Chapter 15

Probability Spaces and Random Variables

We have already introduced the notion of a probability space, namely a measure space $(X, \mathfrak{F}, \mu)$ with the property that $\mu(X) = 1$. Here we look further at some basic notions and results of probability theory.

First, we give some terminology. A set $S \in \mathfrak{F}$ is called an *event*, and $\mu(S)$ is called the probability of the event, often denoted $P(S)$. The image to have in mind is that one picks a point $x \in X$ at random, and $\mu(S)$ is the probability that $x \in S$. A measurable function $f : X \to \mathbb{R}$ is called a (real) *random variable*. (One also speaks of a measurable map $F : (X, \mathfrak{F}) \to (Y, \mathfrak{G})$ as a Y-valued random variable, given a measurable space $(Y, \mathfrak{G})$.) If f is integrable, one sets

$$E(f) = \int_X f \, d\mu, \tag{15.1}$$

called the *expectation* of f, or the *mean* of f. One defines the *variance* of f as

$$\mathrm{Var}(f) = \int_X |f - a|^2 \, d\mu, \quad a = E(f). \tag{15.2}$$

The random variable f has finite variance if and only if $f \in L^2(X, \mu)$.

A random variable $f : X \to \mathbb{R}$ induces a probability measure ν_f on $\mathbb{R}$, called the probability distribution of f:

$$\nu_f(S) = \mu(f^{-1}(S)), \quad S \in \mathfrak{B}(\mathbb{R}), \tag{15.3}$$

where $\mathfrak{B}(\mathbb{R})$ denotes the σ-algebra of Borel sets in $\mathbb{R}$. It is clear that $f \in L^1(X, \mu)$ if and only if $\int |x|\, d\nu_f(x) < \infty$ and that

$$E(f) = \int_{\mathbb{R}} x\, d\nu_f(x). \tag{15.4}$$

We also have

$$\operatorname{Var}(f) = \int_{\mathbb{R}} (x-a)^2\, d\nu_f(x), \quad a = E(f). \tag{15.5}$$

To illustrate some of these concepts, consider coin tosses. We start with

$$X = \{h, t\}, \quad \mu(\{h\}) = \mu(\{t\}) = \frac{1}{2}. \tag{15.6}$$

The event $\{h\}$ is that the coin comes up heads, and $\{t\}$ gives tails. We also form

$$X_k = \prod_{j=1}^{k} X, \quad \mu_k = \mu \times \cdots \times \mu, \tag{15.7}$$

representing the set of possible outcomes of k successive coin tosses. If $H(k, \ell)$ is the event that there are exactly ℓ heads among the k coin tosses, its probability is

$$\mu_k(H(k,\ell)) = 2^{-k} \binom{k}{\ell}. \tag{15.8}$$

If $N_k : X_k \to \mathbb{R}$ yields the number of heads that come up in k tosses, i.e., $N_k(x)$ is the number of h's that occur in $x = (x_1, \dots, x_k) \in X_k$, then

$$E(N_k) = \frac{k}{2}. \tag{15.9}$$

The measure ν_{N_k} on $\mathbb{R}$ is supported on the set $\{\ell \in \mathbb{Z}^+ : 0 \le \ell \le k\}$, and

$$\nu_{N_k}(\{\ell\}) = \mu_k(H(k,\ell)) = 2^{-k}\binom{k}{\ell}. \tag{15.10}$$

A central area of probability theory is the study of the large k behavior of events in spaces such as (X_k, μ_k), particularly various limits as $k \to \infty$. This leads naturally to a consideration of the infinite product space

$$Z = \prod_{j=1}^{\infty} X, \tag{15.11}$$

with σ-algebra $\mathcal{Z}$ and product measure ν, constructed at the end of Chapter 6. In case (X, μ) is given by (15.6), one can consider $\widetilde{N}_k : Z \to \mathbb{R}$, the number of heads that come up in the first k throws; $\widetilde{N}_k = N_k \circ \pi_k$, where $\pi_k : Z \to X_k$ is the natural projection. It is a fundamental fact of probability theory that for almost all infinite sequences of random coin tosses (i.e., with probability 1) the fraction $k^{-1}\widetilde{N}_k$ of them that come up heads tends to $1/2$ as $k \to \infty$. This is a special case of the "law of large numbers," several versions of which will be established below.

Note that $k^{-1}\widetilde{N}_k$ has the form

$$\frac{1}{k}\sum_{j=1}^{k} f_j, \tag{15.12}$$

where $f_j : Z \to \mathbb{R}$ has the form

$$f_j(x) = f(x_j), \quad f : X \to \mathbb{R}, \quad x = (x_1, x_2, x_3, \dots) \in Z. \tag{15.13}$$

The random variables f_j have the important properties of being *independent* and *identically distributed*.

