_n^2\big] = \frac{1}{n^2} \sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\big[X_\ell^2\big] \quad \text{for } n \in \mathbb{Z}^+. \tag{1.2.2}$$

In particular, if

$$M \equiv \sup_{n \in \mathbb{Z}^+} \mathbb{E}^{\mathbb{P}}\big[X_n^2\big] < \infty,$$

then

$$\epsilon^2\, P\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) \leq \mathbb{E}^{\mathbb{P}}\big[\overline{S}_n^{\,2}\big] \leq \frac{M}{n}, \qquad n \in \mathbb{Z}^+ \text{ and } \epsilon > 0; \tag{1.2.3}$$

and so $\overline{S}_n \longrightarrow 0$ in $L^2(P)$ and also in $\mathbb{P}$-probability.

PROOF: To prove the equality in (1.2.2), note that, by by orthognality,

$$\mathbb{E}^{\mathbb{P}}\big[S_n^2\big] = \sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\big[X_\ell^2\big].$$

The rest is just an application of **Chebyshev's inequality**, the estimate which results after integrating the inequality

$$\epsilon^2 \mathbf{1}_{[\epsilon,\infty)}\big(|Y|\big) \leq Y^2 \mathbf{1}_{[\epsilon,\infty)}\big(|Y|\big) \leq Y^2$$

for any random variable Y. $\square$

§1.2.2. Indenpendent Random Variables. Although Lemma 1.2.1 does not use independence, independent random variables provide a ready source of orthogonal functions. Indeed, recall that for any $\mathbb{P}$-integrable random variable X, its **variance** $\operatorname{var}(X)$ satisfies

$$\operatorname{var}(X) \equiv \mathbb{E}^{\mathbb{P}}\Big[\big(X - \mathbb{E}^{\mathbb{P}}[X]\big)^2\Big] = \mathbb{E}^{\mathbb{P}}\big[X^2\big] - \big(\mathbb{E}^{\mathbb{P}}[X]\big)^2 \le \mathbb{E}^{\mathbb{P}}\big[X^2\big].$$

In particular, if the random variables X_n, $n \in \mathbb{Z}^+$, are $\mathbb{P}$-square integral and $\mathbb{P}$-independent, then the random variables

$$\hat{X}_n \equiv X_n - \mathbb{E}^{\mathbb{P}}\big[X_n\big] \qquad n \in \mathbb{Z}^+,$$

are still square $\mathbb{P}$-integrable, have mean-value 0, and therefore satisfy the hypotheses in Lemma 1.2.1. Hence, the following statement is an immediate consequence of that lemma.

Theorem 1.2.4. *Let $\big\{X_n : n \in \mathbb{Z}^+\big\}$ be a sequence of $\mathbb{P}$-independent, square $\mathbb{P}$-integrable random variables with mean-value m and variance dominated by σ^2. Then, for every $n \in \mathbb{Z}^+$ and $\epsilon > 0$:*

$$\epsilon^2\, P\Big(\big|\overline{S}_n - m\big| \ge \epsilon\Big) \le \mathbb{E}^{\mathbb{P}}\Big[\big(\overline{S}_n - m\big)^2\Big] \le \frac{\sigma^2}{n}. \tag{1.2.5}$$

In particular, $\overline{S}_n \longrightarrow m$ in $L^2(P)$ and therefore in $\mathbb{P}$-probability.

As yet we have only made minimal use of independence: all that we have done is subtract off the mean of independent random variables and thereby made them orthogonal. In order to bring the full force of independence into play, one has to exploit the fact that one can compose independent random variables with any (measurable) functions without destroying their independence; in particular, *truncating* independent random variables does not destroy independence. To see how such a property can be brought to bear, we will now consider the problem of extending the last part of Theorem 1.2.4 to X_n's which are less than square $\mathbb{P}$-integrable. In order to understand the statement, recall that a family $\big\{X_i : i \in \mathcal{I}\big\}$ of random variables is said to be **uniformly $\mathbb{P}$-integrable** if

$$\lim_{R \nearrow \infty} \sup_{i \in \mathcal{I}} \mathbb{E}^{\mathbb{P}}\Big[\big|X_i\big|,\ \big|X_i\big| \ge R\Big] = 0. \tag{1.2.6}$$

As the proof of the following theorem illustrates, the importance of this condition is that it allows one to simultaneously approximate the random variables X_i, $i \in \mathcal{I}$, by bounded random variables.

Theorem 1.2.7 (**The Weak Law of Large Numbers**). *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a uniformly $\mathbb{P}$-integrable sequence of $\mathbb{P}$-independent random variables. Then*

$$\frac{1}{n}\sum_1^n \bigl(X_m - \mathbb{E}^{\mathbb{P}}[X_m]\bigr) \longrightarrow 0 \text{ in } L^1(P)$$

and, therefore, also in $\mathbb{P}$-probability. In particular, if $\{X_n : n \in \mathbb{Z}^+\}$ is a sequence of $\mathbb{P}$-independent, $\mathbb{P}$-integrable random variables which are identically distributed, then $\overline{S}_n \longrightarrow \mathbb{E}^{\mathbb{P}}\bigl[X_1\bigr]$ in $L^1(P)$ and $\mathbb{P}$-probability. (Cf. Exercise 1.2.12 below.)

Proof: Without loss in generality, we will assume that $\mathbb{E}^{\mathbb{P}}[X_n] = 0$ for every $n \in \mathbb{Z}^+$.

For each $R \in (0,\infty)$, define $f_R(t) = t\,\mathbf{1}_{[-R,R]}(t)$, $t \in \mathbb{R}$,

$$m_n^{(R)} = \mathbb{E}^{\mathbb{P}}\bigl[f_R \circ X_n\bigr], \quad X_n^{(R)} = f_R \circ X_n - m_n^{(R)}, \quad \text{and} \quad Y_n^{(R)} = X_n - X_n^{(R)},$$

and set

$$\overline{S}_n^{(R)} = \frac{1}{n}\sum_{\ell=1}^n X_\ell^{(R)} \quad \text{and} \quad \overline{T}_n^{(R)} = \frac{1}{n}\sum_{\ell=1}^n Y_\ell^{(R)}.$$

Since $\mathbb{E}[X_n] = 0 \implies m_n^{(R)} = -\mathbb{E}\bigl[X_n,\, |X_n| > R\bigr]$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\Bigl[\bigl|\overline{S}_n\bigr|\Bigr] &\le \mathbb{E}^{\mathbb{P}}\Bigl[\bigl|\overline{S}_n^{(R)}\bigr|\Bigr] + \mathbb{E}^{\mathbb{P}}\Bigl[\bigl|\overline{T}_n^{(R)}\bigr|\Bigr] \\ &\le \mathbb{E}^{\mathbb{P}}\Bigl[\bigl|\overline{S}_n^{(R)}\bigr|^2\Bigr]^{\frac{1}{2}} + 2\max_{1\le\ell\le n}\mathbb{E}^{\mathbb{P}}\Bigl[\bigl|X_\ell\bigr|,\, |X_\ell| \ge R\Bigr] \\ &\le \frac{R}{\sqrt{n}} + 2\max_{\ell\in\mathbb{Z}^+}\mathbb{E}^{\mathbb{P}}\Bigl[\bigl|X_\ell\bigr|,\, |X_\ell| \ge R\Bigr];\end{aligned}$$

and therefore, for each $R > 0$,

$$\varlimsup_{n\to\infty}\mathbb{E}^{\mathbb{P}}\Bigl[\bigl|\overline{S}_n\bigr|\Bigr] \le 2\sup_{\ell\in\mathbb{Z}^+}\mathbb{E}^{\mathbb{P}}\Bigl[\bigl|X_\ell\bigr|,\, |X_\ell| \ge R\Bigr].$$

Hence, because the X_ℓ's are uniformly $\mathbb{P}$-integrable, we get the desired convergence in $L^1(P)$ by letting $R \nearrow \infty$. $\square$

§1.2.3. Approximate Identities. The name of Theorem 1.2.7 comes from a somewhat invidious comparison with the result in Theorem 1.4.9. The reason why the appellation *weak* is not entirely fair is that, although The Weak Law is indeed less *refined* than the result in Theorem 1.4.9, it is every bit as useful as the one in Theorem 1.4.9 and maybe even more important when it comes to applications. What The Weak Law does is provide us with a ubiquitous technique for constructing an **approximate identity** (i.e., a sequence of measures which approximate a point mass) and measuring how fast the approximation is taking place. To illustrate how clever selection of the random variables entering The Weak Law can lead to interesting applications, we will spend the rest of this section discussing S. Bernstein's approach to Weierstrass's approximation theorem.

For a given $p \in [0,1]$, let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of $\mathbb{P}$-independent $\{0,1\}$-valued Bernoulli random variables with mean-value p. Then

$$\mathbb{P}\big(S_n = \ell\big) = \binom{n}{\ell} p^\ell (1-p)^{n-\ell} \quad \text{for} \quad 0 \le \ell \le n.$$

Hence, for any $f \in C\big([0,1];\mathbb{R}\big)$, the nth **Bernstein polynomial**

$$B_n(p;f) \equiv \sum_{\ell=0}^{n} \binom{n}{\ell} f\left(\frac{\ell}{n}\right) p^\ell (1-p)^{n-\ell} \tag{1.2.8}$$

of f at p is equal to

$$\mathbb{E}^{\mathbb{P}}\big[f \circ \overline{S}_n\big].$$

In particular,

$$\begin{aligned}\big|f(p) - B_n(p;f)\big| &= \big|\mathbb{E}^{\mathbb{P}}\big[f(p) - f \circ \overline{S}_n\big]\big| \le \mathbb{E}^{\mathbb{P}}\big[\big|f(p) - f \circ \overline{S}_n\big|\big] \\ &\le 2\|f\|_{\mathrm{u}} P\big(\big|\overline{S}_n - p\big| \ge \epsilon\big) + \rho(\epsilon; f),\end{aligned}$$

where $\|f\|_{\mathrm{u}}$ is the **uniform norm** of f (i.e., the supremum of $|f|$ over the domain of f) and

$$\rho(\epsilon; f) \equiv \sup\big\{|f(t) - f(s)| : 0 \le s < t \le 1 \text{ with } t - s \le \epsilon\big\}$$

is the modulus of continuity of f. Noting that $\mathrm{var}\big(X_n\big) = p(1-p) \le \frac{1}{4}$ and applying (1.2.5), we conclude that, for every $\epsilon > 0$,

$$\big\|f(p) - B_n(p;f)\big\|_{\mathrm{u}} \le \frac{\|f\|_{\mathrm{u}}}{2n\epsilon^2} + \rho(\epsilon; f)$$

In other words, for all $n \in \mathbb{Z}^+$,

$$\big\|f - B_n(\,\cdot\,;f)\big\|_{\mathrm{u}} \le \beta(n;f) \equiv \inf\left\{\frac{\|f\|_{\mathrm{u}}}{2n\epsilon^2} + \rho(\epsilon;f) : \epsilon > 0\right\}. \tag{1.2.9}$$

Obviously, (1.2.9) not only shows that, as $n \to \infty$, $B_n(\,\cdot\,;f) \longrightarrow f$ uniformly on $[0,1]$, but it even provides a rate of convergence in terms of the modulus of continuity of f. Thus, we have done more than simply prove Weierstrass's theorem; we have produced a rather explicit and tractable sequence of approximating polynomials, the sequence $\{B_n(\,\cdot\,;f) : n \in \mathbb{Z}^+\}$. Although this sequence is, by no means, the most efficient one,* as we are about to see, the Bernstein polynomials have a lot to recommend them. In particular, they have the feature that

* See G.G. Lorentz's *Bernstein Polynomials*, Chelsea Publ. Co., New York (1986) for a lot more information.

they provide nonnegative polynomial approximants to nonnegative functions. In fact, the following discussion reveals much deeper nonnegativity preservation properties possessed by the Bernstein approximation scheme.

In order to bring out the virtues of the Bernstein polynomials, it is important to replace (1.2.8) with an expression in which the coefficients of $B_n(\,\cdot\,;f)$ (as polynomials) are clearly displayed. To this end, introduce the **difference operator** Δ_h for $h>0$ given by

$$\big[\Delta_h f\big](t)=\frac{f(t+h)-f(t)}{h}.$$

A straightforward inductive argument (using Pascal's identity for the binomials coefficients) shows that

$$(-h)^m\big[\Delta_h^{(m)}f\big](t)=\sum_{\ell=0}^{m}(-1)^\ell\binom{m}{\ell}f(t+\ell h)\quad\text{for}\quad m\in\mathbb{Z}^+,$$

where $\Delta_h^{(m)}$ denotes the mth iterate of the operator Δ_h. Taking $h=\frac{1}{n}$, we now see that

$$\begin{aligned}B_n(p;f)&=\sum_{\ell=0}^{n}\sum_{k=0}^{n-\ell}\binom{n}{\ell}\binom{n-\ell}{k}(-1)^k f(\ell h)p^{\ell+k}\\&=\sum_{r=0}^{n}p^r\sum_{\ell=0}^{r}\binom{n}{\ell}\binom{n-\ell}{r-\ell}(-1)^{r-\ell}f(\ell h)\\&=\sum_{r=0}^{n}(-p)^r\binom{n}{r}\sum_{\ell=0}^{r}\binom{r}{\ell}(-1)^\ell f(\ell h)\\&=\sum_{r=0}^{n}\binom{n}{r}(ph)^r\big[\Delta_h^{(r)}f\big](0),\end{aligned}$$

where $\big[\Delta_h^0 f\big]\equiv f$. Hence, we have proved that

$$B_n(p;f)=\sum_{\ell=0}^{n}n^{-\ell}\binom{n}{\ell}\left[\Delta_{\frac{1}{n}}^{(\ell)}f\right](0)p^\ell\quad\text{for}\quad p\in[0,1].\tag{1.2.10}$$

The marked resemblance between the expression on the right-hand side of (1.2.10) and a Taylor polynomial is more than coincidental. To demonstrate how one can exploit the relationship between Bernstein and Taylor polynomials, say that a function $\varphi\in C^\infty\big((a,b);\mathbb{R}\big)$ is **absolutely monotone** if its mth derivative $D^m\varphi$ is nonnegative for every $m\in\mathbb{N}$. Also, say that $\varphi\in C^\infty\big([0,1];[0,1]\big)$ is a

probability generating function if there exists a $\{u_n : n \in \mathbb{N}\} \subseteq [0,1]$ such that

$$\sum_{n=0}^{\infty} u_n = 1 \quad \text{and} \quad \varphi(t) = \sum_{n=0}^{\infty} u_n t^n \quad \text{for} \quad t \in [0,1].$$

Obviously, every probability generating function is absolutely monotone on $(0,1)$. The somewhat surprising (remember that most infinitely differentiable functions do not admit power series expansions) fact which we are about to prove is that, apart from a multiplicative constant, the converse is also true. In fact, we do not need to know, *a priori*, that the function is smooth so long as it satisfies a discrete version of absolute monotonicity.

THEOREM 1.2.11. *Let $\varphi \in C\big([0,1];\mathbb{R}\big)$ with $\varphi(1) = 1$ be given. Then the following are equivalent:*
(**i**) *φ is a probability generating function,*
(**ii**) *the restriction of φ to $(0,1)$ is absolutely monotone;*
(**iii**) *$\big[\Delta_{\frac{1}{n}}^{(m)}\varphi\big](0) \geq 0$ for every $n \in \mathbb{N}$ and $0 \leq m \leq n$.*

PROOF: The implication (**i**) $\Longrightarrow$ (**ii**) is trivial. To see that (**ii**) implies (**iii**), first observe that if ψ is absolutely monotone on (a,b) and $h \in (0, b-a)$, then $\big[\Delta_h\psi\big]$ is absolutely monotone on $(a, b-h)$. Indeed, because $\big[D \circ \Delta_h\psi\big] = \big[\Delta_h \circ D\psi\big]$ on $(a, b-h)$, we see that

$$h\big[D^m \circ \Delta_h\psi\big](t) = \int_t^{t+h} D^{m+1}\psi(s)\,ds \geq 0, \quad t \in (a, b-h),$$

for any $m \in \mathbb{N}$. Returning to the function φ, we now know that $\big[\Delta_h^{(m)}\varphi\big]$ is absolutely monotone on $(0, 1-mh)$ for all $m \in \mathbb{N}$ and $h > 0$ with $mh < 1$. In particular,

$$\big[\Delta_h^{(m)}\varphi\big](0) = \lim_{t\searrow 0}\big[\Delta_h^{(m)}\varphi\big](t) \geq 0 \quad \text{if} \quad mh < 1,$$

and so $\big[\Delta_h^{(m)}\varphi\big](0) \geq 0$ when $h = \frac{1}{n}$ and $0 \leq m < n$. Moreover, since

$$\big[\Delta_{\frac{1}{n}}^{(n)}\varphi\big](0) = \lim_{h\nearrow\frac{1}{n}}\big[\Delta_h^{(n)}\varphi\big](0),$$

we also know that $\big[\Delta_h^n\varphi\big](0) \geq 0$ when $h = \frac{1}{n}$, and this completes the proof that (**ii**) implies (**iii**).

