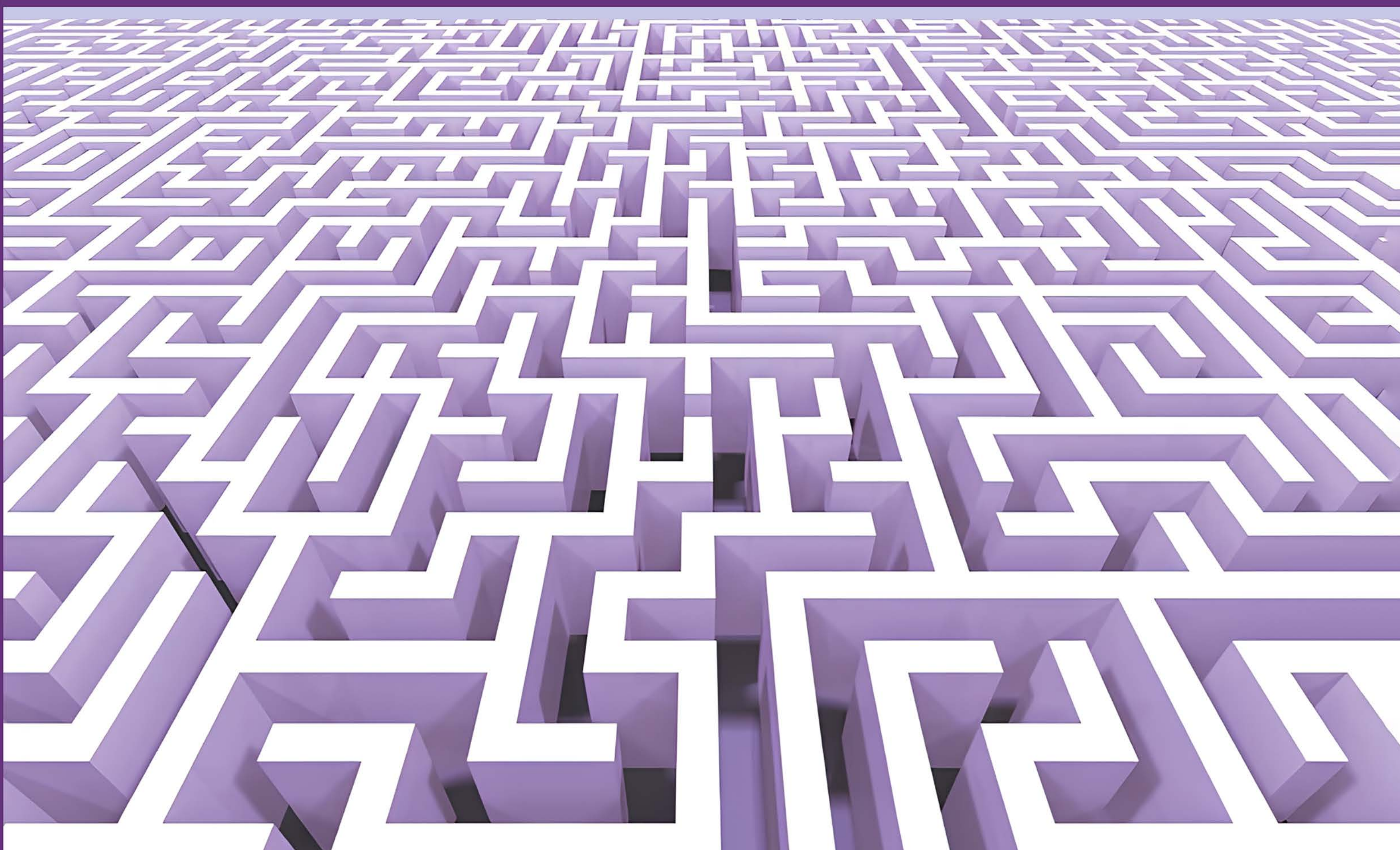
PROBABILITY THEORY

AN ANALYTIC VIEW

THIRD EDITION



DANIEL W. STROOCK

CAMBRIDGE

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 there is no relationship at all! For this reason, I will begin by looking at families of *mutually independent* random variables.[1]

1.1 Independence

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

1.1.1 Mutually Independent σ-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 nonnegative 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 $\mathcal{F}$. I will 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 $A_{i_m} \in \mathcal{F}_{i_m}$, $1 \le m \le n$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_n}). \tag{1.1}$$

In particular, if $\{A_i : i \in \mathcal{I}\}$ is a family of sets from $\mathcal{F}$, I will 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 $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$, the classical definition that 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 nonmeasure theoretic presentations, and therefore the intuition that underlies the elementary notion applies equally well to the definition given here. (See Exercise 1.1.1 for more information about the connection between the present definition and the classical one.)