We define these last two terms in general, for random variables on a probability space $(X, \mathfrak{F}, \mu)$ that need not have a product structure. Say we have random variables $f_1, \dots, f_k : X \to \mathbb{R}$. Extending (15.3), we have a measure ν_{F_k} on $\mathbb{R}^k$ called the joint probability distribution:

$$\nu_{F_k}(S) = \mu(F_k^{-1}(S)), \quad S \in \mathfrak{B}(\mathbb{R}^k), \quad F_k = (f_1, \dots, f_k) : X \to \mathbb{R}^k. \tag{15.14}$$

We say

$$f_1, \dots, f_k \text{ are independent } \Longleftrightarrow \nu_{F_k} = \nu_{f_1} \times \cdots \times \nu_{f_k}. \tag{15.15}$$

We also say

$$f_i \text{ and } f_j \text{ are identically distributed } \Longleftrightarrow \nu_{f_i} = \nu_{f_j}. \tag{15.16}$$

If $\{f_j : j \in \mathbb{N}\}$ is an infinite sequence of random variables on X, we say they are independent if and only if each finite subset is. Equivalently, we can form

$$F = (f_1, f_2, f_3, \dots) : X \to \mathbb{R}^\infty = \prod_{j \ge 1} \mathbb{R}, \tag{15.17}$$

set

$$\nu_F(S) = \mu(F^{-1}(S)), \quad S \in \mathfrak{B}(\mathbb{R}^\infty), \tag{15.18}$$

and then independence is equivalent to

$$\nu_F = \prod_{j \ge 1} \nu_{f_j}. \tag{15.19}$$

It is an easy exercise to show that the random variables in (15.13) are independent and identically distributed.

Here is a simple consequence of independence.

Lemma 15.1. *If $f_1, f_2 \in L^2(X,\mu)$ are independent, then*

$$E(f_1 f_2) = E(f_1)E(f_2). \tag{15.20}$$

Proof. We have $x^2, y^2 \in L^1(\mathbb{R}^2, \nu_{f_1,f_2})$, and hence $xy \in L^1(\mathbb{R}^2, \nu_{f_1,f_2})$. Then

$$\begin{aligned} E(f_1 f_2) &= \int_{\mathbb{R}^2} xy\, d\nu_{f_1,f_2}(x,y) \\ &= \int_{\mathbb{R}^2} xy\, d\nu_{f_1}(x)\, d\nu_{f_2}(y) \\ &= E(f_1)E(f_2). \end{aligned} \tag{15.21}$$

The following is a version of the *weak law of large numbers*. (See the exercises for a more general version.)

Proposition 15.2. *Let $\{f_j : j \in \mathbb{N}\}$ be independent, identically distributed random variables on $(X, \mathfrak{F}, \mu)$. Assume f_j have finite variance, and set $a = E(f_j)$. Then*

$$\frac{1}{k}\sum_{j=1}^{k} f_j \longrightarrow a, \quad \text{in } L^2\text{-norm}, \tag{15.22}$$

and hence in measure, as $k \to \infty$.

Proof. Using Lemma 15.1, we have

$$\Bigl\| \frac{1}{k}\sum_{j=1}^{k} f_j - a \Bigr\|_{L^2}^2 = \frac{1}{k^2}\sum_{j=1}^{k} \|f_j - a\|_{L^2}^2 = \frac{b^2}{k}, \tag{15.23}$$

since $(f_j - a, f_\ell - a)_{L^2} = E(f_j - a)E(f_\ell - a) = 0,\ j \neq \ell$, and $\|f_j - a\|_{L^2}^2 = b^2$ is independent of j. Clearly (15.23) implies (15.22). Convergence in measure then follows by Tchebychev's inequality.

The strong law of large numbers produces pointwise a.e. convergence and relaxes the L^2-hypothesis made in Proposition 15.2. Before proceeding to the general case, we first treat the product case (15.11)–(15.13).

Proposition 15.3. *Let $(Z, \mathcal{Z}, \nu)$ be a product of a countable number of factors of a probability space $(X, \mathfrak{F}, \mu)$, as in* (15.11). *Assume $p \in [1, \infty)$, let $f \in L^p(X, \mu)$, and define $f_j \in L^p(Z, \nu)$, as in* (15.13). *Set $a = E(f)$. Then*

$$\frac{1}{k}\sum_{j=1}^{k} f_j \to a, \quad \text{in } L^p\text{-norm and } \nu\text{-a.e.}, \tag{15.24}$$

as $k \to \infty$.