Finally, assume that (**iii**) holds and set $\varphi_n = B_n(\,\cdot\,;\varphi)$. Then, by (1.2.10) and the equality $\varphi_n(1) = \varphi(1) = 1$, we see that each φ_n is a probability generating function. Thus, in order to complete the proof that (**iii**) implies (**i**), all that we have to do is check that a uniform limit of probability generating functions is itself a probability generating function. To this end, write

$$\varphi_n(t) = \sum_{\ell=0}^{\infty} u_{n,\ell}t^{\ell}, \quad t \in [0,1] \text{ for each } n \in \mathbb{Z}^+.$$

Because the $u_{n,\ell}$'s are all elements of $[0,1]$, we can use a diagonalization procedure to choose $\{n_k : k \in \mathbb{Z}^+\}$ so that

$$\lim_{k\to\infty} u_{n_k,\ell} = u_\ell \in [0,1]$$

exists for each $\ell \in \mathbb{N}$. But, by Lebesgue's Dominated Convergence Theorem, this means that

$$\varphi(t) = \lim_{k\to\infty} \varphi_{n_k}(t) = \sum_{\ell=0}^{\infty} u_\ell t^\ell \quad \text{for every} \quad t \in [0,1).$$

Finally, by the Monotone Convergence Theorem, the preceding extends immediately to $t = 1$, and so φ is a probability generating function. (Notice that the argument just given does not even use the assumed uniform convergence and shows that the pointwise limit of probability generating functions is again a probability generating function.) □

The preceding is only one of many examples in which The Weak Law leads to useful ways of forming an approximate identity. A second example is given in Exercises 1.2.13 and 1.4.22 below. My treatment these is based on that of Wm. Feller,* who provides several other similar applications of The Weak Law, including the ones in the following exercises.

Exercises for § 1.2

EXERCISE 1.2.12. Although, for historical reasons, The Weak Law is usually thought of as a theorem about convergence in $\mathbb{P}$-probability, the forms in which we have presented it are clearly results about convergence in either $\mathbb{P}$-mean or even square $\mathbb{P}$-mean. Thus, it is interesting to discover that one can replace the uniform integrability assumption made in Theorem 1.2.7 with a *weak uniform integrability* assumption if one is willing to settle for convergence in $\mathbb{P}$-probability. Namely, let $X_1, \dots, X_n, \dots$ be mutually $\mathbb{P}$-independent random variables, assume that

$$F(R) \equiv \sup_{n\in\mathbb{Z}^+} RP\Big(|X_n| \geq R\Big) \longrightarrow 0 \quad \text{as } R \nearrow \infty,$$

and set

$$m_n = \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\Big[X_\ell,\ |X_\ell| \leq n\Big], \quad n \in \mathbb{Z}^+.$$

Show that, for each $\epsilon > 0$,

$$\begin{aligned} P\Big(\big|\overline{S}_n - m_n\big| \geq \epsilon\Big) &\leq \frac{1}{(n\epsilon)^2}\sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\Big[X_\ell^2,\ \big|X_\ell\big| \leq n\Big] + P\Big(\max_{1\leq\ell\leq n} \big|X_\ell\big| > n\Big) \\ &\leq \frac{2}{n\epsilon^2}\int_0^n F(t)\,dt + F(n); \end{aligned}$$

* Wm. Feller, *An Introduction to Probability Theory and Its Applications, Vol. II*, J. Wiley Series in Probability and Math. Stat. (1968).

and conclude that $\left|\overline{S}_n - m_n\right| \longrightarrow 0$ in $\mathbb{P}$-probability. (See part (**ii**) of Exercises 1.4.25 and 1.5.12 for a partial converse to this statement.)

Hint: Use the formula

$$\operatorname{var}(Y) \le E^{\mathbb{P}}\left[Y^2\right] = 2\int_{[0,\infty)} t\, P\bigl(|Y| > t\bigr)\, dt.$$

EXERCISE 1.2.13. Show that, for each $T \in [0,\infty)$ and $t \in (0,\infty)$,

$$\lim_{n\to\infty} e^{-nt} \sum_{k \le nT} \frac{(nt)^k}{k!} = \begin{cases} 1 & \text{if} \quad T > t \\ 0 & \text{if} \quad T < t. \end{cases}$$

Hint: Let $X_1, \dots, X_n, \dots$ be $\mathbb{P}$-independent **Poisson random variables on** $\mathbb{N}$ **with mean-value** t. That is, the X_n's are $\mathbb{P}$-independent and

$$\mathbb{P}\bigl(X_n = k\bigr) = e^{-t}\frac{t^k}{k!} \quad \text{for} \quad k \in \mathbb{N}.$$

Show that S_n is a Poisson random variable on $\mathbb{N}$ with mean-value nt, and conclude that, for each $T \in [0,\infty)$ and $t \in (0,\infty)$,

$$e^{-nt} \sum_{k \le nT} \frac{(nt)^k}{k!} = P\bigl(\overline{S}_n \le T\bigr).$$

EXERCISE 1.2.14. Given a right-continuous function $F : [0,\infty) \longrightarrow \mathbb{R}$ of bounded variation with $F(0) = 0$, define its **Laplace transform** $\varphi(\lambda)$, $\lambda \in [0,\infty)$, by the Riemann–Stieltjes integral

$$\varphi(\lambda) = \int_{[0,\infty)} e^{-\lambda t}\, dF(t).$$

Using Exercise 1.2.13, show that

$$\sum_{k \le nT} \frac{(-n)^k}{k!} \left[D^k \varphi\right](n) \longrightarrow F(T) \quad \text{as} \quad n \to \infty$$

for each $T \in [0,\infty)$ at which F is continuous. Conclude, in particular, that F can be recovered from its Laplace transform. Although this is not the most practical recovery method, it is one of the only ones that does not involve complex analysis.

§1.3 Cramér's Theory of Large Deviations

From Theorem 1.2.4, we know that if $\{X_n : n \in \mathbb{Z}^+\}$ is a sequence of $\mathbb{P}$-independent, square $\mathbb{P}$-integrable random variables with mean-value 0, and if the averages $\overline{S}_n$, $n \in \mathbb{Z}^+$, are defined accordingly, then, for every $\epsilon > 0$,

$$\mathbb{P}\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) \leq \frac{\max_{1\leq m\leq n} \operatorname{var}(X_m)}{n\epsilon^2}, \qquad n \in \mathbb{Z}^+.$$

Thus, so long as

$$\frac{\operatorname{var}(X_n)}{n} \longrightarrow 0 \text{ as } n \to \infty,$$

the $\overline{S}_n$'s are becoming more and more concentrated near 0, and the rate at which this concentration is occurring can be estimated in terms of the variances $\operatorname{var}(X_n)$. In this section, we will show that, by placing more stringent integrability requirements on the X_n's, one can gain more information about the rate at which the $\overline{S}_n$'s are concentrating.

In all of this analysis, the trick is to see how independence can be combined with 0 mean-value to produce unexpected cancellations; and, as a preliminary warm-up exercise, we begin with the following.

THEOREM 1.3.1. *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of $\mathbb{P}$-independent, $\mathbb{P}$-integrable random variables with mean-value 0, and assume that*

$$M_4 \equiv \sup_{n\in\mathbb{Z}^+} \mathbb{E}^{\mathbb{P}}\big[X_n^4\big] < \infty.$$

Then, for each $\epsilon > 0$,

$$\epsilon^4 P\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) \leq \mathbb{E}^{\mathbb{P}}\Big[\,\overline{S}_n^{\,4}\Big] \leq \frac{3M_4}{n^2}, \quad n \in \mathbb{Z}^+; \tag{1.3.2}$$

In particular, $\overline{S}_n \longrightarrow 0$ $\mathbb{P}$-almost surely.

PROOF: Obviously, in order to prove (1.3.2), it suffices to check the second inequality, which is equivalent to $\mathbb{E}^{\mathbb{P}}\big[S_n^4\big] \leq 3M_4 n^2$. But

$$\mathbb{E}^{\mathbb{P}}\big[S_n^4\big] = \sum_{m_1,\dots,m_4=1}^{n} \mathbb{E}^{\mathbb{P}}\big[X_{m_1}\cdots X_{m_4}\big],$$

and, by Schwarz's inequality, each of these terms is dominated by M_4. In addition, of these terms, the only ones which do not vanish either all their factors the same or two pairs of equal factors. Thus, the number of non-vanishing terms is $n + 3n(n-1) = 3n^2 - 2n$.

Given (1.3.2), the proof of the last part becomes an easy application of the Borel–Cantelli Lemma. Indeed, for any $\epsilon > 0$, we know from (1.3.2) that

$$\sum_{n=1}^{\infty} P\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) < \infty,$$

and therefore, by (1.1.4), that

$$\mathbb{P}\left(\varlimsup_{n\to\infty} \big|\overline{S}_n\big| \geq \epsilon\right) = 0. \quad \square$$

Remark 1.3.3. The final assertion in Theorem 1.3.1 is a primitive version of The Strong Law of Large Numbers and represents the first time that we have actually used the simultaneous existence of infinitely many mutually independent random variables (previously, and for the rest of this section, it will be enough to know that there are, at any given moment, an arbitrary but finite number). Although The Strong Law will be taken up again, and considerably refined, in Section 1.4, the principle on which its proof here was based is an important one: namely, *control more moments and you will get better estimates; get better estimates and you will reach more refined conclusions.*

With the preceding adage in mind, we will devote the rest of this section to examining what one can say when one has all moments at one's disposal. In fact, from now on, we will be assuming that $X_1, \dots, X_n, \dots$ are independent random variables with common distribution μ having the property that the **moment generating function**

$$M_\mu(\xi) \equiv \int_{\mathbb{R}} e^{\xi\, x}\, \mu(dx) < \infty \quad \text{for all } \xi \in \mathbb{R}. \tag{1.3.4}$$

Obviously, (1.3.4) is more than sufficient to guarantee that the X_n's have moments of all orders. In fact, as an application of Lebesgue's Dominated Convergence Theorem, one sees that $\xi \in \mathbb{R} \longmapsto M(\xi) \in (0,\infty)$ is infinitely differentiable and that

$$\mathbb{E}^{\mathbb{P}}\big[X_1^n\big] = \int_{\mathbb{R}} x^n\, \mu(dx) = \frac{d^n M}{d\xi^n}(0) \quad \text{for all } n \in \mathbb{N}.$$

In the discussion which follows, we will use m and σ^2 to denote, respectively, the common mean-value and variance of the X_n's.

In order to develop some intuition for the considerations which follow, we first consider an example, which, for many purposes, is *the canonical example* in probability theory. Namely, let $g : \mathbb{R} \longrightarrow (0,\infty)$ be the **Gauss kernel**

$$g(y) \equiv \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{|y|^2}{2}\right], \quad y \in \mathbb{R}; \tag{1.3.5}$$

and recall that a random variable X is **standard normal** if

$$\mathbb{P}\big(X \in \Gamma\big) = \int_\Gamma g(y)\,dy, \quad \Gamma \in \mathcal{B}_\mathbb{R}.$$

In spite of their somewhat insultingly bland moniker, standard normal random variables are the building blocks for the most honored family in all of probability theory. Indeed, given $m \in \mathbb{R}$ and $\sigma \in [0,\infty)$, the random variable Y is said to be **normal** (or **Gaussian**) **with mean-value** m **and variance** σ^2 (often this is abbreviated by saying that X is an $\mathcal{N}\big(m,\sigma^2\big)$**-random variable**) if and only if the distribution of Y is γ_{m,σ^2}, where γ_{m,σ^2} is the distribution of variable $\sigma X + m$ when X is standard normal. That is, Y is an $\mathcal{N}(m,\sigma^2)$ random variable if, when $\sigma = 0$, $\mathbb{P}(Y = m) = 1$ and, when $\sigma > 0$,

$$\mathbb{P}\big(Y \in \Gamma\big) = \int_\Gamma \frac{1}{\sigma} g\left(\frac{y-m}{\sigma}\right)\,dy \quad \text{for } \Gamma \in \mathcal{B}_\mathbb{R}.$$

There are two obvious reasons for the honored position held by Gaussian random variables. In the first place, they certainly have finite moment generating functions. In fact, since

$$\int_\mathbb{R} e^{\xi\,y}\,g(y)\,dy = \exp\left(\frac{\xi^2}{2}\right), \quad \xi \in \mathbb{R},$$

it is clear that

$$M_{\gamma_{m,\sigma^2}}(\xi) = \exp\left[\xi m + \frac{\sigma^2\xi^2}{2}\right], \tag{1.3.6}$$

Secondly, *they add nicely*. To be precise, it is a familiar fact from elementary probability theory that *if X is an $\mathcal{N}(m,\sigma^2)$ random variable and $\hat{X}$ is an $\mathcal{N}(\hat{m},\hat{\sigma}^2)$ random variable which is independent of X, then $X + \hat{X}$ is an $\mathfrak{N}\big(m+\hat{m},\sigma^2+\hat{\sigma}^2\big)$ random variable.* In particular, if $X_1,\dots,X_n$ are mutually independent standard normal random variables, then $\overline{S}_n$ is an $\mathcal{N}\left(0,\frac{1}{n}\right)$ random variable. That is,

$$\mathbb{P}\big(\overline{S}_n \in \Gamma\big) = \sqrt{\frac{n}{2\pi}}\int_\Gamma \exp\left[-\frac{n|y|^2}{2}\right]\,dy.$$

Thus (cf. Exercise 1.3.16 below), for any Γ we see that

$$\lim_{n\to\infty}\frac{1}{n}\log\Big[P\big(\overline{S}_n \in \Gamma\big)\Big] = -\text{ess inf}\left\{\frac{|y|^2}{2} : y \in \Gamma\right\}. \tag{1.3.7}$$

where the *ess* in (1.3.7) stands for *essential* and means that what follows is taken *modulo a set of measure* 0. (Hence, apart from a minus sign, the right-hand side of (1.3.7) is the greatest number dominated by $\frac{|y|^2}{2}$ for Lebesgue-almost every $y \in \Gamma$.) In fact, because

$$\int_x^\infty g(y)\,dy \le x^{-1} g(x) \quad \text{for all } x \in (0,\infty),$$

we have the rather precise upper bound

$$\mathbb{P}\big(|\overline{S}_n| \ge \epsilon\big) \le \sqrt{\frac{2}{n\pi\epsilon^2}}\,\exp\left[-\frac{n\epsilon^2}{2}\right] \quad \text{for } \epsilon > 0.$$

At the same time, it is clear that, for $0 < \epsilon < |a|$,

$$P\big(|\overline{S}_n - a| < \epsilon\big) \ge \sqrt{\frac{2\epsilon n}{\pi}}\,\exp\left[-\frac{n(|a|+\epsilon)^2}{2}\right].$$

More generally, if the X_n's are mutually independent $\mathcal{N}(m,\sigma^2)$-random variables, then one finds that

$$\mathbb{P}\big(|\overline{S}_n - m| \ge \sigma\epsilon\big) \le \sqrt{\frac{2}{n\pi\epsilon^2}}\,\exp\left[-\frac{n\epsilon^2}{2}\right] \quad \text{for } \epsilon > 0,$$

and, for $0 < \epsilon < |a|$ and sufficiently large n's

$$P\big(|\overline{S}_n - (m+a)| < \sigma\epsilon\big) \ge \sqrt{\frac{2\epsilon n}{\pi}}\,\exp\left[-\frac{n(|a|+\epsilon)^2}{2}\right].$$

Of course, in general one cannot hope to get such explicit expressions for the distribution of $\overline{S}_n$. Nonetheless, on the basis of the preceding, one can start to see what is going on. Namely, when the distribution μ falls off rapidly outside of compacts, averaging n independent random variables with distribution μ has the effect of *building an exponentially deep well in which the mean-value m lies at the bottom.* More precisely, if one believes that the Gaussian random variables *are normal* in the sense that they are typical, then one should conjecture that, even when the random variables are not normal, the behavior of $\mathbb{P}\big(\big|\overline{S}_n - m\big| \ge \epsilon\big)$ for large n's should resemble that of Gaussians with the same variance; and it is in the verification of this conjecture that the moment generating function M_μ plays a central rôle. Namely, although an expression in terms of μ for the distribution of S_n is seldom readily available, the moment generating function for S_n is easily expressed in terms of M_μ. To wit, as a trivial application of independence, we have:

$$\mathbb{E}^{\mathbb{P}}\left[e^{\xi S_n}\right] = M_\mu(\xi)^n, \quad \xi \in \mathbb{R}.$$

Hence, by Markov's inequality applied to $e^{\xi S_n}$, we see that, for any $a \in \mathbb{R}$,

$$\mathbb{P}\big(\overline{S}_n \geq a\big) \leq e^{-n\xi a}\, M_\mu(\xi)^n = \exp\big[-n\big(\xi a - \Lambda_\mu(\xi)\big)\big], \quad \xi \in [0,\infty),$$

where

$$\Lambda_\mu(\xi) \equiv \log\big(M_\mu(\xi)\big) \tag{1.3.8}$$

is the **logarithmic moment generating function of** μ. The preceding relation is one of those lovely situations in which a single quantity is dominated by a whole family of quantities, with the result that one should optimize by minimizing over the dominating quantities. Thus, we now have

$$\mathbb{P}\big(\overline{S}_n \geq a\big) \leq \exp\left[-n \sup_{\xi\in[0,\infty)} \big(\xi a - \Lambda_\mu(\xi)\big)\right]. \tag{1.3.9}$$

Notice that (1.3.9) is really very good. For instance, when the X_n's are $\mathcal{N}(m,\sigma^2)$-random variables and $\sigma > 0$, then (cf. (1.3.6)) the preceding leads quickly to the estimate

$$\mathbb{P}\big(\overline{S}_n - m \geq \epsilon\big) \leq \exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right),$$

which is essentially the upper bound at which we arrived before.