[1] Most authors truncate "mutually independent" to "independent." However, as my advisor Mark Kac pointed out, unless one is talking about random variables that are constant, the random variables are not independent of themselves but of each other. That is, they are *mutually* independent.

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

$$\mathcal{F}_\Lambda = \bigvee_{i\in\Lambda} \mathcal{F}_i \equiv \sigma\left(\bigcup_{i\in\Lambda} \mathcal{F}_i\right)$$

be the σ-algebra generated by $\bigcup_{i\in\Lambda} \mathcal{F}_i$ (i.e., $\mathcal{F}_\Lambda$ is the smallest σ-algebra containing $\bigcup_{i\in\Lambda} \mathcal{F}_i$). Next, define the **tail σ-algebra** $\mathcal{T}$ to be the intersection over all finite $\Lambda \subseteq \mathcal{I}$ of the σ-algebras $\mathcal{F}_{\mathcal{I}\setminus\Lambda}$. 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 $\mathcal{F}_i$ are $\mathbb{P}$-independent. To see this, for a given nonempty $\Lambda \subseteq \mathcal{I}$, let C_Λ denote the collection of sets of the form $A_{i_1} \cap \cdots \cap A_{i_n}$, where $\{i_1, \ldots, i_n\}$ are distinct elements of Λ and $A_{i_m} \in \mathcal{F}_{i_m}$ for each $1 \le m \le n$. Clearly C_Λ is closed under intersection and $\mathcal{F}_\Lambda = \sigma(C_\Lambda)$. In addition, by assumption, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in C_\Lambda$ and $B \in C_{\mathcal{I}\setminus\Lambda}$. Hence, by Exercise 1.1.4, $\mathcal{F}_\Lambda$ is independent of $\mathcal{F}_{\mathcal{I}\setminus\Lambda}$. But this means that $\mathcal{T}$ is independent of $\mathcal{F}_F$ for every finite $F \subseteq \mathcal{I}$, and therefore, again by Exercise 1.1.4, $\mathcal{T}$ is independent of

$$\mathcal{F} = \sigma\left(\bigcup\{\mathcal{F}_F : F \text{ a finite subset of } \Lambda\}\right).$$

Since $\mathcal{T} \subseteq \mathcal{F}$, this implies 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 I have now proved the following famous result.

Theorem 1.1.1 (Kolmogorov's 0–1 Law) *Let $\{\mathcal{F}_i : i \in \mathcal{I}\}$ be a family of mutually $\mathbb{P}$-independent sub-σ-algebras of $(\Omega, \mathcal{F}, \mathbb{P})$ and define the tail σ-algebra $\mathcal{T}$ accordingly, as discussed earlier. Then, for every $A \in \mathcal{T}$, $\mathbb{P}(A)$ is either* 0 *or* 1.

To develop a feeling for the kind of conclusions that can be drawn from Kolmogorov's 0–1 Law (cf. Exercises 1.1.10 and 1.1.11 as well), let $\{A_n : n \ge 1\}$ be a sequence of subsets of Ω, and recall the notion of the limit superior of sets

$$\varlimsup_{n\to\infty} A_n \equiv \bigcap_{m=1}^{\infty} \bigcup_{n\ge m} A_n = \left\{\omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{Z}^+\right\}.$$

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 are mutually $\mathbb{P}$-independent elements of $\mathcal{F}$, then

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

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

Lemma 1.1.2 (Borel–Cantelli Lemma) *Let $\{A_n : n \in \mathbb{Z}^+\} \subseteq \mathcal{F}$ be given. Then*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\left(\overline{\lim_{n\to\infty}} A_n\right) = 0. \tag{1.2}$$

In fact, if the A_n are $\mathbb{P}$-independent sets, then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \iff \mathbb{P}\left(\overline{\lim_{n\to\infty}} A_n\right) = 1. \tag{1.3}$$