Proof. This follows from ergodic theorems established in Chapter 14. In fact, note that $f_j = T^{j-1} f_1$, where $Tg(x) = f(\varphi(x))$, for

$$\varphi : Z \to Z, \quad \varphi(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots). \tag{15.25}$$

By Proposition 14.11, φ is ergodic. We see that

$$\frac{1}{k}\sum_{j=1}^{k} f_j = \frac{1}{k}\sum_{j=0}^{k-1} T^j f_1 = A_k f_1, \tag{15.26}$$

as in (14.4). The convergence $A_k f_1 \to a$ asserted in (15.24) now follows from Proposition 14.8.

We now establish the following strong law of large numbers.

Theorem 15.4. *Let $(X, \mathfrak{F}, \mu)$ be a probability space, and let $\{f_j : j \in \mathbb{N}\}$ be independent, identically distributed random variables in $L^p(X, \mu)$, with $p \in [1, \infty)$. Set $a = E(f_j)$. Then*

$$\frac{1}{k}\sum_{j=1}^{k} f_j \to a, \quad \text{in } L^p\text{-norm and } \mu\text{-a.e.}, \tag{15.27}$$

as $k \to \infty$.

Proof. Our strategy is to reduce this to Proposition 15.3. We have a map $F : X \to \mathbb{R}^\infty$ as in (15.17), yielding a measure ν_F on $\mathbb{R}^\infty$, as in (15.18), which is actually a product measure, as in (15.19). We have coordinate functions

$$\xi_j : \mathbb{R}^\infty \longrightarrow \mathbb{R}, \quad \xi_j(x_1, x_2, x_3, \dots) = x_j, \tag{15.28}$$

and

$$f_j = \xi_j \circ F. \tag{15.29}$$

Note that $\xi_j \in L^p(\mathbb{R}^\infty, \nu_F)$ and

$$\int_{\mathbb{R}^\infty} \xi_j \, d\nu_F = \int_{\mathbb{R}} x_j \, d\nu_{f_j} = a. \tag{15.30}$$

Now Proposition 15.3 implies

$$\frac{1}{k} \sum_{j=1}^{k} \xi_j \to a, \quad \text{in } L^p\text{-norm and } \nu_F\text{-a.e.}, \tag{15.31}$$

on $(\mathbb{R}^\infty, \nu_F)$, as $k \to \infty$. Since (15.29) holds and F is measure-preserving, (15.27) follows.

Note that if f_j are independent random variables on X and $F_k = (f_1, \dots, f_k) : X \to \mathbb{R}^k$, then

$$\begin{aligned} \int_X G(f_1, \dots, f_k) \, d\mu &= \int_{\mathbb{R}^k} G(x_1, \dots, x_k) \, d\nu_{F_k} \\ &= \int_{\mathbb{R}^k} G(x_1, \dots, x_k) \, d\nu_{f_1}(x_1) \cdots d\nu_{f_k}(x_k). \end{aligned} \tag{15.32}$$

In particular, we have for the sum

$$S_k = \sum_{j=1}^{k} f_k \tag{15.33}$$

and a Borel set $B \subset \mathbb{R}$ that

$$\begin{aligned} \nu_{S_k}(B) &= \int_X \chi_B(f_1 + \cdots + f_k) \, d\mu \\ &= \int_{\mathbb{R}^k} \chi_B(x_1 + \cdots + x_k) \, d\nu_{f_1}(x_1) \cdots d\nu_{f_k}(x_k). \end{aligned} \tag{15.34}$$

Recalling the definition (9.60) of convolution of measures, we see that

$$\nu_{S_k} = \nu_{f_1} * \cdots * \nu_{f_k}. \tag{15.35}$$

Given a random variable $f : X \to \mathbb{R}$, the function

$$\chi_f(\xi) = E(e^{-i\xi f}) = \int_{\mathbb{R}} e^{-ix\xi} \, d\nu_f(x) = \sqrt{2\pi} \, \hat{\nu}_f(\xi) \tag{15.36}$$

is called the *characteristic function* of f (not to be confused with the characteristic function of a set). Combining (15.35) and (9.64), we see that if $\{f_j\}$ are independent and S_k is given by (15.33), then

$$\chi_{S_k}(\xi) = \chi_{f_1}(\xi) \cdots \chi_{f_k}(\xi). \tag{15.37}$$

There is a special class of probability distributions on $\mathbb{R}$ called *Gaussian*. They have the form

$$d\gamma_a^\sigma(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-a)^2/2\sigma}\, dx. \tag{15.38}$$