Taking a hint from the preceding, we now introduce the **Legendre transform**

$$I_\mu(x) \equiv \sup\big\{\xi x - \Lambda_\mu(\xi) : \xi \in \mathbb{R}\big\}, \qquad x \in \mathbb{R}, \tag{1.3.10}$$

of Λ_μ and, before proceeding further, make some elementary observations about the structure of the functions Λ_μ and I_μ.

LEMMA 1.3.11. *The function Λ_μ is infinitely differentiable. In addition, for each $\xi \in \mathbb{R}$, the probability measure ν_ξ on $\mathbb{R}$ given by*

$$\nu_\xi(\Gamma) = \frac{1}{M_\mu(\xi)} \int_\Gamma e^{\xi x}\, \mu(dx) \quad \text{for } \Gamma \in \mathcal{B}_\mathbb{R}$$

has moments of all orders,

$$\int_\mathbb{R} x\, \nu_\xi(dx) = \Lambda_\mu'(\xi), \quad \text{and} \quad \int_\mathbb{R} x^2\, \nu_\xi(dx) - \left(\int_\mathbb{R} x\, \nu_\xi(dx)\right)^2 = \Lambda_\mu''(\xi).$$

Next, the function I_μ is a $[0,\infty]$-valued, lower semicontinuous, convex function which vanishes at m . Moreover,

$$I_\mu(x) = \sup\big\{\xi x - \Lambda_\mu(\xi) : \xi \geq 0\big\} \quad \text{for} \quad x \in [m,\infty)$$

and

$$I_\mu(x) = \sup\{\xi x - \Lambda_\mu(\xi) : \xi \le 0\} \quad \textit{for} \quad x \in (-\infty, m].$$

Finally, if

$$\alpha = \inf\{x \in \mathbb{R} : \mu\big((-\infty, x]\big) > 0\}$$

and

$$\beta = \sup\{x \in \mathbb{R} : \mu\big([x, \infty)\big) > 0\},$$

then I_μ *is smooth on* (α, β) *and identically* $+\infty$ *off of* $[\alpha, \beta]$*. In fact, either* $\mu(\{m\}) = 1$ *and* $\alpha = m = \beta$*; or* $m \in (\alpha, \beta)$ *and* Λ'_μ *is a smooth, strictly increasing mapping from* $\mathbb{R}$ *onto* (α, β)*,*

$$I_\mu(x) = \Xi_\mu(x)\, x - \Lambda_\mu\big(\Xi_\mu(x)\big),\ x \in (\alpha, \beta), \quad \textit{where} \quad \Xi_\mu = \big(\Lambda'_\mu\big)^{-1}$$

is the inverse of Λ'_μ*,* $\mu(\{\alpha\}) = e^{-I_\mu(\alpha)}$ *if* $\alpha > -\infty$*, and* $\mu(\{\beta\}) = e^{-I_\mu(\beta)}$ *if* $\beta < \infty$*.*

Proof: For notational convenience, we will drop the subscript "μ" during the proof. Further, we remark that the smoothness of Λ follows immediately from the positivity and smoothness of M, and the identification of $\Lambda'(\xi)$ and $\Lambda''(\xi)$ with the mean and variance of ν_ξ is elementary calculus combined with the remark following (1.3.4). Thus, we will concentrate on the properties of the function I.

As the pointwise supremum of functions which are linear, I is certainly lower semicontinuous and convex. Also, because $\Lambda(0) = 0$, it is obvious that $I \ge 0$. Next, by Jensen's inequality,

$$\Lambda(\xi) \ge \xi \int_{\mathbb{R}} x\, \mu(dx) = \xi\, m,$$

and, therefore, $\xi x - \Lambda(\xi) \le 0$ if $x \le m$ and $\xi \ge 0$ or if $x \ge m$ and $\xi \le 0$. Hence, because I is nonnegative, this proves the one-sided extremal characterization of $I_\mu(x)$.

Turning to the final part, note first that there is nothing more to do in the case when $\mu(\{m\}) = 1$. Thus, we will assume that $\mu(\{m\}) < 1$, in which case it is clear that $m \in (\alpha, \beta)$ and that none of the measures ν_ξ is degenerate. In particular, because $\Lambda''(\xi)$ is the variance of the ν_ξ, we know that $\Lambda'' > 0$ everywhere. Hence, Λ' is strictly increasing and therefore admits a smooth inverse Ξ on its image. Furthermore, because $\Lambda'(\xi)$ is the mean of ν_ξ, it is clear that the image of Λ' is contained in (α, β). At the same time, given an $x \in [m, \beta)$, choose $y \in (x, \beta)$ and note that, for $\xi \ge 0$,

$$\Lambda(\xi) \ge \xi\, y - \kappa \quad \text{where} \quad \kappa = -\log\big[\mu\big([y, \infty)\big)\big].$$

After combining this with the fact (already established) that $\xi x - \Lambda(\xi) \le 0$ for $\xi \le 0$, we conclude that $\xi \in \mathbb{R} \longmapsto \xi x - \Lambda(\xi)$ achieves its absolute maximum somewhere in the interval $\left[0, \frac{\kappa}{y-x}\right]$ and therefore that $\Lambda'(\xi) = x$ for some ξ in that interval. Since an analogous argument applies when $x \in (\alpha, m]$, we now know that (α, β) is precisely the image of Λ'. Finally, because (by convexity) $I(x) = \xi x - \Lambda(\xi)$ if and only if $\Lambda'(\xi) = x$, we have also proved that I is given on (α, β) by the asserted expression.

To complete the proof, suppose that $\beta < \infty$. Then

$$e^{\xi\beta}\mu\big(\{\beta\}\big) \le M(\xi), \quad \xi \in \mathbb{R}.$$

Thus, on the one hand, we have that $\mu(\{\beta\}) \le e^{-I(\beta)}$. On the other hand, because

$$e^{-I(\beta)} \le \int_{\mathbb{R}} e^{\xi(x-\beta)}\,\mu(dx) \quad \text{for} \quad \xi \in [0, \infty)$$

and

$$\int_{\mathbb{R}} e^{\xi(x-\beta)}\,\mu(dx) \searrow \mu(\{\beta\}) \quad \text{as} \quad \xi \nearrow \infty,$$

we also see that $\mu(\{\beta\}) \ge e^{-I(\beta)}$. Finally, if $x \in (\beta, \infty)$, then $I(x) = \infty$ follows immediately from $\Lambda(\xi) \le \xi\,\beta$, $\xi \in [0, \infty)$.

Since the same reasoning applies when $\alpha > -\infty$, we are done. $\square$

THEOREM 1.3.12 (**Cramér's Theorem**). *Let $\{X_n\}_1^\infty$ be a sequence of $\mathbb{P}$-independent random variables with common distribution μ, assume that the associated moment generating function M_μ satisfies (1.3.4), set $m = \int_{\mathbb{R}} x\,\mu(dx)$, and define I_μ accordingly, as in (1.3.10). Then,*

$$P\big(\overline{S}_n \ge a\big) \le e^{-nI_\mu(a)} \quad \textit{for all } a \in [m, \infty),$$
$$P\big(\overline{S}_n \le a\big) \le e^{-nI_\mu(a)} \quad \textit{for all } a \in (-\infty, m].$$

Moreover, for $a \in (\alpha, \beta)$ (cf. Lemma 1.3.11), $\epsilon > 0$, and $n \in \mathbb{Z}^+$,

$$\mathbb{P}\Big(\big|\overline{S}_n - a\big| < \epsilon\Big) \ge \left(1 - \frac{\Lambda''_\mu\big(\Xi_\mu(a)\big)}{n\epsilon^2}\right)\exp\Big[-n\Big(I_\mu(a) + \epsilon|\Xi_\mu(a)|\Big)\Big],$$

where Λ_μ is the function given in (1.3.8) and $\Xi_\mu \equiv \big(\Lambda_\mu{}'\big)^{-1}$.

PROOF: To prove the first part, suppose that $a \in [m, \infty)$, and apply the second part of Lemma 1.3.11 to see that the exponent in (1.3.9) equals $I_\mu(a)$, and, after replacing $\{X_n\}_1^\infty$ by $\{-X_n\}_1^\infty$, also get the desired estimate when $a \le m$.

To prove the lower bound, let $a \in [m, \beta)$ be given, and set $\xi = \Xi_\mu(a) \in [0, \infty)$. Next, recall the probability measure ν_ξ described in Lemma 1.3.11, and remember that ν_ξ has mean $a = \Lambda'_\mu(\xi)$ and variance $\Lambda''_\mu(\xi)$. Further, if $\{Y_n : n \in \mathbb{Z}^+\}$ is a sequence of independent, identically distributed random variables with common distribution ν_ξ, then it is an easy matter to check that, for any $n \in \mathbb{Z}^+$ and every $\mathcal{B}_{\mathbb{R}^n}$-measurable $F : \mathbb{R}^n \longrightarrow [0,\infty)$,

$$\mathbb{E}^{\mathbb{P}}\Big[F\big(Y_1,\dots,Y_n\big)\Big] = \frac{1}{M_\mu(\xi)^n}\mathbb{E}^{\mathbb{P}}\Big[e^{\xi S_n}\, F\big(X_1,\dots,X_n\big)\Big].$$

In particular, if

$$T_n = \sum_{\ell=1}^{n} Y_\ell \quad \text{and} \quad \overline{T}_n = \frac{T_n}{n},$$

then, because $I_\mu(a) = \xi a - \Lambda_\mu(\xi)$,

$$\begin{aligned} P\Big(\big|\overline{S}_n - a\big| < \epsilon\Big) &= M(\xi)^n \mathbb{E}^{\mathbb{P}}\Big[e^{-\xi T_n},\ \big|\overline{T}_n - a\big| < \epsilon\Big] \\ &\geq e^{-n\xi(a+\epsilon)}\, M(\xi)^n\, P\Big(\big|\overline{T}_n - a\big| < \epsilon\Big) \\ &= \exp\Big[-n\Big(I_\mu(a) + \xi\epsilon\Big)\Big] P\Big(\big|\overline{T}_n - a\big| < \epsilon\Big). \end{aligned}$$

But, because the mean-value and variance of the Y_n's are, respectively, a and $\Lambda''_\mu(\xi)$, (1.2.5) leads to

$$P\Big(\big|\overline{T}_n - a\big| \geq \epsilon\Big) \leq \frac{\Lambda''_\mu(\xi)}{n\epsilon^2}.$$

The case when $a \in (\alpha, m]$ is handled in the same way. $\square$

Results like the ones obtained in Theorem 1.3.12 are examples of a class of results known as **large deviations estimates.** They are *large deviations* because the probability of their occurrence is exponentially small. Although large deviation estimates are available in a variety of circumstances,* in general one has to settle for the cruder sort of information contained in the following.

Corollary 1.3.13. *For any* $\Gamma \in \mathcal{B}_{\mathbb{R}}$,

$$\begin{aligned} -\inf_{x\in\Gamma^\circ} I_\mu(x) &\leq \varliminf_{n\to\infty} \frac{1}{n}\log\Big[P\big(\overline{S}_n \in \Gamma\big)\Big] \\ &\leq \varlimsup_{n\to\infty} \frac{1}{n}\log\Big[P\big(\overline{S}_n \in \Gamma\big)\Big] \leq -\inf_{x\in\overline{\Gamma}} I_\mu(x). \end{aligned}$$

(We use Γ° and $\overline{\Gamma}$ to denote the interior and closure of a set Γ. Also, recall that we take the infimum over the empty set to be $+\infty$.)

---

* In fact, see, for example, J.-D. Deuschel and D. Stroock, *Large Deviations*, Academic Press Pure Math Series, **137** (1989); some people have written entire books on the subject.

Proof: To prove the upper bound, let Γ be a closed set and define $\Gamma_+ = \Gamma \cap [m, \infty)$ and $\Gamma_- = \Gamma \cap (-\infty, m]$. Clearly,

$$\mathbb{P}\big(\overline{S}_n \in \Gamma\big) \leq 2P\big(\overline{S}_n \in \Gamma_+\big) \vee P\big(\overline{S}_n \in \Gamma_-\big).$$

Moreover, if $\Gamma_+ \neq \emptyset$ and $a_+ = \min\{x : x \in \Gamma_+\}$, then, by Lemma 1.3.11 and Theorem 1.3.12,

$$I_\mu(a_+) = \inf\big\{I_\mu(x) : x \in \Gamma_+\big\} \quad \text{and} \quad P\big(\overline{S}_n \in \Gamma_+\big) \leq e^{-nI_\mu(a_+)}.$$

Similarly, if $\Gamma_- \neq \emptyset$ and $a_- = \max\{x : x \in \Gamma_-\}$, then

$$I_\mu(a_-) = \inf\big\{I_\mu(x) : x \in \Gamma_-\big\} \quad \text{and} \quad P\big(\overline{S}_n \in \Gamma_-\big) \leq e^{-nI_\mu(a_-)}.$$

Hence, either $\Gamma = \emptyset$, and there is nothing to do anyhow, or

$$\mathbb{P}\big(\overline{S}_n \in \Gamma\big) \leq 2\exp\left[-n\inf\big\{I_\mu(x) : x \in \Gamma\big\}\right], \quad n \in \mathbb{Z}^+,$$

which certainly implies the asserted upper bound.

To prove the lower bound, assume that Γ is a nonempty open set. What we have to show is that

$$\varliminf_{n\to\infty} \frac{1}{n}\log\Big[P\big(\overline{S}_n \in \Gamma\big)\Big] \geq -I_\mu(a)$$

for every $a \in \Gamma$. If $a \in \Gamma \cap (\alpha, \beta)$, choose $\delta > 0$ so that $(a - \delta, a + \delta) \subseteq \Gamma$ and use the second part of Theorem 1.3.12 to see that

$$\varliminf_{n\to\infty} \frac{1}{n}\log\Big[P\big(\overline{S}_n \in \Gamma\big)\Big] \geq -I_\mu(a) - \epsilon\big|\Xi_\mu(a)\big|$$

for every $\epsilon \in (0, \delta)$. If $a \notin [\alpha, \beta]$, then $I_\mu(a) = \infty$, and so there is nothing to do. Finally, if $a \in \{\alpha, \beta\}$, then $\mu(\{a\}) = e^{-I_\mu(a)}$ and therefore

$$\mathbb{P}\big(\overline{S}_n \in \Gamma\big) \geq P\big(\overline{S}_n = a\big) \geq e^{-nI_\mu(a)}. \quad \square$$

Remark 1.3.14. The upper bound in Theorem 1.3.12 is often called **Chernoff's Inequality**. The idea underlying this estimate is rather mundane by comparison to the subtle one used in the proof of the lower bound. Indeed, it may not be immediately obvious what that idea was! Thus, consider once again the second part of the proof of Theorem 1.3.12. What we had to do is estimate the probability that $\overline{S}_n$ lies in a neighborhood of a. When a is the mean-value m, such an estimate is provided by The Weak Law. On the other hand, when $a \neq m$, The Weak Law for the X_n's has very little to contribute. Thus, what we did is replace the original X_n's by random variables Y_n, $n \in \mathbb{Z}^+$, whose mean-value is a. Furthermore, the transformation from the X_n's to the Y_n's was sufficiently simple that it was easy to estimate X_n-probabilities in terms of Y_n-probabilities. Finally, The Weak Law applied to the Y_n's gave strong information about the rate of approach of $\frac{1}{n}\sum_{\ell=1}^n Y_\ell$ to a.

We close this section by verifying the conjecture (cf. the discussion preceding Lemma 1.3.11) that the Gaussian case is *normal*. In particular, we want to check that the *well around* m in which the distribution of $\overline{S}_n$ becomes concentrated looks Gaussian, and, in view of Theorem 1.3.12, this comes down to the following.