(See part (**iii**) *of Exercise 5.2.3 and Lemma 11.4.6 for generalizations.)*

Proof The first assertion, which is due to E. Borel, 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 complete the proof of (1.3) when the A_n are mutually independent, note that, by countable additivity, $\mathbb{P}\left(\overline{\lim}_{n\to\infty} A_n\right) = 1$ if and only if

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

But, by independence and another application of countable additivity, for any given $m \geq 1$, we have that

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

if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. (In the preceding equation, I have used the trivial inequality $1 - t \leq e^{-t}$, $t \in [0, \infty)$.) This part is due to F. Cantelli. □

A second, and perhaps more transparent, way of dealing with the contents of the preceding result is to introduce the nonnegative random variable $N(\omega) \in \mathbb{Z}^+ \cup \{\infty\}$, which counts the number of $n \in \mathbb{Z}^+$ such that $\omega \in A_n$. Then, by Tonelli's Theorem,[2] $\mathbb{E}^{\mathbb{P}}[N] = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$, and so Borel's contribution is equivalent to the $\mathbb{E}^{\mathbb{P}}[N] < \infty \implies \mathbb{P}(N < \infty) = 1$, which is obvious, whereas, when combined with Kolmogorov's 0–1 Law, Cantelli's contribution is that, for mutually independent A_n, $\mathbb{P}(N < \infty) > 0 \implies \mathbb{E}^{\mathbb{P}}[N] < \infty$, which is not obvious.

1.1.2 Mutually Independent Functions

Having described what it means for the σ-algebras to be $\mathbb{P}$-independent, I will 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**

[2] Throughout this book, I 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 $A = \Omega$, I will write $\mathbb{E}^{\mathbb{P}}[X]$. Tonelli's Theorem is the version of Fubini's Theorem for nonnegative functions. Its virtue is that it applies whether or not the integrand is integrable.

variable (i.e., a measurable function on $(\Omega, \mathcal{F})$ with values in the measurable space $(E_i, \mathcal{B}_i)$). I will say that the random variables X_i, $i \in \mathcal{I}$, are **mutually** $\mathbb{P}$-**independent** if the σ-algebras

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

are mutually $\mathbb{P}$-independent. If $B(E; \mathbb{R})$ denotes the space of bounded measurable $\mathbb{R}$-valued functions on the measurable space $(E, \mathcal{B})$, then it should be clear that $\mathbb{P}$-independence of $\{X_i : i \in \mathcal{I}\}$ is equivalent to the statement that

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

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

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

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

Thus far I 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 $(E_i, \mathcal{B}_i, \mu_i)$ 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

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

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, I have chosen to defer the details of the product construction to Exercises 1.1.6 and 1.1.7 and to, instead, spend the rest of this section developing a more hands-on approach to constructing sequences of mutually independent, real-valued random variables. Indeed, although the product method is more ubiquitous and has become the construction of choice, the one that I am about to present has the advantage that it shows independent random variables can arise "naturally" and even in a familiar places.

1.1.3 The Rademacher Functions

Until further notice, take $(\Omega, \mathcal{F}) = \big([0,1), \mathcal{B}_{[0,1)}\big)$ (when E is a metric space, I use $\mathcal{B}_E$ to denote the Borel field over E) and $\mathbb{P}$ to be the restriction $\lambda_{[0,1)}$ of Lebesgue measure $\lambda_{\mathbb{R}}$ to $[0,1)$. Next define the **Rademacher functions** R_n, $n \in \mathbb{Z}^+$, on Ω as follows. Take $\lfloor t \rfloor$ for $t \in \mathbb{R}$ to be the **integer part** (i.e., the largest integer dominated by t) and consider the function $R : \mathbb{R} \longrightarrow \{-1, 1\}$ given by

$$R(t) = \begin{cases} -1 & \text{if} \quad t - \lfloor t \rfloor \in \left[0, \tfrac{1}{2}\right), \\ \ \ 1 & \text{if} \quad t - \lfloor t \rfloor \in \left[\tfrac{1}{2}, 1\right). \end{cases}$$

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

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

I will now show that the Rademacher functions are mutually $\mathbb{P}$-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 I 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 $n = 1$, I will assume that it holds for n and need only check 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 $F\colon \{-1, 1\}^n \longrightarrow \mathbb{R}$. 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 $I_{m,n}$, we get the desired result.