That this is a probability distribution follows from Exercise 1 of Chapter 7. One computes

$$\int x\, d\gamma_a^\sigma = a, \quad \int (x-a)^2 d\gamma_a^\sigma = \sigma. \tag{15.39}$$

The distribution (15.38) is also called *normal*, with mean a and variance σ. (Frequently one sees σ^2 in place of σ in these formulas.) A random variable f on $(X, \mathfrak{F}, \mu)$ is said to be Gaussian if ν_f is Gaussian. The computation (9.43)–(9.48) shows that

$$\sqrt{2\pi}\, \hat{\gamma}_a^\sigma(\xi) = e^{-\sigma\xi^2/2 - ia\xi}. \tag{15.40}$$

Hence $f : X \to \mathbb{R}$ is Gaussian with mean a and variance σ if and only if

$$\chi_f(\xi) = e^{-\sigma\xi^2/2 - ia\xi}. \tag{15.41}$$

We also see that

$$\gamma_a^\sigma * \gamma_b^\tau = \gamma_{a+b}^{\sigma+\tau} \tag{15.42}$$

and that if f_j are independent Gaussian random variables on X, the sum $S_k = f_1 + \cdots + f_k$ is also Gaussian.

Gaussian distributions are often approximated by distributions of the sum of a large number of independent randon variables, suitably rescaled. Theorems to this effect are called Central Limit Theorems. We present one here.

Let $\{f_j : j \in \mathbb{N}\}$ be independent, identically distributed random variables on a probability space $(X, \mathfrak{F}, \mu)$, with

$$E(f_j) = a, \quad E((f_j - a)^2) = \sigma < \infty. \tag{15.43}$$

The appropriate rescaling of $f_1 + \cdots + f_k$ is suggested by the computation (15.23). We have

$$g_k = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} (f_j - a) \Longrightarrow \|g_k\|_{L^2}^2 \equiv \sigma. \tag{15.44}$$

Note that if ν_1 is the probability distribution of $f_j - a$, then for any Borel set $B \subset \mathbb{R}$,

$$\nu_{g_k}(B) = \nu_k(\sqrt{k}B), \quad \nu_k = \nu_1 * \cdots * \nu_1 \quad (k \ \text{ factors}). \tag{15.45}$$

We have

$$\int x^2 \, d\nu_1 = \sigma, \quad \int x \, d\nu_1 = 0. \tag{15.46}$$

We are prepared to prove the following version of the Central Limit Theorem.

Proposition 15.5. *If $\{f_j : j \in \mathbb{N}\}$ are independent, identically distributed random variables on $(X, \mathfrak{F}, \mu)$, satisfying* (15.43)*, and g_k is given by* (15.44)*, then*

$$\nu_{g_k} \to \gamma_0^\sigma, \quad \textit{weak}^* \ \ \textit{in} \ \ \mathfrak{M}(\mathbb{R}) = C_*(\mathbb{R})'. \tag{15.47}$$

Proof. By (15.45) we have

$$\chi_{g_k}(\xi) = \chi(k^{-1/2}\xi)^k, \tag{15.48}$$

where $\chi(\xi) = \chi_{f_1 - a}(\xi) = \sqrt{2\pi}\hat{\nu}_1(\xi)$. By (15.46) we have $\chi \in C^2(\mathbb{R})$, $\chi'(0) = 0$, and $\chi''(0) = -\sigma$. Hence

$$\chi(\xi) = 1 - \frac{\sigma}{2}\xi^2 + r(\xi), \quad r(\xi) = o(\xi^2) \text{ as } \xi \to 0. \tag{15.49}$$

Equivalently,

$$\chi(\xi) = e^{-\sigma\xi^2/2 + \rho(\xi)}, \quad \rho(\xi) = o(\xi^2). \tag{15.50}$$

Hence

$$\chi_{g_k}(\xi) = e^{-\sigma\xi^2/2 + \rho_k(\xi)}, \tag{15.51}$$

where

$$\rho_k(\xi) = k\rho(k^{-1/2}\xi) \to 0 \ \text{ as } \ k \to \infty, \quad \forall\, \xi \in \mathbb{R}. \tag{15.52}$$

In other words,

$$\lim_{k\to\infty} \hat{\nu}_{g_k}(\xi) = \hat{\gamma}_0^\sigma(\xi), \quad \forall\, \xi \in \mathbb{R}. \tag{15.53}$$

Note that the functions in (15.53) are uniformly bounded by $\sqrt{2\pi}$. Making use of (15.53), the Fourier transform identity (9.58), and the Dominated Convergence Theorem, we obtain, for each $v \in \mathcal{S}(\mathbb{R})$,

$$\begin{aligned} \int v\, d\nu_{g_k} &= \int \tilde{v}(\xi)\hat{\nu}_{g_k}(\xi)\, d\xi \\ &\to \int \tilde{v}(\xi)\hat{\gamma}_0^\sigma(\xi)\, d\xi \\ &= \int v\, d\gamma_0^\sigma. \end{aligned} \tag{15.54}$$

Since $\mathcal{S}(\mathbb{R})$ is dense in $C_*(\mathbb{R})$ and all these measures are probability measures, this implies the asserted weak* convergence in (15.47).