THEOREM 1.3.15. *Let everything be as in Lemma 1.3.11 and assume that the variance $\sigma^2 > 0$. There exists a $\delta > 0$ and a $K \in (0,\infty)$ such that $[m-\delta, m+\delta] \subseteq (\alpha, \beta)$ (cf. Lemma 1.3.11), $\left|\Lambda''_\mu\big(\Xi(x)\big)\right| \le K$,*

$$\left|\Xi_\mu(x)\right| \le K|x-m|, \quad \text{and} \quad \left|I_\mu(x) - \frac{(x-m)^2}{2\sigma^2}\right| \le K|x-m|^3$$

for all $x \in [m-\delta, m+\delta]$. In particular, if $0 < \epsilon < \delta$, then

$$\mathbb{P}\big(|\overline{S}_n - m| \ge \epsilon\big) \le 2\exp\left[-n\left(\frac{\epsilon^2}{2\sigma^2} - K\epsilon^3\right)\right],$$

and if $|a-m| < \delta$ and $\epsilon > 0$, then

$$\mathbb{P}\big(|\overline{S}_n - a| < \epsilon\big) \ge \left(1 - \frac{K}{n\epsilon^2}\right)\exp\left[-n\left(\frac{|a-m|^2}{2\sigma^2} + K|a-m|\big(\epsilon + |a-m|^2\big)\right)\right].$$

PROOF: Without loss in generality (cf. Exercise 1.3.17 below), we will assume that $m = 0$ and $\sigma^2 = 1$. Since, in this case, $\Lambda_\mu(0) = \Lambda'_\mu(0) = 0$ and $\Lambda''_\mu(0) = 1$, it follows that $\Xi_\mu(0) = 0$ and $\Xi'_\mu(0) = 1$. Hence, we can find an $M \in (0,\infty)$ and a $\alpha < -\delta < \delta < \beta$ for which $\left|\Xi_\mu(x) - x\right| \le M|x|^2$ and $\left|\Lambda_\mu(\xi) - \frac{\xi^2}{2}\right| \le M|\xi|^3$ whenever $|x| \le \delta$ and $|\xi| \le (M+1)\delta$, respectively. In particular, this leads immediately to $\left|\Xi_\mu(x)\right| \le (M+1)|x|$ for $|x| \le \delta \wedge 1$; and the estimate for I_μ comes easily from the preceding combined with equation $I_\mu(x) = \Xi(x)x - \Lambda_\mu\big(\Xi_\mu(x)\big)$. $\square$

Exercises for § 1.3

EXERCISE 1.3.16. Let $\big(E, \mathcal{F}, \mu\big)$ be a measurable space and f a nonnegative, $\mathcal{F}$-measurable function. If either $\mu(E) < \infty$ or f is μ-integrable, show that

$$\|f\|_{L^p(\mu)} \longrightarrow \|f\|_{L^\infty(\mu)} \quad \text{as} \quad p \to \infty.$$

Hint: Handle the case $\mu(E) < \infty$ first, and handle the case when $f \in L^1(\mu)$ by considering the measure $\nu(dx) = f(x)\,\mu(dx)$.

EXERCISE 1.3.17. Referring to the notation used in this section, assume that μ is a nondegenerate (i.e., it is not concentrated at a single point) probability measure on $\mathbb{R}$ for which (1.3.4) holds. Next, let m and σ^2 be the mean and variance of μ, use ν to denote the distribution of

$$x \in \mathbb{R} \longmapsto \frac{x-m}{\sigma} \in \mathbb{R} \quad \text{under} \quad \mu,$$

and define Λ_ν, I_ν, and Ξ_ν accordingly. Show that

$$\begin{aligned}
\Lambda_\mu(\xi) &= \xi m + \Lambda_\nu(\sigma\xi), \qquad \xi \in \mathbb{R},\\
I_\mu(x) &= I_\nu\left(\frac{x-m}{\sigma}\right), \qquad x \in \mathbb{R},\\
\operatorname{Image}\bigl(\Lambda'_\mu\bigr) &= m + \sigma \operatorname{Image}\bigl(\Lambda'_\nu\bigr),\\
\Xi_\mu(x) &= \frac{1}{\sigma}\Xi_\nu\left(\frac{x-m}{\sigma}\right), \quad x \in \operatorname{Image}\bigl(\Lambda'_\mu\bigr).
\end{aligned}$$

Exercise 1.3.18. Continue with the same notation.

(**i**) Show that $I_\nu \le I_\mu$ if $M_\mu \le M_\nu$.

(**ii**) Show that

$$I_\mu(x) = \frac{(x-m)^2}{2\sigma^2}, \qquad x \in \mathbb{R},$$

when μ is the $\mathcal{N}\bigl(m,\sigma^2\bigr)$ distribution and show that

$$I_\mu(x) = \frac{x-a}{b-a}\log\frac{x-a}{(1-p)(b-a)} + \frac{b-x}{b-a}\log\frac{b-x}{p(b-a)}, \qquad x \in (a,b),$$

when $a < b$, $p \in (0,1)$, and $\mu(\{a\}) = 1 - \mu(\{b\}) = p$.

(**iii**) When μ is the centered Bernoulli distribution given by $\mu\bigl(\{\pm1\}\bigr) = \frac{1}{2}$, show that $M_\mu(\xi) \le \exp\left[\frac{\xi^2}{2}\right]$, $\xi \in \mathbb{R}$, and conclude that $I_\mu(x) \ge \frac{x^2}{2}$, $x \in \mathbb{R}$. More generally, given $n \in \mathbb{Z}^+$, $\bigl\{\sigma_k\bigr\}_1^n \subseteq \mathbb{R}$, and independent random variables $X_1,\dots,X_n$ with this μ as their common distribution, let ν denote the distribution of $S \equiv \sum_1^n \sigma_k X_k$ and show that $I_\nu(x) \ge \frac{x^2}{2\Sigma^2}$, where $\Sigma^2 \equiv \sum_1^n \sigma_k^2$. In particular, conclude that

$$\mathbb{P}\Bigl(|S| \ge a\Bigr) \le 2\exp\left[-\frac{a^2}{2\Sigma^2}\right], \quad a \in [0,\infty).$$

Exercise 1.3.19. Although it is not exactly the direction in which we have been going, it seems appropriate to include here a derivation of **Stirling's formula**. Namely, recall **Euler's Gamma function**

$$\Gamma(t) \equiv \int_{[0,\infty)} x^{t-1}e^{-x}\,dx, \qquad t \in (-1,\infty). \tag{1.3.20}$$

What we want to prove is that

$$\Gamma(t+1) \sim \sqrt{2\pi t}\left(\frac{t}{e}\right)^t \quad \text{as} \quad t \nearrow \infty, \tag{1.3.21}$$

where the *tilde* "$\sim$" means that the two sides are **asymptotic** to one another in the sense that their ratio tends to 1. (See Exercise 2.1.39 for another approach.)

The first step is to make the problem look like one to which Exercise 1.3.16 is applicable. Thus, make the substitution $x = ty$ and apply Exercise 1.3.16 to see that

$$\left(\frac{\Gamma(t+1)}{t^{t+1}}\right)^{\frac{1}{t}} = \left(\int_{[0,\infty)} y^t\, e^{-ty}\, dy\right)^{\frac{1}{t}} \longrightarrow e^{-1}.$$

This is, of course, far less than we need to know. However, it does show that all the *action* is going to take place near $y = 1$ and that the principal factor in the asymptotics of $\frac{\Gamma(t+1)}{t^{t+1}}$ is e^{-t}. In order to highlight these observations, make the substitution $y = z + 1$ and obtain

$$\frac{\Gamma(t+1)}{t^{t+1}e^{-t}} = \int_{(-1,\infty)} (1+z)^t\, e^{-tz}\, dz.$$

Before taking the next step, introduce the function $R(z) = \log(1+z) - z + \frac{z^2}{2}$ for $z \in (-1,1)$, and check that $R(z) \le 0$ if $z \le 0$ and that $|R(z)| \le \frac{|z|^3}{3(1-|z|)}$. Now let $\delta \in (0,1)$ be given, and show that

$$\int_{-1}^{-\delta} \big(1+z\big)^t\, e^{-tz}\, dz \le (1-\delta)\big[(1-\delta)e^{-\delta}\big] \le \exp\left[-\frac{t\delta^2}{2}\right],$$

and

$$\begin{aligned}\int_{\delta}^{\infty} \big(1+z\big)^t\, e^{-tz}\, dz &\le \Big[\big(1+\delta\big)e^{-\delta}\Big]^{t-1} \int_{\delta}^{\infty} (1+z)e^{-z}\, dz \\ &\le \exp\left[1 - \frac{t\delta^2}{2} + \frac{\delta^3}{3(1-\delta)}\right].\end{aligned}$$

Next, write $(1+z)^t e^{-tz} = e^{-\frac{tz^2}{2}} e^{tR(z)}$. Then

$$\int_{|z|\le\delta} \big(1+z\big)^t\, e^{-tz}\, dz = \int_{|z|\le\delta} e^{-\frac{tz^2}{2}}\, dz + E(t,\delta),$$

where

$$E(t,\delta) = \int_{|z|\le\delta} e^{-\frac{tz^2}{2}} \big(e^{tR(z)} - 1\big)\, dz.$$

Check that

$$\left|\int_{|z|\le\delta} e^{-\frac{tz^2}{2}}\, dz - \sqrt{\frac{2\pi}{t}}\right| = t^{-\frac{1}{2}} \int_{|z|\ge t^{\frac{1}{2}}\delta} e^{-\frac{z^2}{2}}\, dz \le \frac{2}{t^{\frac{1}{2}}\delta} e^{-\frac{t\delta^2}{2}}.$$

At the same time, show that

$$|E(t,\delta)| \le t\int_{|z|\le\delta} |R(z)|e^{-\frac{tz^2}{2}+|R(z)|}\,dz \le t\int_{|z|\le\delta} |z|^3 e^{-\frac{tz^2}{2}\frac{3-5\delta}{3(1-\delta)}}\,dz \le \frac{12(1-\delta)}{(3-5\delta)^2 t}$$

as long as $\delta < \frac{3}{5}$. Finally, take $\delta = \sqrt{2t^{-1}\log t}$, combine these to conclude that there there is a $C < \infty$ such that

$$\left|\frac{\Gamma(t+1)}{\sqrt{2\pi t}\left(\frac{t}{e}\right)^t} - 1\right| \le \frac{C}{t}, \quad t \in [1,\infty).$$

EXERCISE 1.3.22. Here is a rather different sort of application of large deviation estimates. Namely, inspired by T.H. Carne,* we will show that for each $n \in \mathbb{Z}^+$ and $1 \le m < n$ there exists an $(m-1)$st order polynomial $p_{m,n}$ with the property that

$$\left|x^n - p_{m,n}(x)\right| \le 2\exp\left[-\frac{m^2}{2n}\right] \quad \text{for } x \in [-1,1].$$

(**i**) Given a $\mathbb{C}$-valued f on $\mathbb{Z}$, define $\Delta f : \mathbb{Z} \longrightarrow \mathbb{C}$ by

$$\mathcal{A}f(n) = \frac{f(n+1)+f(n-1)}{2}, \quad n \in \mathbb{Z},$$

and show that, for any $n \ge 1$, $\mathcal{A}^n f = \mathbb{E}^{\mathbb{P}}\big[f(S_n)\big]$, where S_n is the sum of n $\mathbb{P}$-independent, Bernoulli random variables.

(**ii**) Show that, for each $z \in \mathbb{C}$, there is a unique sequence $\{Q(m,z) : m \in \mathbb{Z}\} \subseteq \mathbb{C}$ satisfying $Q(0,z) = 1$,

$$Q(-m,z) = Q(m,z), \quad \text{and } \big[\mathcal{A}Q(\,\cdot\,,z)\big](m) = zQ(m,z) \text{ for all } m \in \mathbb{Z}.$$

In fact, show that, for each $m \in \mathbb{Z}^+$: $Q(m,\,\cdot\,)$ is a polynomial of degree m and

$$Q(m,\cos\theta) = \cos(m\theta), \quad \theta \in \mathbb{C}.$$

In particular, this means that $|Q(n,x)| \le 1$ for all $x \in [-1,1]$. (It also means that $Q(n,\,\cdot\,)$ is the nth **Chebychev polynomial**.)

* T.H. Carne, "A transformation formula for Markov chains," *Bull. Sc. Math.*, **109**: 399–405 (1985). As Carne points out, what he is doing is the discrete analog of Hadamard's representation, via the Weierstrass transform, of solutions to heat equations in terms of solutions to the wave equations.

(**iii**) Using induction on $n \in \mathbb{Z}^+$, show that

$$\big[\mathcal{A}^n Q(\,\cdot\,, z)\big](m) = z^n Q(m, z), \quad m \in \mathbb{Z} \text{ and } z \in \mathbb{C},$$

and conclude that

$$z^n = \mathbb{E}\Big[Q\big(S_n, z\big)\Big], \quad n \in \mathbb{Z}^+ \quad \text{and} \quad z \in \mathbb{C},$$

where S_n is the sum of n mutually independent, standard, $\{-1, 1\}$-valued Bernoulli random variables. In particular, if

$$p_{m,n}(z) \equiv \mathbb{E}\Big[Q\big(S_n, z\big),\ \big|S_n\big| < m\Big] = 2^{-n} \sum_{|2\ell - n| < m} \binom{n}{\ell} Q(2\ell - n, z),$$

conclude that (cf. Exercise 1.3.18)

$$\sup_{x \in [-1,1]} |x^n - p_{m,n}(x)| \le P\big(|S_n| \ge m\big) \le 2 \exp\left[-\frac{m^2}{2n}\right] \quad \text{for all } 1 \le m \le n.$$

(**iv**) Suppose that A is a self-adjoint contraction on the Hilbert space H (i.e., $(f, Ag)_H = \overline{(g, Af)}_H$ and $\|Af\|_H \le \|f\|_H$ for all $f,\, g \in H$). Next, assume that $\big(f, A^\ell g\big)_H = 0$ for some $f,\, g \in H$ and each $0 \le \ell < m$. Show that

$$\big|\big(f, A^n g\big)_H\big| \le 2\|f\|_H \|g\|_H \exp\left[-\frac{m^2}{2n}\right] \quad \text{for } n \ge m.$$

(See Exercise 2.2.27 for an application.)

Hint: Note that $\big(f, p_{m,n}(A)g\big)_H = 0$, and use the Spectral Theorem to see that, for any polynomial p,

$$\|p(A)f\|_H \le \sup_{x \in [-1,1]} |p(x)|\, \|f\|_H, \quad f \in H.$$

§1.4 The Strong Law of Large Numbers

In this section we will discuss a few almost sure convergence properties of partial sums of independent random variables. Thus, once again, $\big\{X_n\big\}_1^\infty$ will be a sequence of independent random variables on a probability space $\big(\Omega, \mathcal{F}, P\big)$, and S_n and $\overline{S}_n$ will be, respectively, the sum and average of $X_1, \dots, X_n$. Throughout this section, the reader should notice how much more immediately important a rôle independence (as opposed to orthogonality) plays than it did in Section 1.2.

To get started, we point out that, for both $\big\{S_n\big\}_1^\infty$ and $\big\{\overline{S}_n\big\}_1^\infty$, the set on which convergence occurs has $\mathbb{P}$-measure either 0 or 1. In fact, we have the following simple application of Kolmogorov's 0–1 Law (Theorem 1.1.2).

Lemma 1.4.1. *For any sequence* $\{a_n : n \in \mathbb{Z}^+\} \subseteq \mathbb{R}$ *and any sequence* $\{b_n : n \in \mathbb{Z}^+\} \subseteq (0,\infty)$ *which converges to an element of* $(0,\infty]$, *the set on which*

$$\lim_{n\to\infty} \frac{S_n - a_n}{b_n} \quad \text{exists in} \quad \mathbb{R}$$

has $\mathbb{P}$*-measure either* 0 *or* 1. *In fact, if* $b_n \longrightarrow \infty$ *as* $n \to \infty$, *then both*

$$\varlimsup_{n\to\infty} \frac{S_n - a_n}{b_n} \quad \text{and} \quad \varliminf_{n\to\infty} \frac{S_n - a_n}{b_n}$$

are $\mathbb{P}$*-almost surely constant.*

Proof: Simply observe that all of the events and functions involved can be expressed in terms of $\{S_{m+n} - S_m\}_{n=1}^\infty$ for each $m \in \mathbb{Z}^+$ and are therefore tail-measurable. □

Our basic result about the almost sure convergence properties of both $\{S_n\}_1^\infty$ and $\{\overline{S}_n\}_1^\infty$ is the following beautiful statement, which was proved originally by Kolmogorov.

Theorem 1.4.2. *If the* X_n*'s are independent, square* $\mathbb{P}$*-integrable random variables and if*

$$\sum_{n=1}^\infty \operatorname{var}(X_n) < \infty, \tag{1.4.3}$$

then

$$\sum_{n=1}^\infty \Big(X_n - \mathbb{E}^{\mathbb{P}}[X_n]\Big) \quad \text{converges } \mathbb{P}\text{-almost surely.}$$

Note that, since

$$\sup_{n\ge N} P\left(\left|\sum_{\ell=N}^n \Big(X_\ell - \mathbb{E}^{\mathbb{P}}[X_\ell]\Big)\right| \ge \epsilon\right) \le \frac{1}{\epsilon^2}\sum_{\ell=N}^\infty \operatorname{var}(X_\ell), \tag{1.4.4}$$

(1.4.4) certainly implies that the series $\sum_{n=1}^\infty \Big(X_n - \mathbb{E}^{\mathbb{P}}[X_n]\Big)$ converges in $\mathbb{P}$-measure. Thus, all that we are trying to do here is replace a convergence in measure statement with an almost sure one. Obviously, this replacement would be trivial if the "$\sup_{n\ge N}$" in (1.4.4) appeared on the other side of $\mathbb{P}$. The remarkable fact which we are about to prove is that, in the present situation, the "$\sup_{n\ge N}$" can be brought inside!