At this point I have produced a countably infinite sequence of mutually independent **Bernoulli random variables** (i.e., two-valued random variables whose range is usually either $\{-1, 1\}$ or $\{0, 1\}$) with mean value 0. Of course, for all $a < b$, any sequence of mutually independent, $\{-1, 1\}$-valued Bernoulli random variables can be transformed into $\{a, b\}$-valued ones via a linear transformation.

In order to get more general random variables, I will combine Bernoulli random variables together in a clever way.

Recall that a random variable U is said to be **uniform** 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 1.1.3 *Let* $\{X_\ell : \ell \in \mathbb{Z}^+\}$ *be a sequence of mutually* $\mathbb{P}$*-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 I prove it for a particular realization of independent, mean value $\frac{1}{2}$, $\{0, 1\}$-valued Bernoulli random variables. Thus, by the preceding discussion, I 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 (cf. part (**i**) of Exercise 1.1.8), 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. □

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

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

Clearly, each $Y_{k,\ell}$ is a $\{0, 1\}$-valued, Bernoulli random variable with mean value $\frac{1}{2}$, and the random variables $\left\{Y_{k,\ell} \colon (k, \ell) \in \left(\mathbb{Z}^+\right)^2\right\}$ are mutually $\mathbb{P}$-independent. Hence, by Lemma 1.1.3, 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 are obviously mutually independent. Hence, I have now produced a sequence of mutually independent random variables, each of which is uniformly distributed on $[0, 1)$. To complete our program, I use the time-honored transformation that 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 that 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,

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

(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:

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

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

Theorem 1.1.4 *Let* $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}_{[0,1)}$, *and* $\mathbb{P} = \lambda_{[0,1)}$. *Then, for any sequence* $\{F_k \colon k \in \mathbb{Z}^+\}$ *of distribution functions on* $\mathbb{R}$, *there exists a sequence* $\{X_k \colon k \in \mathbb{Z}^+\}$ *of mutually* $\mathbb{P}$-*independent random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with the property that* $\mathbb{P}\big(X_k \le t\big) = F_k(t)$, $t \in \mathbb{R}$, *for each* $k \in \mathbb{Z}^+$.

1.1.4 Exercises for §1.1

Exercise 1.1.1 As I 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

$$\mathbb{P}\big(A_1 \cap A_2 \cap A_3\big) = \mathbb{P}\big(A_1\big)\mathbb{P}\big(A_2\big)\mathbb{P}\big(A_3\big),$$

show that it is not necessarily true that the three σ-algebras generated by A_1, A_2, and A_3 are mutually $\mathbb{P}$-independent.

Exercise 1.1.2 This exercise deals with three elementary, but important, properties of mutually 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 mutually $\mathbb{P}$-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 $F\colon E_1 \times E_2 \longrightarrow \mathbb{R}$ that is bounded below, use Tonelli'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 mutually $\mathbb{P}$-independent, real-valued random variables. If each of the X_m is $\mathbb{P}$-integrable, show that $X_1 \cdots X_n$ is also $\mathbb{P}$-integrable and that

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

(**iii**) Let $\{X_n\colon n \in \mathbb{Z}^+\}$ be a sequence of mutually independent random variables taking values in some separable metric space E. If $\mathbb{P}(X_n = x) = 0$ for all $x \in E$ and $n \in \mathbb{Z}^+$, show that $\mathbb{P}\big(X_m = X_n \text{ for some } m \neq n\big) = 0$.

Exercise 1.1.3 As an application of Lemma 1.1.3 and part (**ii**) of Exercise 1.1.2, prove the identity

$$\sin z = z \prod_{n=1}^{\infty} \cos\big(2^{-n} z\big) \quad \text{for all } z \in \mathbb{C}.$$

Exercise 1.1.4 Given a nonempty set Ω, recall[3] that a collection $\mathcal{C}$ of subsets of Ω is called a π-**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 $\{A_n\colon n \geq 1\}$ is a nondecreasing sequence of members of $\mathcal{L}$. Finally, Lemma 2.1.12 in [61] says that if $\mathcal{C}$ is a π-system, then the σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ 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 $\mathcal{C}$. Next use this to see that if $\{\mathcal{C}_i\colon i \in \mathcal{I}\}$ is a family of π-systems contained in $\mathcal{F}$ and if (1.1) holds when the A_i are from the $\mathcal{C}_i$, then the σ-algebras $\{\sigma(\mathcal{C}_i)\colon i \in \mathcal{I}\}$ are mutually $\mathbb{P}$-independent.