Theorem 1.4.5 (**(Kolmogorov's Inequality).** *If the* X_n*'s are independent and square* $\mathbb{P}$*-integrable, then*

$$\mathbb{P}\left(\sup_{n\ge 1}\left|\sum_{\ell=1}^n \Big(X_\ell - \mathbb{E}^{\mathbb{P}}[X_\ell]\Big)\right| \ge \epsilon\right) \le \frac{1}{\epsilon^2}\sum_{n=1}^\infty \operatorname{var}(X_n) \tag{1.4.6}$$

for each $\epsilon > 0$.

PROOF: Without loss in generality, assume that each X_n has mean-value 0.

Given $1 \le n < N$, note that

$$S_N^2 - S_n^2 = \big(S_N - S_n\big)^2 + 2\big(S_N - S_n\big)S_n \ge 2\big(S_N - S_n\big)S_n;$$

and therefore, since $S_N - S_n$ has mean-value 0 and is independent of the σ-algebra $\sigma\big(X_1, \dots, X_n\big)$,

$$(*) \qquad \mathbb{E}^{\mathbb{P}}\big[S_N^2,\, A_n\big] \ge \mathbb{E}^{\mathbb{P}}\big[S_n^2,\, A_n\big] \quad \text{for any } A_n \in \sigma\big(X_1, \dots, X_n\big).$$

In particular, if $A_1 = \big\{|S_1| > \epsilon\big\}$ and

$$A_{n+1} = \Big\{\big|S_{n+1}\big| > \epsilon \text{ and } \max_{1\le \ell \le n} \big|S_\ell\big| \le \epsilon\Big\}, \qquad n \in \mathbb{Z}^+,$$

then, the A_n's are mutually disjoint,

$$B_N \equiv \left\{\max_{1\le n\le N} \big|S_n\big| > \epsilon\right\} = \bigcup_{n=1}^N A_n,$$

and so (*) implies that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[S_N^2,\, B_N\big] &= \sum_{n=1}^N \mathbb{E}^{\mathbb{P}}\big[S_N^2,\, A_n\big] \ge \sum_{n=1}^N \mathbb{E}^{\mathbb{P}}\big[S_n^2,\, A_n\big] \\ &\ge \epsilon^2 \sum_{n=1}^N P\big(A_n\big) = \epsilon^2 P\big(B_N\big).\end{aligned}$$

In particular,

$$\begin{aligned}\epsilon^2 P\left(\sup_{n\ge1} \big|S_n\big| > \epsilon\right) &= \lim_{N\to\infty} \epsilon^2 P\big(B_N\big) \\ &\le \lim_{N\to\infty} \mathbb{E}^{\mathbb{P}}\big[S_N^2\big] \le \sum_{n=1}^\infty \mathbb{E}^{\mathbb{P}}\big[X_n^2\big],\end{aligned}$$

and so the result follows after one takes left limits with respect to $\epsilon > 0$. $\square$

PROOF OF 1.4.2: Again we assume that the X_n's have mean-value 0. By (1.4.6) applied to $\big\{X_{N+n} : n \in \mathbb{Z}^+\big\}$, we see that (1.4.4) implies

$$\mathbb{P}\left(\sup_{n>N} \big|S_n - S_N\big| \ge \epsilon\right) \le \frac{1}{\epsilon^2} \sum_{n=N+1}^\infty \mathbb{E}^{\mathbb{P}}\big[X_n^2\big] \longrightarrow 0 \quad \text{as} \quad N \to \infty$$

for every $\epsilon > 0$; and this is equivalent to the $\mathbb{P}$-almost sure Cauchy convergence of $\big\{S_n\big\}_1^\infty$. $\square$

In order to convert the conclusion in Theorem 1.4.2 into a statement about $\big\{\overline{S}_n\big\}_1^\infty$, we will need the following elementary *summability* fact about sequences of real numbers.

Lemma 1.4.7 (**Kronecker**). *Let $\{b_n : n \in \mathbb{Z}^+\}$ be a nondecreasing sequence of positive numbers which tends to ∞, and set $\beta_n = b_n - b_{n-1}$, where $b_0 \equiv 0$. If $\{s_n\}_1^\infty \subseteq \mathbb{R}$ is a sequence which converges to $s \in \mathbb{R}$, then*

$$\frac{1}{b_n}\sum_{\ell=1}^n \beta_\ell s_\ell \longrightarrow s.$$

In particular, if $\{x_n\}_1^\infty \subseteq \mathbb{R}$, then

$$\sum_{n=1}^\infty \frac{x_n}{b_n} \text{ converges in } \mathbb{R} \implies \frac{1}{b_n}\sum_{\ell=1}^n x_\ell \longrightarrow 0 \text{ as } n \to \infty.$$

Proof: To prove the first part, assume that $s = 0$, and for given $\epsilon > 0$ choose $N \in \mathbb{Z}^+$ so that $|s_\ell| < \epsilon$ for $\ell \geq N$. Then, with $M = \sup_{n\geq 1}|s_n|$,

$$\left|\frac{1}{b_n}\sum_{\ell=1}^n \beta_\ell s_\ell\right| \leq \frac{M b_N}{b_n} + \epsilon \longrightarrow \epsilon$$

as $n \to \infty$.

Turning to the second part, set $y_\ell = \frac{x_\ell}{b_\ell}$, $s_0 = 0$, and $s_n = \sum_{\ell=1}^n y_\ell$. After summation by parts,

$$\frac{1}{b_n}\sum_{\ell=1}^n x_\ell = s_n - \frac{1}{b_n}\sum_{\ell=1}^n \beta_\ell s_{\ell-1};$$

and so, since $s_n \longrightarrow s \in \mathbb{R}$ as $n \to \infty$, the first part gives the desired conclusion. □

After combining Theorem 1.4.2 with Lemma 1.4.7, we arrive at the following interesting statement.

Corollary 1.4.8. *Assume that $\{b_n\}_1^\infty \subseteq (0,\infty)$ increases to infinity as $n \to \infty$, and suppose that $\{X_n\}_1^\infty$ is a sequence of independent, square $\mathbb{P}$-integrable random variables. If*

$$\sum_{n=1}^\infty \frac{\operatorname{var}(X_n)}{b_n^2} < \infty,$$

then

$$\frac{1}{b_n}\sum_{\ell=1}^n \Big(X_\ell - \mathbb{E}^{\mathbb{P}}\big[X_\ell\big]\Big) \longrightarrow 0 \quad P\text{-almost surely.}$$

As an immediate consequence of the preceding, we see that $\overline{S}_n \longrightarrow m$ $\mathbb{P}$-almost surely if the X_n's are identically distributed and square $\mathbb{P}$-integrable. In fact, without very much additional effort, we can also prove the following much more significant refinement of the last part of Theorem 1.3.1.

Theorem 1.4.9 (**Kolmogorov's Strong Law**). *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of $\mathbb{P}$-independent, identically distributed random variables. If X_1 is $\mathbb{P}$-integrable and has mean-value m, then, as $n \to \infty$, $\overline{S}_n \longrightarrow m$ $\mathbb{P}$-almost surely and in $L^1(P)$. Conversely, if $\overline{S}_n$ converges (in $\mathbb{R}$) on a set of positive $\mathbb{P}$-measure, then X_1 is $\mathbb{P}$-integrable.*

Proof: Assume that X_1 is $\mathbb{P}$-integrable and that $\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0$. Next, set $Y_n = X_n \mathbf{1}_{[0,n]}\big(|X_n|\big)$, and note that

$$\begin{aligned}\sum_{n=1}^{\infty} P\big(Y_n \neq X_n\big) &= \sum_{n=1}^{\infty} P\big(|X_n| > n\big)\\ &\leq \sum_{n=1}^{\infty} \int_{n-1}^{n} P\big(|X_1| > t\big)\, dt = \mathbb{E}^{\mathbb{P}}\big[|X_1|\big] < \infty.\end{aligned}$$

Thus, by the first part of the Borel–Cantelli Lemma,

$$\mathbb{P}\Big(\big(\exists n \in \mathbb{Z}^+\big)\big(\forall N \geq n\big) Y_N = X_N\Big) = 1.$$

In particular, if $\overline{T}_n = \frac{1}{n}\sum_{\ell=1}^{n} Y_\ell$ for $n \in \mathbb{Z}^+$, then, for $\mathbb{P}$-almost every $\omega \in \Omega$, $\overline{T}_n(\omega) \longrightarrow 0$ if and only if $\overline{S}_n(\omega) \longrightarrow 0$. Finally, to see that $\overline{T}_n \longrightarrow 0$ $\mathbb{P}$-almost surely, first observe that, because $\mathbb{E}^{\mathbb{P}}[X_1] = 0$, by the first part of Lemma 1.4.7,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\big[Y_\ell\big] = \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\Big[X_1,\ \big|X_1\big| \leq n\Big] = 0,$$

and therefore, by Corollary 1.4.8, it suffices for us to check that

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}^{\mathbb{P}}\big[Y_n^2\big]}{n^2} < \infty.$$

To this end, set

$$C = \sup_{\ell \in \mathbb{Z}^+} \ell \sum_{n=\ell}^{\infty} \frac{1}{n^2},$$

and note that

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{\mathbb{E}^{\mathbb{P}}\big[Y_n^2\big]}{n^2} &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\Big[X_1^2,\ \ell - 1 < \big|X_1\big| \leq \ell\Big]\\ &= \sum_{\ell=1}^{\infty} \mathbb{E}^{\mathbb{P}}\Big[X_1^2,\ \ell - 1 < \big|X_1\big| \leq \ell\Big] \sum_{n=\ell}^{\infty} \frac{1}{n^2}\\ &\leq C \sum_{\ell=1}^{\infty} \frac{1}{\ell} \mathbb{E}^{\mathbb{P}}\Big[X_1^2,\ \ell - 1 < \big|X_1\big| \leq \ell\Big] \leq C\, \mathbb{E}^{\mathbb{P}}\big[|X_1|\big] < \infty.\end{aligned}$$

Thus, the $\mathbb{P}$-almost sure convergence is now established, and the $L^1(P)$-convergence result was proved already in Theorem 1.2.7.

Turning to the converse assertion, first note that (by Lemma 1.4.1) if $\overline{S}_n$ converges in $\mathbb{R}$ on a set of positive $\mathbb{P}$-measure, then it converges $\mathbb{P}$-almost surely to some $m \in \mathbb{R}$. In particular,

$$\varlimsup_{n\to\infty} \frac{|X_n|}{n} = \varlimsup_{n\to\infty} \left|\overline{S}_n - \overline{S}_{n-1}\right| = 0 \quad P\text{-almost surely;}$$

and so, if $A_n \equiv \{|X_n| > n\}$, then $\mathbb{P}\left(\varlimsup_{n\to\infty} A_n\right) = 0$. But the A_n's are mutually independent; and therefore, by the second part of the Borel–Cantelli Lemma, we now know that $\sum_{n=1}^{\infty} P\left(A_n\right) < \infty$. Hence,

$$\mathbb{E}^{\mathbb{P}}\left[|X_1|\right] = \int_0^\infty P\left(|X_1| > t\right) dt \le 1 + \sum_{n=1}^{\infty} P\left(|X_n| > n\right) < \infty. \quad \square$$

Remark 1.4.10. A reason for being interested in the converse part of Theorem 1.4.9 is that it provides a reconciliation between the measure theory vs. frequency schools of probability theory.

Although Theorem 1.4.9 is the centerpiece of this section, we still want to give another approach to the study of the almost sure convergence properties of $\{S_n\}_1^\infty$. In fact, following P. Lévy, we are going to show that $\{S_n\}_1^\infty$ *converges* $\mathbb{P}$-*almost surely if it converges in* $\mathbb{P}$-measure. Hence, for example, Theorem 1.4.2 can be proved as a direct consequence of (1.4.4), without appeal to Kolmogorov's Inequality.

The key to Lévy's analysis lies in a version of the *reflection principle*, whose statement requires the introduction of a new concept. Given an $\mathbb{R}$-valued random variable Y, we say that $\alpha \in \mathbb{R}$ is a **median** of Y and write $\alpha \in \operatorname{med}(Y)$, if

$$\mathbb{P}\left(Y \le \alpha\right) \wedge P\left(Y \ge \alpha\right) \ge \frac{1}{2}. \tag{1.4.11}$$

Notice that (as distinguished from a *mean-value*) every Y admits a median; for example, it is easy to check that

$$\alpha \equiv \inf\left\{t \in \mathbb{R} :\, P\left(Y \le t\right) \ge \frac{1}{2}\right\}$$

is a median of Y. In addition, it is clear that

$$\operatorname{med}(\beta + Y) = \beta + \operatorname{med}(Y) \quad \text{for all } \beta \in \mathbb{R}.$$

On the other hand, the notion of median is flawed by the fact that, in general, a random variable will admit an entire nondegenerate interval of medians. In addition, it is neither easy to compute the medians of a sum in terms of the medians of the summands nor to relate the *medians* of an integrable random variable to its *mean-value*. Nonetheless, at least if $Y \in L^p(P)$ for some $p \in [1,\infty)$, the following estimate provides some information. Namely, since, for $\alpha \in \operatorname{med}(Y)$ and $\beta \in \mathbb{R}$,

$$\frac{|\alpha-\beta|^p}{2} \le |\alpha-\beta|^p P\big(Y \ge \alpha\big) \wedge P\big(Y \le \alpha\big) \le \mathbb{E}^{\mathbb{P}}\big[|Y-\beta|^p\big],$$

we see that, for any $p \in [1,\infty)$ and $Y \in L^p(P)$,

$$|\alpha-\beta| \le \Big(2\mathbb{E}^{\mathbb{P}}\big[|Y-\beta|^p\big]\Big)^{\frac{1}{p}} \text{ for all } \beta \in \mathbb{R} \quad \text{and} \quad \alpha \in \operatorname{med}(Y).$$

In particular, if $Y \in L^2(P)$ and m is the mean-value of Y, then

$$|\alpha - m| \le \sqrt{2\operatorname{var}(Y)}. \tag{1.4.12}$$

Theorem 1.4.13 (**Lévy's Reflection Principle**). *Let $\{X_n : n \in \mathbb{Z}^+\}$ be a sequence of $\mathbb{P}$-independent random variables, and, for $k \le \ell$, choose $\alpha_{\ell,k} \in \operatorname{med}\big(S_\ell - S_k\big)$. Then, for any $N \in \mathbb{Z}^+$ and $\epsilon > 0$,*

$$\mathbb{P}\left(\max_{1\le n\le N}\big(S_n + \alpha_{N,n}\big) \ge \epsilon\right) \le 2P\big(S_N \ge \epsilon\big); \tag{1.4.14}$$

and therefore

$$\mathbb{P}\left(\max_{1\le n\le N}\big|S_n + \alpha_{N,n}\big| \ge \epsilon\right) \le 2P\big(|S_N| \ge \epsilon\big).$$

Proof: Clearly 1.4.13 follows by applying (1.4.14) to both the sequences $\{X_n\}_1^\infty$ and $\{-X_n\}_1^\infty$ and then adding the two results.

To prove (1.4.14), set $A_1 = \{S_1 + \alpha_{N,1} \ge \epsilon\}$ and

$$A_{n+1} = \left\{\max_{1\le \ell\le n}\big(S_\ell + \alpha_{N,\ell}\big) < \epsilon \text{ and } S_{n+1} + \alpha_{N,n+1} \ge \epsilon\right\}$$

for $1 \le n < N$. Obviously, the A_n's are mutually disjoint and

$$\bigcup_{n=1}^N A_n = \left\{\max_{1\le n\le N}\big(S_n + \alpha_{N,n}\big) \ge \epsilon\right\}.$$

In addition,

$$\{S_N \ge \epsilon\} \supseteq A_n \cap \{S_N - S_n \ge \alpha_{N,n}\} \quad \text{for each } 1 \le n \le N.$$

Hence,

$$P\big(S_N \ge \epsilon\big) \ge \sum_{n=1}^{N} P\Big(A_n \cap \big\{S_N - S_n \ge \alpha_{N,n}\big\}\Big)$$
$$\ge \frac{1}{2}\sum_{n=1}^{N} P\big(A_n\big) = \frac{1}{2}P\left(\max_{1\le n\le N}\big(S_n + \alpha_{N,n}\big) \ge \epsilon\right),$$

where, in the passage to the last line, we have used the independence of the sets A_n and $\big\{S_N - S_n \ge \alpha_{N,n}\big\}$. $\square$

Corollary 1.4.15. *Let $\big\{X_n : n \in \mathbb{Z}^+\big\}$ be a sequence of independent random variables, and assume that $\big\{S_n : n \in \mathbb{Z}^+\big\}$ converges in $\mathbb{P}$-measure to an $\mathbb{R}$-valued random variable S. Then $S_n \longrightarrow S$ $\mathbb{P}$-almost surely. (Cf. Exercise 1.4.24 as well.)*

Proof: What we must show is that, for each $\epsilon > 0$, there is an $M \in \mathbb{Z}^+$ such that

$$\sup_{N\ge 1} P\Big(\max_{1\le n\le N}\big|S_{n+M} - S_M\big| \ge \epsilon\Big) < \epsilon.$$

To this end, let $0 < \epsilon < 1$ be given, and choose $M \in \mathbb{Z}^+$ so that

$$\mathbb{P}\left(\big|S_{n+M} - S_{k+M}\big| \ge \frac{\epsilon}{2}\right) < \frac{\epsilon}{2} \quad \text{for all } 1 \le k < n.$$

Next, for a given $N \in \mathbb{Z}^+$, choose $\alpha_{N,n} \in \operatorname{med}\big(S_{M+N} - S_{M+n}\big)$ for $0 \le n \le N$. Then $|\alpha_{N,n}| \le \frac{\epsilon}{2}$, and so, by 1.4.13 applied to $\{X_{M+n}\}_{n=1}^{\infty}$,

$$P\left(\max_{1\le n\le N}\big|S_{M+n} - S_M\big| \ge \epsilon\right) \le P\left(\max_{1\le n\le N}\big|S_{M+n} - S_M + \alpha_{N,n}\big| \ge \frac{\epsilon}{2}\right)$$
$$\le 2P\Big(\big|S_{M+N} - S_M\big| \ge \frac{\epsilon}{2}\Big) < \epsilon. \quad \square$$

Remark 1.4.16. The most beautiful and startling feature of Lévy's line of reasoning is that it requires *no* integrability assumptions. Of course, in many applications of Corollary 1.4.15, integrability considerations enter into the proof that $\big\{S_n\big\}_1^{\infty}$ converges in $\mathbb{P}$-measure. Finally, a word of caution may be in order. Namely, the result in Corollary 1.4.15 applies to the quantities S_n themselves; it does *not* apply to associated quantities like $\overline{S}_n$! Indeed, suppose that $\{X_n\}_1^{\infty}$ is a sequence of independent random variables with common distribution satisfying

$$\mathbb{P}\big(X_n \le -t\big) = P\big(X_n \ge t\big) = \Big(\big(1+t^2\big)\,\log\big(e^4 + t^2\big)\Big)^{-\frac{1}{2}} \quad \text{for all } t \ge 0.$$

On the one hand, by Exercise 1.2.12, we know that the associated averages $\overline{S}_n$ tend to 0 in probability. On the other hand, by the second part of Theorem 1.4.9, we know that the sequence $\big\{\overline{S}_n\big\}_1^{\infty}$ diverges almost surely.

Exercises for § 1.4

EXERCISE 1.4.17. Let X and Y be nonnegative random variables, and suppose that

$$\mathbb{P}\big(X \geq t\big) \leq \frac{1}{t}\,\mathbb{E}^{\mathbb{P}}\Big[Y,\, X \geq t\Big], \quad t \in (0,\infty). \tag{1.4.18}$$

Show that

$$\Big(\mathbb{E}^{\mathbb{P}}\big[X^p\big]\Big)^{\frac{1}{p}} \leq \frac{p}{p-1}\Big(\mathbb{E}^{\mathbb{P}}\big[Y^p\big]\Big)^{\frac{1}{p}}, \quad p \in (1,\infty). \tag{1.4.19}$$

Hint: First, reduce to the case when X is bounded. Next, recall that, for any measure space $\big(E,\mathcal{F},\mu\big)$, any nonnegative, measurable f on $\big(E,\mathcal{F}\big)$, and any $\alpha \in (0,\infty)$,

$$\int_E f(x)^\alpha\,\mu(dx) = \alpha\int_{(0,\infty)} t^{\alpha-1}\,\mu\big(f > t\big)\,dt.$$

Use this together with (1.4.18) to justify the relation

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[X^p\big] &\leq p\int_{(0,\infty)} t^{p-2}\,\mathbb{E}^{\mathbb{P}}\Big[Y,\, X \geq t\Big]\\ &= p\mathbb{E}^{\mathbb{P}}\left[X\int_0^X t^{p-2}\,dt\right] = \frac{p}{p-1}\,\mathbb{E}^{\mathbb{P}}\Big[X^{p-1}\,Y\Big];\end{aligned}$$

and arrive at (1.4.19) after an application of Hölder's inequality.

EXERCISE 1.4.20. Let $\big\{X_n\big\}_1^\infty$ be a sequence of mutually independent, square $\mathbb{P}$-integrable random variables with mean value 0, and assume that $\sum_1^\infty E\big[X_n^2\big] < \infty$. Let S denote the random variable (guaranteed by Theorem 1.4.2) to which $\big\{S_n\big\}_1^\infty$ converges $\mathbb{P}$-almost surely, and, using elementary orthogonality considerations, check that $S_n \longrightarrow S$ in $L^2(P)$ as well. Next, after examining the proof of Kolmogorov's inequality (cf. (1.4.6)), show that

$$\mathbb{P}\left(\sup_{n\in\mathbb{Z}^+}\big|S_n\big|^2 \geq t\right) \leq \frac{1}{t}\mathbb{E}^{\mathbb{P}}\left[S^2,\ \sup_{n\in\mathbb{Z}^+}\big|S_n\big|^2 \geq t\right], \quad t > 0.$$

Finally, by applying (1.4.19), show that

$$\mathbb{E}^{\mathbb{P}}\left[\sup_{n\in\mathbb{Z}^+}\big|S_n\big|^{2p}\right] \leq \left(\tfrac{p}{p-1}\right)^p\mathbb{E}^{\mathbb{P}}\Big[\big|S\big|^{2p}\Big], \quad p \in (1,\infty); \tag{1.4.21}$$

and conclude from this that, for each $p \in (2,\infty)$, $\big\{S_n\big\}_1^\infty$ converges to S in $L^p(P)$ if and only if $S \in L^p(P)$.

EXERCISE 1.4.22. If $X \in L^2(P)$, then it is easy to characterize its mean m as the $c \in \mathbb{R}$ which minimizes $\mathbb{E}^{\mathbb{P}}\big[(X-c)^2\big]$. Assuming that $X \in L^1(P)$, show that $\alpha \in \mathrm{med}(X)$ if and only if

$$\mathbb{E}^{\mathbb{P}}\big[|X-\alpha|\big] = \min_{c\in\mathbb{R}} \mathbb{E}^{\mathbb{P}}\big[|X-c|\big].$$

Hint: Show that, for any a, $b \in \mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}}\big[|X-b|\big] - \mathbb{E}^{\mathbb{P}}\big[|X-a|\big] = \int_a^b \big[P(X \le t) - P(X \ge t)\big]\,dt.$$

EXERCISE 1.4.23. Let $\{X_n : n \ge 1\}$ be a sequence of random variables which converges in probability to the random variable X, and assume that $\sup_{n\ge 1} \mathrm{var}(X_n) < \infty$. Show that X is square integrable and that $\mathbb{E}^{\mathbb{P}}\big[|X_n - X|\big] \longrightarrow 0$. In particular, if, in addition, $\mathrm{var}(X_n) \longrightarrow \mathrm{var}(X)$, the $\mathbb{E}^{\mathbb{P}}\big[|X_n - X|^2\big] \longrightarrow 0$.

Hint: Let $\alpha_n \in \mathrm{med}(X_n)$, and show that $\alpha_+ = \overline{\lim}_{n\to\infty} \alpha_n$ and $\alpha_- = \underline{\lim}_{n\to\infty} \alpha_n$ are both elements of $\mathrm{med}(X)$. Combine this with (1.4.12) to conclude that $\sup_{n\ge 1}\big|\mathbb{E}^{\mathbb{P}}[X_n]\big| < \infty$ and therefore that $\sup_{n\ge 1} \mathbb{E}^{\mathbb{P}}[X^2] < \infty$.

EXERCISE 1.4.24. The following variant of Theorem 1.4.13 is sometimes useful and has the advantage that it avoids the introduction of medians. Namely show that for any $t \in (0,\infty)$ and $n \in \mathbb{Z}^+$:

$$\mathbb{P}\Big(\max_{1\le m\le n} |S_n| \ge 2t\Big) \le \frac{P\big(|S_n| > t\big)}{1 - \max\limits_{1\le m\le n} P\big(|S_n - S_m| > t\big)}.$$

Note that this can be used in place of 1.4.13 when proving results like the one in Corollary 1.4.15.

EXERCISE 1.4.25. A random variable X is said to by **symmetric** if $-X$ has the same distribution as X itself. Obviously, the most natural choice of median for a symmetric random variable is 0; and thus, because sums of independent, symmetric random variables are again symmetric, (1.4.14) and 1.4.13 are particularly interesting when the X_n's are symmetric, since the $\alpha_{\ell,k}$'s can then be taken to be 0. In this connection, we present the following interesting variation on the theme of Theorem 1.4.13.

(**i**) Let $X_1, \dots, X_n, \dots$ be independent, symmetric random variables, set $M_n(\omega) = \max_{1\le \ell\le n} |X_\ell(\omega)|$, let $\tau_n(\omega)$ be the smallest $1 \le \ell \le n$ with the property that $\big|X_\ell(\omega)\big| = M_n(\omega)$, and define

$$Y_n(\omega) = X_{\tau_n(\omega)}(\omega) \quad \text{and} \quad \hat{S}_n = S_n - Y_n.$$

Show that

$$\omega \in \Omega \longmapsto \big(\hat{S}_n(\omega), Y_n(\omega)\big) \in \mathbb{R}^2 \quad \text{and} \quad \omega \in \Omega \longmapsto \big(-\hat{S}_n(\omega), Y_n(\omega)\big) \in \mathbb{R}^2$$

have the same distribution, and conclude first that

$$\begin{aligned} P\big(Y_n \geq t\big) &\leq P\Big(Y_n \geq t \;\&\; \hat{S}_n \geq 0\Big) + P\Big(Y_n \geq t \;\&\; \hat{S}_n \leq 0\Big) \\ &= 2P\Big(Y_n \geq t \;\&\; \hat{S}_n \geq 0\Big) \leq 2P\big(S_n \geq t\big), \end{aligned}$$

for all $t \in \mathbb{R}$; and then that

$$\mathbb{P}\left(\max_{1 \leq \ell \leq n} \big|X_\ell\big| \geq t\right) \leq 2P\Big(\big|S_n\big| \geq t\Big), \quad t \in [0, \infty).$$

(**ii**) Continuing in the same setting, add the assumption that the X_n's are identically distributed, and use part (**ii**) to show that

$$\begin{aligned} &\lim_{n \to \infty} P\Big(\big|\overline{S}_n\big| \leq C\Big) = 1 \quad \text{for some } C \in (0, \infty) \\ &\qquad \Longrightarrow \lim_{n \to \infty} nP\Big(\big|X_1\big| \geq n\Big) = 0. \end{aligned}$$

Hint: Note that

$$\mathbb{P}\big(\max_{1 \leq \ell \leq n} |X_\ell| \geq t\big) = 1 - \mathbb{P}(|X_1 \geq t)^n$$

and that $\frac{1-(1-x)^n}{x} \longrightarrow n$ as $x \searrow 0$.

In conjunction with Exercise 1.2.12, this proves that *if* $\{X_n\}_1^\infty$ *is a sequence of independent, identically distributed symmetric random variables, then* $\overline{S}_n \longrightarrow 0$ *in* $\mathbb{P}$*-probability if and only if* $\lim_{n\to\infty} nP\big(|X_1| \geq n\big) = 0$.

Exercise 1.4.26. Let $X_1, \dots, X_n, \dots$ be a sequence of mutually independent, identically distributed, $\mathbb{P}$-integrable random variables with mean-value m. As we already know, when $m > 0$, the partial sums S_n tend, $\mathbb{P}$-almost surely, to $+\infty$ at an asymptotic linear rate m; and, of course, when $m < 0$ the situation is similar at $-\infty$. Moreover, when $m = 0$, we know that, if $|S_n|$ tends to ∞ at all, then, $\mathbb{P}$-almost surely, it does so at a strictly sublinear rate. In this exercise, we will sharpen this statement by proving that

$$m = 0 \implies \varliminf_{n \to \infty} |S_n| < \infty \quad P\text{-almost surely.}$$

The beautiful argument given below is due to Y. Guivarc'h, but it's full power cannot be appreciated in the present context (cf. Exercise (6.2.3?)). Indeed, a classic result (cf. Exercise (5.2.11?) below) due to K.L. Chung and W.H. Fuchs shows that $\varliminf_{n\to\infty} |S_n| = 0$ $\mathbb{P}$-almost surely.

In order to prove the assertion here, assume that $\lim_{n\to\infty} |S_n| = \infty$ with positive $\mathbb{P}$-probability, use Kolmogorov's 0–1 Law to see that $|S_n| \longrightarrow \infty$ $\mathbb{P}$-almost surely, and proceed as follows.

(**i**) Show that there must exist an $\epsilon > 0$ with the property that

$$\mathbb{P}\Big(\forall \ell > k \ \big|S_\ell - S_k\big| \geq \epsilon\Big) \geq \epsilon$$

for some $k \in \mathbb{Z}^+$ and therefore that

$$\mathbb{P}(A) \geq \epsilon \quad \text{where } A \equiv \Big\{\omega : \forall \ell \in \mathbb{Z}^+ \ \big|S_\ell(\omega)\big| \geq \epsilon\Big\}.$$

(**ii**) For each $\omega \in \Omega$ and $n \in \mathbb{Z}^+$, set

$$\Gamma_n(\omega) = \Big\{t \in \mathbb{R} : \exists 1 \leq \ell \leq n \ \big|t - S_\ell(\omega)\big| < \tfrac{\epsilon}{2}\Big\}$$

and

$$\Gamma_n'(\omega) = \Big\{t \in \mathbb{R} : \exists 1 \leq \ell \leq n \ \big|t - S_\ell'(\omega)\big| < \tfrac{\epsilon}{2}\Big\},$$

where $S_n' \equiv \sum_{\ell=1}^n X_{\ell+1}$. Next, let $R_n(\omega)$ and $R_n'(\omega)$ denote the Lebesgue measure of $\Gamma_n(\omega)$ and $\Gamma_n'(\omega)$, respectively; and, using the translation invariance of Lebesgue's measure, show that

$$R_{n+1}(\omega) - R_n'(\omega) \geq \epsilon \mathbf{1}_{A'}(\omega),$$
$$\text{where } A' \equiv \Big\{\omega : \forall \ell \geq 2 \ \big|S_\ell(\omega) - S_1(\omega)\big| \geq \epsilon\Big\}.$$

On the other hand, show that

$$\mathbb{E}^{\mathbb{P}}\big[R_n'\big] = \mathbb{E}^{\mathbb{P}}\big[R_n\big] \quad \text{and} \quad P(A') = P(A);$$

and conclude first that

$$\epsilon P(A) \leq \mathbb{E}^{\mathbb{P}}\big[R_{n+1} - R_n\big], \quad n \in \mathbb{Z}^+,$$

and then that

$$\epsilon P(A) \leq \varliminf_{n\to\infty} \frac{1}{n}\mathbb{E}^{\mathbb{P}}\big[R_n\big].$$

(**iii**) In view of parts (**i**) and (**ii**), we will be done once we show that

$$m = 0 \implies \lim_{n\to\infty} \frac{1}{n}\mathbb{E}^{\mathbb{P}}\big[R_n\big] = 0.$$

But clearly, $0 \leq R_n(\omega) \leq n\epsilon$. Thus, it is enough for us to show that, when $m = 0$, $\frac{R_n}{n} \longrightarrow 0$ $\mathbb{P}$-almost surely; and, to this end, first check that

$$\frac{S_n(\omega)}{n} \longrightarrow 0 \implies \frac{R_n(\omega)}{n} \longrightarrow 0,$$

and, finally, apply The Strong Law of Large Numbers.

EXERCISE 1.4.27. As we have already said, for many applications the Weak Law of Large Numbers is just as good as and even preferable to the Strong Law. Nonetheless, here is an application in which the full strength of Strong Law plays an essential role. Namely, we are going to use the Strong Law to produce examples of continuous, strictly increasing functions F on $[0,1]$ with the property that their derivative

$$F'(t) \equiv \lim_{y \to x} \frac{F(y) - F(x)}{y - x} = 0 \quad \text{at Lebesgue almost every } x \in (0,1).$$

By familiar facts about functions of a real variable, one knows that such functions F are in one-to-one correspondence with non-atomic, Borel probability measures μ on $[0,1]$ which charge every non-empty open subset but are singular to Lebesgue's measure. Namely, F is the distribution function determined by μ: $F(x) = \mu\big((-\infty, x]\big)$.

(**i**) Set $\Omega = \{0,1\}^{\mathbb{Z}^+}$, and, for each $p \in (0,1)$, take $M_p = (\beta_p)^{\mathbb{Z}^+}$ where β_p on $\{0,1\}$ is the Bernoulli measure with $\beta_p(\{1\}) = p = 1 - \beta_p(\{0\})$. Next, define

$$\omega \in \Omega \longmapsto Y(\omega) \equiv \sum_{n=1}^{\infty} 2^{-n}\omega_n \in [0,1],$$

and let μ_p denote the M_p-distribution of Y. Given $n \in \mathbb{Z}^+$ and $0 \le m < 2^n$, show that

$$\mu_p\big([m2^{-n}, (m+1)2^{-n}]\big) = p^{\ell_{m,n}}(1-p)^{n-\ell_{m,n}},$$

where $\ell_{m,n} = \sum_{k=1}^n \omega_k$ and $(\omega_1, \dots, \omega_n) \in \{0,1\}^n$ is determined by $m2^{-n} = \sum_{k=1}^n 2^{-k}\omega_k$. Conclude, in particular, that μ_p is non-atomic and charges every non-empty open subset of $[0,1]$.