Exercise 1.1.5 In this exercise I discuss two criteria for determining when random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ are mutually independent.

(**i**) Let $X_1, \dots, X_n$ be bounded, real-valued random variables. Using Weierstrass's Approximation Theorem (cf. §1.2.3), show that the X_m are mutually $\mathbb{P}$-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**) For those who are unfamiliar with Fourier analysis, attempting this exercise should be

[3] See, for example, §3.1 in the author's [61].

postponed until the content of §2.3.1 has been read. Let $\mathbf{X}\colon \Omega \longrightarrow \mathbb{R}^m$ and $\mathbf{Y}\colon \Omega \longrightarrow \mathbb{R}^n$ be random variables. Show that $\mathbf{X}$ and $\mathbf{Y}$ are mutually $\mathbb{P}$-independent if and only if

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\exp\left[\imath\left(\left(\boldsymbol{\alpha},\mathbf{X}\right)_{\mathbb{R}^m}+\left(\boldsymbol{\beta},\mathbf{Y}\right)_{\mathbb{R}^n}\right)\right]\right]\\ &\quad=\mathbb{E}^{\mathbb{P}}\left[\exp\left[\imath\left(\boldsymbol{\alpha},\mathbf{X}\right)_{\mathbb{R}^m}\right]\right]\mathbb{E}^{\mathbb{P}}\left[\exp\left[\imath\left(\boldsymbol{\beta},\mathbf{Y}\right)_{\mathbb{R}^n}\right]\right]\end{aligned}$$

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 mutually independent if

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

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

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

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 1.1.6 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 $\left(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\right)$, 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 $\left(E_1 \times E_2, \mathcal{B}_1 \times \mathcal{B}_2\right)$ such that

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

More generally, for any $n \geq 2$ and measurable spaces $\{(E_i, \mathcal{B}_i)\colon 1 \leq i \leq n\}$, one takes $\prod_1^n \mathcal{B}_i$ to be the σ-algebra over $\prod_1^n E_i$ generated by the sets $\prod_1^n \Gamma_i$, $\Gamma_i \in \mathcal{B}_i$. Since $\prod_1^{n+1} E_i$ and $\prod_1^{n+1} \mathcal{B}_i$ can be identified with $\left(\prod_1^n E_i\right) \times E_{n+1}$ and $\left(\prod_1^n \mathcal{B}_i\right) \times \mathcal{B}_{n+1}$, 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 $\left(\prod_1^n E_i, \prod_1^n \mathcal{B}_i\right)$ such that

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

The purpose of this exercise is to generalize the preceding construction to infinite collections. Thus, let $\mathcal{I}$ be an infinite index set, and, for each $i \in \mathcal{I}$, let $(E_i, \mathcal{B}_i)$ be a measurable space. Given $\emptyset \neq \Lambda \subseteq \mathcal{I}$, 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 $\mathbf{E}_{\mathcal{I}}$ onto $\mathbf{E}_\Lambda$. Further, let $\mathcal{B}_{\mathcal{I}} = \prod_{i\in\mathcal{I}} \mathcal{B}_i$ stand for the σ-algebra over $\mathbf{E}_{\mathcal{I}}$ 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 $\mathcal{I}$ (abbreviated by $\emptyset \neq F \subset\subset \mathcal{I}$). In the following steps, I outline a proof that, for every choice of probability measures μ_i on the $(E_i, \mathcal{B}_i)$, there is a unique probability measure $\prod_{i\in\mathcal{I}} \mu_i$ on $(\mathbf{E}_{\mathcal{I}}, \mathcal{B}_{\mathcal{I}})$ with the property that

$$\prod_{i\in\mathcal{I}} \mu_i \left(\pi_F^{-1} \left(\prod_{i\in F} \Gamma_i \right) \right) = \prod_{i\in F} \mu_i(\Gamma_i), \quad \Gamma_i \in \mathcal{B}_i, \tag{1.4}$$

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

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

is called the **product** over $\mathcal{I}$ 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 $(E^{\mathcal{I}}, \mathcal{B}^{\mathcal{I}}, \mu^{\mathcal{I}})$, and if, in addition, $\mathcal{I} = \{1, \dots, N\}$, one uses $(E^N, \mathcal{B}^N, \mu^N)$.