(**iii**) Given $x \in [0,1)$ and $n \in \mathbb{Z}^+$, define

$$\epsilon_n(x) = \begin{cases} 1 & \text{if } 2^{n-1}x - [2^{n-1}x] \ge \frac{1}{2} \\ 0 & \text{if } 2^{n-1}x - [2^{n-1}x] < \frac{1}{2}, \end{cases}$$

where $[s]$ denotes the integer part of s. If $\{\epsilon_n : n \ge 1\} \subseteq \{0,1\}$ satisfies $x = \sum_1^\infty 2^{-m}\epsilon_m$ show that $\epsilon_m = \epsilon_m(x)$ for all $m \ge 1$ if and only if $\epsilon_m = 0$ for infinitely many $m \ge 1$. In particular, conclude first that $\omega_n = \epsilon_n\big(Y(\omega)\big)$, $n \in \mathbb{Z}^+$, for M_p-almost every $\omega \in \Omega$ and second, by the Strong Law, that

$$\frac{1}{n}\sum_{m=1}^{n} \epsilon_n(x) \longrightarrow p \quad \text{for } \mu_p\text{-almost every } x \in [0,1].$$

Thus, $\mu_{p_1} \perp \mu_{p_2}$ whenever $p_1 \ne p_2$.

(**iv**) By Lemma 1.1.6, we know that $\mu_{\frac{1}{2}}$ is Lebesgue measure $\lambda_{[0,1]}$ on $[0,1]$. Hence, we now know that $\mu_p \perp \lambda_{[0,1]}$ when $p \neq \frac{1}{2}$. In view of the introductory remarks, this completes the proof that, for each $p \in (0,1) \setminus \{\frac{1}{2}\}$, the function $F_p(x) = \mu_p\big((-\infty, x]\big)$ is a strictly increasing, continuous function on $[0,1]$ whose derivative vanishes at Lebesgue almost every point. Here, we can do better. Namely, referring to part (**iii**), let Δ_p denote the set of $x \in [0,1)$ such that

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_n(x) = p \quad \text{where } \Sigma_n(x) \equiv \sum_{m=1}^{n} \epsilon_m(x).$$

We know that $\Delta_{\frac{1}{2}}$ has Lebesgue measure 1. Show that, for each $x \in \Delta_{\frac{1}{2}}$ and $p \in (0,1) \setminus \{\frac{1}{2}\}$, F_p is differentiable with derivative 0 at x.

Hint: Given $x \in [0,1)$, define

$$L_n(x) = \sum_{m=1}^{n} 2^{-m}\epsilon_m(x) \quad \text{and} \quad R_n(x) = L_n(x) + 2^{-n}.$$

Show that

$$F_p\big(R_n(x)\big) - F_p\big(L_n(x)\big) = M_p\left(\sum_{m=1}^{n} 2^{-m}\omega_m = L_n(x)\right) = p^{\Sigma_n(x)}(1-p)^{n-\Sigma_n(x)}.$$

When $p \in (0,1) \setminus \{\frac{1}{2}\}$ and $x \in \Delta_{\frac{1}{2}}$, use this together with $4p(1-p) < 1$ to show that

$$\lim_{n\to\infty} n \log\left(\frac{F_p\big(R_n(x)\big) - F_p\big(L_n(x)\big)}{R_n(x) - L_n(x)}\right) < 0.$$

To complete the proof, for given $x \in \Delta_{\frac{1}{2}}$ and $n \geq 2$ such that $\Sigma_n(x) \geq 2$, let $M_n(x)$ denote the largest $m < n$ such that $\epsilon_m(x) = 1$, and show that $\frac{M_n(x)}{n} \longrightarrow 1$ as $n \to \infty$. Hence, since $2^{-n-1} < h \leq 2^{-n}$ implies that

$$\frac{F_p(x) - F_p(x-h)}{h} \leq 2^{n-M_n(x)+1}\,\frac{F_p(\big(R_n(x)\big) - F_p(\big(L_n(x)\big)}{R_n(x) - L_n(x)},$$

one concludes that F_p is left-differentiable at x and has left derivative equal to 0 there. To get the same conclusion about right derivatives, simply note that $F_p(x) = 1 - F_{1-p}(1-x)$.

(**v**) Again let $p \in (0,1) \setminus \{\frac{1}{2}\}$ be given, but this time choose $x \in \Delta_p$. Show that

$$\lim_{h\searrow 0} \frac{F_p(x+h) - F_p(x)}{h} = +\infty.$$

The argument is similar to the one used to handle part (**iv**). However, this time the role played by the inequality $4pq < 1$ is played here by $(2p)^p(2q)^q > 1$ when $q = 1-p$.

§1.5 Law of the Iterated Logarithm

Let $X_1, \dots, X_n, \dots$ be a sequence of independent, identically distributed random variables with mean-value 0 and variance 1. In this section, we will investigate exactly how large $\{S_n : n \in \mathbb{Z}^+\}$ can become as $n \to \infty$. To get a feeling for what we should be expecting, first note that, by Corollary 1.4.8, for any nondecreasing $\{b_n\}_1^\infty \subseteq (0,\infty)$,

$$\frac{S_n}{b_n} \longrightarrow 0 \quad P\text{-almost surely if} \quad \sum_{n=1}^{\infty} \frac{1}{b_n^2} < \infty.$$

Thus, for example, S_n grows more slowly than $n^{\frac{1}{2}} \log n$. On the other hand, if the X_n's are $\mathcal{N}(0,1)$-random variables, then so are the random variables $\frac{S_n}{\sqrt{n}}$; and therefore, for every $R \in (0,\infty)$,

$$\begin{aligned} P\left(\overline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{n}} \geq R\right) &= \lim_{N\to\infty} P\left(\bigcup_{n\geq N} \left\{\frac{S_n}{\sqrt{n}} \geq R\right\}\right) \\ &\geq \lim_{N\to\infty} P\left(\frac{S_N}{\sqrt{N}} \geq R\right) > 0. \end{aligned}$$

Hence, at least for normal random variables, we can use Lemma 1.4.1 to see that

$$\overline{\lim_{n\to\infty}} \frac{S_n}{\sqrt{n}} = \infty \quad P\text{-almost surely;}$$

and so S_n grows faster than $n^{\frac{1}{2}}$.

If, as we did in Section 1.3, we proceed on the assumption that Gaussian random variables are typical, we should expect the growth rate of the S_n's to be something between $n^{\frac{1}{2}}$ and $n^{\frac{1}{2}} \log n$. What, in fact, turns out to be the precise growth rate is

$$\Lambda_n \equiv \sqrt{2n \log_{(2)}(n \vee 3)} \tag{1.5.1}$$

where $\log_{(2)} x \equiv \log(\log x)$ (*not* the logarithm with base 2) for $x \in [e,\infty)$. That is, one has the **Law of the Iterated Logarithm**:

$$\overline{\lim_{n\to\infty}} \frac{S_n}{\Lambda_n} = 1 \quad P\text{-almost surely.} \tag{1.5.2}$$

This remarkable fact was discovered first for Bernoulli random variables by Khinchine, was extended by Kolmogorov to random variables possessing $2+\epsilon$ moments, and eventually achieved its final form in the work of Hartman and Wintner. The approach which we will adopt here is based on ideas (taught to the author by M. Ledoux) introduced originally to handle generalizations of (1.5.2) to random variables with values in a Banach space.* This approach consists of two steps. The first establishes a preliminary version of (1.5.2) which, although it is far cruder than (1.5.2) itself, will allow us to justify a reduction of the general case to the case of bounded random variables. In the second step, we deal with bounded random variables and more or less follow Khinchine's strategy for deriving (1.5.2) once one has estimates like the ones provided by Theorem 1.3.12.

In what follows, we will use $[\beta] \equiv \max\{n \in \mathbb{Z} : n \leq \beta\}$ to denote the integer part of $\beta \in \mathbb{R}$ and will define

$$\Lambda_\beta = \Lambda_{[\beta]} \quad \text{and} \quad \tilde{S}_\beta = \frac{S_{[\beta]}}{\Lambda_\beta} \quad \text{for} \quad \beta \in [3, \infty).$$

Lemma 1.5.3. *Let $\{X_n\}_1^\infty$ be a sequence of independent, identically distributed random variables with mean-value* 0 *and variance* 1*. Then, for any $a \in (0, \infty)$ and $\beta \in (1, \infty)$,*

$$\varlimsup_{n\to\infty} \left|\tilde{S}_n\right| \leq a \quad \text{(a.s., } P) \quad \textit{if} \quad \sum_{m=1}^{\infty} P\Big(\left|\tilde{S}_{\beta^m}\right| \geq a\,\beta^{-\frac{1}{2}}\Big) < \infty.$$

Proof: Let $\beta \in (1, \infty)$ be given and, for each $m \in \mathbb{N}$ and $1 \leq n \leq \beta^m$, let $\alpha_{m,n}$ be a median (cf. (1.4.11)) of $S_{[\beta^m]} - S_n$. Noting that, by (1.4.12), $\left|\alpha_{m,n}\right| \leq \sqrt{2\beta^m}$, we see that

$$\begin{aligned}
\varlimsup_{n\to\infty} \left|\tilde{S}_n\right| &= \varlimsup_{m\to\infty} \max_{\beta^{m-1}\leq n \leq \beta^m} \left|\tilde{S}_n\right| \\
&\leq \beta^{\frac{1}{2}} \varlimsup_{m\to\infty} \max_{\beta^{m-1}\leq n \leq \beta^m} \frac{\left|S_n\right|}{\Lambda_{\beta^m}} \\
&\leq \beta^{\frac{1}{2}} \varlimsup_{m\to\infty} \max_{n\leq\beta^m} \frac{\left|S_n + \alpha_{m,n}\right|}{\Lambda_{\beta^m}};
\end{aligned}$$

and therefore,

$$\mathbb{P}\Big(\varlimsup_{n\to\infty} \left|\tilde{S}_n\right| \geq a\Big) \leq P\left(\varlimsup_{m\to\infty} \max_{n\leq\beta^m} \frac{\left|S_n + \alpha_{m,n}\right|}{\Lambda_{\beta^m}} \geq a\,\beta^{-\frac{1}{2}}\right).$$

But, by Theorem 1.4.13,

$$P\left(\max_{n\leq\beta^m} \frac{\left|S_n + \alpha_{m,n}\right|}{\Lambda_{\beta^m}} \geq a\beta^{-\frac{1}{2}}\right) \leq 2P\left(\left|\tilde{S}_{\beta^m}\right| \geq a\beta^{-\frac{1}{2}}\right),$$

and so the desired result follows from the Borel–Cantelli Lemma. □

---

* See M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer–Verlag Ergebnisse Series 3.Folge·Band 23 (1991).

LEMMA 1.5.4. *For any sequence $\{X_n\}_1^\infty$ of independent, identically distributed random variables with mean-value* 0 *and variance* σ^2,

$$\overline{\lim_{n\to\infty}} \left|\tilde{S}_n\right| \leq 8\sigma \quad (\text{a.}s., \mathbb{P}). \tag{1.5.5}$$

PROOF: Without loss in generality, we assume throughout that $\sigma = 1$; and, for the moment, we will also assume that the X_n's are symmetric (cf. Exercise 1.4.24). By Lemma 1.5.3, we will know that (1.5.5) holds with 8 replaced by 4 once we show that

$$\sum_{m=0}^{\infty} P\left(\left|\tilde{S}_{2^m}\right| \geq 2^{\frac{3}{2}}\right) < \infty. \tag{*}$$

In order to take maximal advantage of symmetry, let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which the X_n's are defined, use $\{R_n\}_1^\infty$ to denote the sequence of Rademacher functions on $[0,1)$ introduced in Section 1.1, and set $Q = \lambda_{[0,1)} \times P$ on $\big([0,1) \times \Omega, \mathcal{B}_{[0,1)} \times \mathcal{F}\big)$. It is then an easy matter to check that symmetry of the X_n's is equivalent to the statement that

$$\omega \in \Omega \longrightarrow \big(X_1(\omega), \dots, X_n(\omega), \dots\big) \in \mathbb{R}^{\mathbb{Z}^+}$$

has the same distribution under $\mathbb{P}$ as

$$(t, \omega) \in [0,1) \times \Omega \longmapsto \big(R_1(t)X_1(\omega), \dots, R_n(t)X_n(\omega), \dots\big) \in \mathbb{R}^{\mathbb{Z}^+}$$

does under Q. Next, using the last part of (**iii**) in Exercise 1.3.18 with $\sigma_k = X_k(\omega)$, note that:

$$\begin{aligned} \lambda_{[0,1)} \left(\left\{ t \in [0,1) : \left| \sum_{n=1}^{2^m} R_n(t) X_n(\omega) \right| \geq a \right\} \right) \\ \leq 2 \exp\left[-\frac{a^2}{2\sum_{n=1}^{2^m} X_n(\omega)^2} \right], \quad a \in [0,\infty) \text{ and } \omega \in \Omega. \end{aligned}$$

Hence, if

$$A_m \equiv \left\{ \omega \in \Omega : \frac{1}{2^m} \sum_{n=1}^{2^m} X_m(\omega)^2 \geq 2 \right\}$$

and

$$F_m(\omega) \equiv \lambda_{[0,1)} \left(\left\{ t \in [0,1) : \left| \sum_{n=1}^{2^m} R_n(t) X_n(\omega) \right| \geq 2^{\frac{3}{2}} \Lambda_{2^m} \right\} \right),$$

then, by Tonelli's Theorem,

$$P\Big(\Big\{\omega\in\Omega:\,\big|S_{2^m}(\omega)\big|\ge 2^{\frac{3}{2}}\Lambda_{2^m}\Big\}\Big)=\int_\Omega F_m(\omega)\,P(d\omega)$$

$$\le 2\int_\Omega \exp\left[-\frac{8\Lambda_{2^m}^2}{2\sum_{n=1}^{2^m}X_n(\omega)^2}\right]P(d\omega)\le 2\exp\Big[-4\log_{(2)}2^m\Big]+2P\big(A_m\big).$$

Thus, (*) comes down to proving that $\sum_{m=0}^\infty P\big(A_m\big)<\infty$, and in order to check this we argue in much the same way as we did when we proved the converse statement in Kolmogorov's Strong Law. Namely, set

$$T_m=\sum_{n=1}^{2^m}X_n^2,\quad B_m=\left\{\frac{T_{m+1}-T_m}{2^m}\ge 2\right\},\quad\text{and}\quad \overline{T}_m=\frac{T_m}{2^m}$$

for $m\in\mathbb{N}$. Clearly, $\mathbb{P}\big(A_m\big)=P\big(B_m\big)$. Moreover, the sets B_m, $m\in\mathbb{N}$, are mutually independent; and therefore, by the Borel–Cantelli Lemma, we need only check that

$$\mathbb{P}\left(\overline{\lim_{m\to\infty}}B_m\right)=P\left(\overline{\lim_{m\to\infty}}\frac{T_{m+1}-T_m}{2^m}\ge 2\right)=0.$$

But, by The Strong Law, we know that $\overline{T}_m\longrightarrow 1$ (a.s., $\mathbb{P}$), and therefore it is clear that

$$\frac{T_{m+1}-T_m}{2^m}\longrightarrow 1\quad\text{(a.s., }\mathbb{P}).$$

We have now proved (1.5.5) with 4 replacing 8 for symmetric random variables. To eliminate the symmetry assumption, again let $(\Omega,\mathcal{F},\mathbb{P})$ be the probability space on which the X_n's are defined, let $\big(\Omega',\mathcal{F}',P'\big)$ be a second copy of the same space, and consider the random variables

$$(\omega,\omega')\in\Omega\times\Omega'\longmapsto Y_n\big(\omega,\omega'\big)\equiv\frac{X_n(\omega)-X_n(\omega')}{\sqrt{2}}$$

under the measure $Q\equiv P\times P'$. Since the Y_n's are obviously symmetric, the result which we have already proved says that

$$\overline{\lim_{n\to\infty}}\frac{\big|S_n(\omega)-S_n(\omega')\big|}{\Lambda_n}\le 2^{\frac{5}{2}}\le 8\quad\text{for }Q\text{-almost every }(\omega,\omega')\in\Omega\times\Omega'.$$

Now suppose that $\overline{\lim}_{n\to\infty}\frac{|S_n|}{\Lambda_n}>8$ on a set of positive $\mathbb{P}$-measure. Then, by Kolmogorov's 0–1 Law, there would exist an $\epsilon>0$ such that

$$\overline{\lim_{n\to\infty}}\frac{|S_n(\omega)|}{\Lambda_n}\ge 8+\epsilon\quad\text{for }\mathbb{P}\text{-almost every }\omega\in\Omega;$$

and so, by Fubini's Theorem,* we would have that, for Q-almost every $(\omega,\omega') \in \Omega\times\Omega'$, there is a $\{n_m(\omega) : m\in\mathbb{Z}^+\}\subseteq\mathbb{Z}^+$ such that $n_m(\omega)\nearrow\infty$ and

$$\begin{aligned}&\varliminf_{m\to\infty}\frac{\left|S_{n_m(\omega)}(\omega')\right|}{\Lambda_{n_m(\omega)}}\\ &\qquad\geq \varliminf_{m\to\infty}\frac{\left|S_{n_m(\omega)}(\omega)\right|}{\Lambda_{n_m(\omega)}} - \varlimsup_{m\to\infty}\frac{\left|S_{n_m(\omega)}(\omega)-S_{n_m(\omega)}(\omega')\right|}{\Lambda_{n_m(\omega)}}\geq\epsilon.\end{aligned}$$

But, again by Fubini's Theorem, this would mean that there exists a $\{n_m : m\in\mathbb{Z}^+\}\subseteq\mathbb{Z}^+$ such that $n_m\nearrow\infty$ and $\varliminf_{m\to\infty}\frac{\left|S_{n_m}(\omega')\right|}{\Lambda_{n_m}}\geq\epsilon$ for $\mathbb{P}'$-almost every $\omega'\in\Omega'$; and obviously this contradicts

$$\mathbb{E}^{P'}\left[\left(\frac{S_n}{\Lambda_n}\right)^2\right]=\frac{1}{2\log_{(2)}n}\longrightarrow 0.\quad\square$$

We have now got the *crude statement* alluded to above. In order to get the more precise statement contained in (1.5.2), we will need the following application of the results in Section 3.

LEMMA 1.5.6. *Let $\{X_n\}_1^\infty$ be a sequence of independent random variables with mean-value 0, variance 1, and common distribution μ. Further, assume that (1.3.4) holds. Then, for each $R\in(0,\infty)$ there is an $N(R)\in\mathbb{Z}^+$ such that*

$$\mathbb{P}\Big(\big|\tilde{S}_n\big|\geq R\Big)\leq 2\exp\left[-\left(1-K\sqrt{\frac{8R\log_{(2)}n}{n}}\right)R^2\log_{(2)}n\right]\tag{1.5.7}$$

for $n\geq N(R)$. In addition, for each $\epsilon\in(0,1]$, there is an $N(\epsilon)\in\mathbb{Z}^+$ such that, for all $n\geq N(\epsilon)$ and $|a|\leq\frac{1}{\epsilon}$,

$$\mathbb{P}\Big(\big|\tilde{S}_n-a\big|<\epsilon\Big)\geq\frac{1}{2}\exp\Big[-\Big(a^2+4K|a|\epsilon\Big)\log_{(2)}n\Big].\tag{1.5.8}$$

In both (1.5.7) and (1.5.8), the constant $K\in(0,\infty)$ is the one in Theorem 1.3.15.

PROOF: Set

$$\lambda_n=\frac{\Lambda_n}{n}=\left(\frac{2\log_{(2)}(n\vee 3)}{n}\right)^{\frac{1}{2}}.$$

* This is Fubini at his best and subtlest. Namely, we are using Fubini to switch between *horizontal* and *vertical* sets of measure 0.

To prove (1.5.7), simply apply the upper bound in the last part of Theorem 1.3.15 to see that, for sufficiently large $n \in \mathbb{Z}^+$,

$$P\Big(\big|\tilde{S}_n\big| \geq R\Big) = P\Big(\big|\overline{S}_n\big| \geq R\lambda_n\Big) \\ \leq 2\exp\left[-n\left(\frac{(R\lambda_n)^2}{2} - K\big(R\lambda_n\big)^3\right)\right].$$

To prove (1.5.8), first note that

$$\mathbb{P}\Big(\big|\tilde{S}_n - a\big| < \epsilon\Big) = P\Big(\big|\overline{S}_n - a_n\big| < \epsilon_n\Big),$$

where $a_n = a\lambda_n$ and $\epsilon_n = \epsilon\lambda_n$. Thus, by the lower bound in the last part of Theorem 1.3.15,

$$P\Big(\big|\tilde{S}_n - a\big| < \epsilon\Big) \\ \geq \left(1 - \frac{K}{n\epsilon_n^2}\right)\exp\left[-n\left(\frac{a_n^2}{2} + K|a_n|\big(\epsilon_n + a_n^2\big)\right)\right] \\ \geq \left(1 - \frac{K}{2\epsilon^2\log_{(2)} n}\right)\exp\left[-\Big(a^2 + 2K|a|\big(\epsilon + a^2\lambda_n\big)\Big)\log_{(2)} n\right]$$

for sufficiently large n's. $\square$

Theorem 1.5.9 (**Law of Iterated Logarithm**). *The equation* (1.5.2) *holds for any sequence* $\{X_n\}_1^\infty$ *of independent, identically distributed random variables with mean-value* 0 *and variance* 1. *In fact,* $\mathbb{P}$*-almost surely, the set of limit points of* $\left\{\frac{S_n}{\Lambda_n}\right\}_1^\infty$ *coincides with the entire interval* $[-1,1]$. *Equivalently, for any* $f \in C\big(\mathbb{R};\mathbb{R}\big)$,

$$\text{(1.5.10)} \qquad \overline{\lim_{n\to\infty}} f\left(\frac{S_n}{\Lambda_n}\right) = \sup_{t\in[-1,1]} f(t) \quad (\text{a.s.}, \mathbb{P}).$$

(Cf. Exercise 1.5.13 below for a converse statement.)

Proof: We begin with the observation that, because of (1.5.5), we may restrict our attention to the case when the X_n's are bounded random variables. Indeed, for any X_n's and any $\epsilon > 0$, an easy truncation procedure allows us to find an $\psi \in C_{\mathrm{b}}(\mathbb{R};\mathbb{R})$ such that $Y_n \equiv \psi \circ X_n$ again has mean-value 0 and variance 1 while $Z_n \equiv X_n - Y_n$ has variance less than ϵ^2. Hence, if the result is known when the random variables are bounded, then, by (1.5.5) applied to the Z_n's:

$$\overline{\lim_{n\to\infty}} \big|\tilde{S}_n(\omega)\big| \leq 1 + \overline{\lim_{n\to\infty}} \left|\frac{\sum_{m=1}^n Z_m(\omega)}{\Lambda_n}\right| \leq 1 + 8\epsilon,$$

and, for $a \in [-1,1]$,

$$\varliminf_{n\to\infty} \left|\tilde{S}_n(\omega) - a\right| \le \varlimsup_{n\to\infty} \left|\frac{\sum_{m=1}^n Z_m(\omega)}{\Lambda_n}\right| \le 8\epsilon$$

for $\mathbb{P}$-almost every $\omega \in \Omega$.

In view of the preceding, from now on we may and will assume that the X_n's are bounded. To prove that $\varlimsup_{n\to\infty} \tilde{S}_n \le 1$ (a.s., $\mathbb{P}$), let $\beta \in (1,\infty)$ be given and use (1.5.7) to see that

$$\mathbb{P}\Big(\left|\tilde{S}_{\beta^m}\right| \ge \beta^{\frac{1}{2}}\Big) \le 2\exp\Big[-\beta^{\frac{1}{2}}\log_{(2)}\big[\beta^m\big]\Big]$$

for all sufficiently large $m \in \mathbb{Z}^+$. Hence, by Lemma 1.5.3 with $a = \beta$, we see that $\varlimsup_{n\to\infty}\left|\tilde{S}_n\right| \le \beta$ (a.s., $\mathbb{P}$) for every $\beta \in (1,\infty)$. To complete the proof, we must still show that, for every $a \in (-1,1)$ and $\epsilon > 0$,

$$\mathbb{P}\Big(\varliminf_{n\to\infty}\left|\tilde{S}_n - a\right| < \epsilon\Big) = 1.$$

Because we want to get this conclusion as an application of the second part of the Borel–Cantelli Lemma, it is important that we be dealing with independent events; and for this purpose, we use the result just proved to see that, for every integer $k \ge 2$,

$$\begin{aligned}\varliminf_{n\to\infty}\left|\tilde{S}_n - a\right| &\le \varlimsup_{k\to\infty}\varliminf_{m\to\infty}\left|\tilde{S}_{k^m} - a\right| \\ &= \varlimsup_{k\to\infty}\varliminf_{m\to\infty}\left|\frac{S_{k^m} - S_{k^{m-1}}}{\Lambda_{k^m}} - a\right| \quad P\text{-almost surely.}\end{aligned}$$

Thus, because the events

$$A_{k,m} \equiv \left\{\left|\frac{S_{k^m} - S_{k^{m-1}}}{\Lambda_{k^m}} - a\right| < \epsilon\right\}, \quad m \in \mathbb{Z}^+,$$

are independent for each $k \ge 2$, all that we need to do is check that

$$\sum_{m=1}^{\infty} P\big(A_{k,m}\big) = \infty \quad \text{for sufficiently large } k \ge 2.$$

But

$$\mathbb{P}\big(A_{k,m}\big) = P\left(\left|\tilde{S}_{k^m - k^{m-1}} - \frac{\Lambda_{k^m} a}{\Lambda_{k^m - k^{m-1}}}\right| < \frac{\Lambda_{k^m}\epsilon}{\Lambda_{k^m - k^{m-1}}}\right),$$

and, because

$$\lim_{k\to\infty}\max_{m\in\mathbb{Z}^+}\left|\frac{\Lambda_{k^m}}{\Lambda_{k^m - k^{m-1}}} - 1\right| = 0,$$

everything reduces to showing that

$$(*)\qquad \sum_{m=1}^{\infty} P\Big(\big|\tilde{S}_{k^m-k^{m-1}} - a\big| < \epsilon\Big) = \infty$$

for each $k \geq 2$, $a \in (-1, 1)$, and $\epsilon > 0$. Finally, referring to (1.5.8), choose $\epsilon_0 > 0$ so small that $\rho \equiv a^2 + 4K\epsilon_0|a| < 1$, and conclude that, when $0 < \epsilon < \epsilon_0$,

$$\mathbb{P}\Big(\big|\tilde{S}_n - a\big| < \epsilon\Big) \geq \frac{1}{2}\exp\Big[-\rho \log_{(2)} n\Big]$$

for all sufficiently large n's; from which (*) is easy. □

REMARK 1.5.11. The reader should notice that the Law of the Iterated Logarithm provides a *naturally occurring* sequence of functions which converge in measure but not almost everywhere. Indeed, it is obvious that $\tilde{S}_n \longrightarrow 0$ in $L^2(P)$, but the Law of the Iterated Logarithm says that $\big\{\tilde{S}_n\big\}_1^\infty$ is wildly divergent when looked at in terms of $\mathbb{P}$-almost sure convergence.

Exercises for § 1.5

EXERCISE 1.5.12. Let X and X' be a pair of independent random variables which have the same distribution, let α be a median of X, and set $Y = X - X'$.

(**i**) Show that Y is symmetric and that

$$\mathbb{P}\Big(\big|X - \alpha\big| \geq t\Big) \leq 2P\Big(\big|Y\big| \geq t\Big) \quad \text{for all} \quad t \in [0, \infty),$$

and conclude that, for any $p \in (0, \infty)$,

$$2^{-\frac{1}{p}\vee 1}\mathbb{E}^{\mathbb{P}}\big[Y^p\big]^{\frac{1}{p}} \leq \mathbb{E}^{\mathbb{P}}\big[X^p\big]^{\frac{1}{p}} \leq 2^{(\frac{1}{p}-1)^+}\left(2\mathbb{E}^{\mathbb{P}}\big[Y^p\big]^{\frac{1}{p}} + |\alpha|\right).$$

In particular, $|X^p$ is integrable if and only if $|Y|^p$ is.

(**ii**) As an initial application of (**i**), we give our final refinement of The Weak Law of Large Numbers. Namely, let $\big\{X_n\big\}_1^\infty$ be a sequence of independent, identically distributed random variables. By combining Exercise 1.2.12, part (**ii**) in Exercise 1.4.25, and part (**i**) above, show that*

$$\begin{aligned} &\lim_{n\to\infty} P\Big(\big|\overline{S}_n\big| \leq C\Big) = 1 \quad \text{for some } C \in (0, \infty) \\ &\qquad\Longrightarrow \lim_{n\to\infty} nP\big(|X_1| \geq n\big) = 0 \\ &\Longrightarrow \overline{S}_n - \mathbb{E}^{\mathbb{P}}\big[X_1,\, |X_1| \leq n\big] \longrightarrow 0 \text{ in } \mathbb{P}\text{-probability.} \end{aligned}$$

---

* These ideas are taken from the book by Wm. Feller cited at the end of §1.2. They become even more elegant when combined with a theorem due to E.J.G. Pitman (cf. *ibid.*).

EXERCISE 1.5.13. Let $X_1, \dots, X_n, \dots$ be a sequence of independent, identically distributed random variables for which

$$\mathbb{P}\left(\varlimsup_{n\to\infty} \frac{|S_n|}{\Lambda_n} < \infty\right) > 0. \tag{1.5.14}$$

In this exercise we will show* that X_1 is square $\mathbb{P}$-integrable, $\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0$, and

$$\varlimsup_{n\to\infty} \frac{S_n}{\Lambda_n} = -\varliminf_{n\to\infty} \frac{S_n}{\Lambda_n} = \mathbb{E}^{\mathbb{P}}\big[X_1^2\big]^{\frac{1}{2}} \quad (\text{a.s.}, \mathbb{P}). \tag{1.5.15}$$

(**i**) Using Lemma 1.4.1, show that there is a $\sigma \in [0, \infty)$ such that

$$\varlimsup_{n\to\infty} \frac{\big|S_n\big|}{\Lambda_n} = \sigma \quad (\text{a.s.}, \mathbb{P}). \tag{1.5.16}$$

Next, assuming that X_1 is square $\mathbb{P}$-integrable, use The Strong Law of Large Numbers together with Theorem 1.5.9 to show that $\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0$ and

$$\sigma = \mathbb{E}^{\mathbb{P}}\big[X_1^2\big]^{\frac{1}{2}} = \varlimsup_{n\to\infty} \frac{S_n}{\Lambda_n} = -\varliminf_{n\to\infty} \frac{S_n}{\Lambda_n} \quad (\text{a.s.}, \mathbb{P}).$$

In other words, everything comes down to proving that (1.5.14) implies that X_1 is square $\mathbb{P}$-integrable.

(**ii**) Assume that the X_n's are symmetric. For $t \in (0, \infty)$, set

$$\check{X}_n^t = X_n\,\mathbf{1}_{[0,t]}\big(|X_n|\big) - X_n\,\mathbf{1}_{(t,\infty)}\big(|X_n|\big),$$

and show that

$$\big(\check{X}_1^t, \dots, \check{X}_n^t, \dots\big) \quad \text{and} \quad \big(X_1, \dots, X_n, \dots\big)$$

have the same distribution. Conclude first that, for all $t \in [0, 1)$,

$$\varlimsup_{n\to\infty} \frac{\big|\sum_{m=1}^n X_n\,\mathbf{1}_{[0,t]}\big(|X_n|\big)\big|}{\Lambda_n} \le \sigma \quad (\text{a.s.}, \mathbb{P}),$$

where σ is the number in (1.5.16), and second that

$$\mathbb{E}^{\mathbb{P}}\big[X_1^2\big] = \lim_{t\nearrow\infty} \mathbb{E}^{\mathbb{P}}\Big[X_1^2,\, \big|X_1\big| \le t\Big] \le \sigma^2.$$

Hint: Use the equation

$$X_n\,\mathbf{1}_{[0,t]}\big(|X_n|\big) = \frac{X_n + \check{X}_n^t}{2},$$

and apply part (**i**).

---

* We follow Wm. Feller "An extension of the law of the iterated logarithm," *J. Math. Mech.* **18**, although V. Strassen was the first to prove the result.

(**iii**) For general $\{X_n\}_1^\infty$, produce an independent copy $\{X_n'\}_1^\infty$ (as in the proof of Lemma 1.5.4), and set $Y_n = X_n - X_n'$. After checking that

$$\overline{\lim_{n\to\infty}} \frac{\left|\sum_{m=1}^n Y_m\right|}{\Lambda_n} \le 2\sigma \quad (\text{a.s.}, \mathbb{P}),$$

conclude first that $\mathbb{E}^{\mathbb{P}}\big[Y_1^2\big] \le 4\sigma^2$ and then (cf. part (**i**) of Exercise1.5.12) that $\mathbb{E}^{\mathbb{P}}\big[X_1^2\big] < \infty$. Finally, apply (**i**) to arrive at $\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0$ and (1.5.15).

Exercise 1.5.17. Let $\{\tilde{s}_n\}_1^\infty$ be a sequence of real numbers which possess the properties that

$$\overline{\lim_{n\to\infty}} \tilde{s}_n = 1, \quad \underline{\lim_{n\to\infty}} \tilde{s}_n = -1, \quad \text{and} \quad \lim_{n\to\infty} \big|\tilde{s}_{n+1} - \tilde{s}_n\big| = 0.$$

Show that the set of sub-sequential limit points of $\{\tilde{s}_n\}_1^\infty$ coincides with $[-1,1]$. Apply this observation to show that in order to get the final statement in Theorem 1.5.9, we need only have proved (1.5.10) for the function $f(x) = x$, $x \in \mathbb{R}$.

Hint: In proving the last part, use the square integrability of X_1 to see that

$$\sum_{n=1}^\infty P\left(\frac{X_n^2}{n} \ge 1\right) < \infty,$$

and apply the Borel–Cantelli Lemma to conclude that $\tilde{S}_n - \tilde{S}_{n-1} \longrightarrow 0$ (a.s., $\mathbb{P}$).