(**i**) Because (cf. Exercise 1.1.4) two probability measures that 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 $(\mathbf{E}_{\mathcal{I}}, \mathcal{B}_{\mathcal{I}})$ that satisfies the condition in (1.4). Hence, the problem is purely one of existence.

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

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

for all $\emptyset \neq F \subset\subset \mathcal{I}$. Hence, all that one has to do is check that μ admits a σ-additive extension to $\mathcal{B}_{\mathcal{I}}$, and, by a standard extension theorem,[4] this comes down to checking that $\mu(A_n) \searrow 0$ whenever $\{A_n\colon n \geq 1\} \subseteq \mathcal{A}$ and $A_n \searrow \emptyset$. Thus, let $\{A_n\colon n \geq 1\}$ be a nonincreasing sequence from $\mathcal{A}$, and assume that $\mu(A_n) \geq \epsilon$ for some $\epsilon > 0$ and all $n \in \mathbb{Z}^+$. One 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 to assume that $A_n = \pi_{F_n}^{-1}(\Gamma_{F_n})$, where, for each $n \in \mathbb{Z}^+$, $\emptyset \neq F_n \subset\subset \mathcal{I}$ and $\Gamma_{F_n} \in \mathcal{B}_{F_n}$. In addition, show that one may assume that $F_1 = \{i_1\}$ and that $F_n = F_{n-1} \cup \{i_n\}$, $n \geq 2$, where $\{i_n\colon n \geq 1\}$ is a sequence of distinct elements of $\mathcal{I}$. Now, make these assumptions, and show that it suffices to find $a_\ell \in E_{i_\ell}$, $\ell \in \mathbb{Z}^+$, with the property that, 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 $g_{m,n}\colon \mathbf{E}_{F_m} \longrightarrow [0,1]$ 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

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

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

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

[4] For example, Theorem 8.2.6 in my [61].

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

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

In addition, note that

$$\int_{E_{i_1}} g_1\big(x_{i_1}\big)\,\mu_{i_1}\big(dx_{i_1}\big) = \lim_{n\to\infty}\int_{E_{i_1}} g_{1,n}\big(x_{i_1}\big)\,\mu_{i_1}\big(dx_{i_1}\big) = \lim_{n\to\infty}\mu(A_n) \ge \epsilon,$$

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

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

Finally, check that $\{a_m : m \ge 1\}$ is a sequence of the sort for which we were looking at the end of part (**iii**).

Exercise 1.1.7 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 $\Phi_*\mu$ (sometimes denoted by $\mu\circ\Phi^{-1}$) defined on $(E',\mathcal{B}')$ by

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

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

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

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

Exercise 1.1.8 Define $\{\epsilon_n(\omega) : n \ge 1\}$ for $\omega \in [0,1)$ as in the proof of Lemma 1.1.3.

(**i**) Show that $\{\epsilon_n(\omega) : n \ge 1\}$ is the unique sequence $\{\alpha_n : n \ge 1\} \subseteq \{0,1\}^{\mathbb{Z}^+}$ such that $\omega - \sum_{m=1}^n 2^{-m}\alpha_m < 2^{-n}$, and conclude that $\epsilon_1(\omega) = \lfloor 2\omega \rfloor$ and $\epsilon_{n+1}(\omega) = \lfloor 2^{n+1}\omega\rfloor - 2\lfloor 2^n\omega\rfloor$ for $n \ge 1$.

(**ii**) Define $F\colon [0,1) \longrightarrow [0,1)^2$ by

$$F(\omega) = \left(\sum_{n=1}^\infty 2^{-n}\epsilon_{2n-1}(\omega), \sum_{n=1}^\infty 2^{-n}\epsilon_{2n}(\omega)\right),$$

and show that $\lambda_{[0,1)^2} = F_*\lambda_{[0,1)}$. That is, $\lambda_{[0,1)}\big(\{\omega : F(\omega) \in \Gamma\}\big) = \lambda^2_{[0,1)}(\Gamma)$ for all $\Gamma \in \mathcal{B}_{[0,1)^2}$.

(**iii**) Define $G\colon [0,\infty)^2 \longrightarrow [0,1)$ by

$$G\big((\omega_1,\omega_2)\big) = \sum_{n=1}^\infty \frac{2\epsilon_n(\omega_1) + \epsilon_n(\omega_2)}{4^n},$$

and show that $\lambda_{[0,1)} = G_*\lambda_{[0,1)^2}$.
Hint: Set $X_n = 2\epsilon_n(\omega_1) + \epsilon(\omega_2)$, and show that the X_n are mutually independent, $\{0, 1, 2, 3\}$-valued random variable that takes each of its values with probability $\frac{1}{4}$. Then, argue as in the proof of Lemma 1.1.3 that $\sum_{n=1}^{\infty} 4^{-n} X_n$ has distribution $\lambda_{[0,1)}$.

Parts (**ii**) and (**iii**) are special cases of a general principle that says, under very general circumstances, measures can be transformed into one another.

Exercise 1.1.9 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 $\mathbb{P}$-square 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

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

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

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

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

(**ii**) Let $n \in \mathbb{Z}^+$, and suppose that f and g are $\mathbb{R}$-valued, Borel measurable functions on $\mathbb{R}^n$ that are nondecreasing with respect to each coordinate (separately). Show that if $\mathbf{X} = \big(X_1, \ldots, X_n\big)$ 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 $\mathbb{P}$-square 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 1.1.10 A σ-algebra is said to be **countably generated** if it contains a countable collection of sets that generate it. The purpose of this exercise is to 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 probability space and $\{A_n : n \in \mathbb{Z}^+\} \subseteq \mathcal{F}$ a sequence of mutually $\mathbb{P}$-independent sub-subsets of $\mathcal{F}$ with the property that $\alpha \leq \mathbb{P}(A_n) \leq 1 - \alpha$ for some $\alpha \in (0, 1)$. Let $\mathcal{F}_n$ be the sub-σ-algebra generated by A_n. Show that the tail σ-algebra $\mathcal{T}$ determined by $\{\mathcal{F}_n : n \in \mathbb{Z}^+\}$ cannot be countably generated.

Hint: Show that $C \in \mathcal{T}$ 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, $\mathbb{P}(B_\ell) \in \{0, 1\}$ 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 1.1.11 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 1.1.3 and part (**i**) of Exercise 1.1.8, define the transformations $T_n : [0, 1) \longrightarrow [0, 1)$ for $n \in \mathbb{Z}^+$ so that

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

and notice (cf. the proof of Lemma 1.1.3) 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 to 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 Lebesgue 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, I will be studying what happens when one averages mutually 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 that, 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 the discussion, $(\Omega, \mathcal{F}, \mathbb{P})$ will be a probability space on which there is a sequence $\{X_n : n \geq 1\}$ of real-valued random variables. Given $n \in \mathbb{Z}^+$, 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

The first result is a very general one; in fact, it even applies to random variables that are not necessarily independent and do not necessarily have mean 0.

Lemma 1.2.1 *Assume that*

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

Then, for each $\epsilon > 0$*,*

$$\epsilon^2\, \mathbb{P}\Big(\big|\overline{S}_n\big| \geq \epsilon\Big) \leq \mathbb{E}^{\mathbb{P}}\big[\,\overline{S}_n^2\big] = \frac{1}{n^2} \sum_{\ell=1}^{n} \mathbb{E}^{\mathbb{P}}\big[X_\ell^2\big] \quad \textit{for } n \in \mathbb{Z}^+. \tag{1.5}$$

In particular, if

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

then

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

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

Proof To prove the equality in (1.5), note that, by orthogonality,

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

The rest is just an application of **Chebyshev's inequality**, the estimate that 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. □

1.2.2 Mutually Independent Random Variables

Although Lemma 1.2.1 does not require independence, mutually independent random variables provide a ready source of orthogonal functions. To wit, recall that for any $\mathbb{P}$-square integrable random variable X, its **variance** $\mathrm{Var}(X)$ satisfies

$$\mathrm{Var}(X) \equiv \mathbb{E}^{\mathbb{P}}\Big[\big(X - \mathbb{E}^{\mathbb{P}}[X]\big)^2\Big] = \mathbb{E}^{\mathbb{P}}\big[X^2\big] - \big(\mathbb{E}^{\mathbb{P}}[X]\big)^2 \leq \mathbb{E}^{\mathbb{P}}\big[X^2\big].$$