Chapter 16 is devoted to the construction and study of a very important probability measure, known as Wiener measure, on the space of continuous paths in $\mathbb{R}^n$. There are many naturally occurring Gaussian random variables on this space.

We return to the strong law of large numbers and generalize Theorem 15.4 to a setting in which the f_j need not be independent. A sequence $\{f_j : j \in \mathbb{N}\}$ of real-valued random variables on $(X, \mathfrak{F}, \mu)$, giving a map $F : X \to R^\infty$ as in (15.17), is called a *stationary process* provided the probability measure ν_F on $\mathbb{R}^\infty$ given by (15.18) is invariant under the shift map

$$\theta : \mathbb{R}^\infty \longrightarrow \mathbb{R}^\infty, \quad \theta(x_1, x_2, x_3, \dots) = (x_2, x_3, x_4, \dots). \tag{15.55}$$

An equivalent condition is that for each $k, n \in \mathbb{N}$, the n-tuples $\{f_1, \dots, f_n\}$ and $\{f_k, \dots, f_{k+n-1}\}$ are identically distributed $\mathbb{R}^n$-valued random variables. Clearly a sequence of independent, identically distributed random variables is stationary, but there are many other stationary processes. (See the exercises.)

To see what happens to the averages $k^{-1}\sum_{j=1}^k f_j$ when one has a stationary process, we can follow the proof of Theorem 15.4. This time, an application of Theorem 14.6 and Proposition 14.7 to the action of θ on $(\mathbb{R}^\infty, \nu_F)$ gives

$$\frac{1}{k}\sum_{j=1}^k \xi_j \to P\xi_1, \quad \nu_F\text{-a.e. and in } L^p\text{-norm}, \tag{15.56}$$

provided

$$\xi_1 \in L^p(\mathbb{R}^\infty, \nu_F), \quad \text{i.e.,} \quad f_1 \in L^p(X, \mu), \quad p \in [1, \infty). \tag{15.57}$$

Here the map $P : L^2(\mathbb{R}^\infty, \nu_F) \to L^2(\mathbb{R}^\infty, \nu_F)$ is the orthogonal projection of $L^2(\mathbb{R}^\infty, \nu_F)$ onto the subspace consisting of θ-invariant functions, which, by Proposition 14.3, extends uniquely to a continuous projection on $L^p(\mathbb{R}^\infty, \nu_F)$ for each $p \in [1, \infty]$. Since $F : (X, \mathfrak{F}, \mu) \to (\mathbb{R}^\infty, \mathfrak{B}(\mathbb{R}^\infty), \nu_F)$ is measure preserving, the result (15.56) yields the following.

Proposition 15.6. *Let $\{f_j : j \in \mathbb{N}\}$ be a stationary process, consisting of $f_j \in L^p(X, \mu)$, with $p \in [1, \infty)$. Then*

$$\frac{1}{k} \sum_{j=1}^{k} f_j \to (P\xi_1) \circ F, \quad \mu\text{-a.e. and in } L^p\text{-norm.} \tag{15.58}$$

The right side of (15.58) can be written as a *conditional expectation.* See Exercise 12 in Chapter 17.

Exercises

1. The second computation in (15.39) is equivalent to the identity

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

Verify this identity.
Hint. Differentiate the identity

$$\int_{-\infty}^{\infty} e^{-sx^2}\, dx = \sqrt{\pi} s^{-1/2}.$$

2. Given a probability space $(X, \mathfrak{F}, \mu)$ and $A_j \in \mathfrak{F}$, we say the sets A_j, $1 \le j \le K$, are independent if and only if their characteristic functions χ_{A_j} are independent, as defined in (15.15). Show that such a collection of sets is independent if and only if, for any distinct $i_1, \dots, i_j$ in $\{1, \dots, K\}$,

$$\mu(A_{i_1} \cap \cdots \cap A_{i_j}) = \mu(A_{i_1}) \cdots \mu(A_{i_j}). \tag{15.59}$$

3. Let f_1, f_2 be random variables on $(X, \mathfrak{F}, \mu)$. Show that f_1 and f_2 are independent if and only if

$$E(e^{-i(\xi_1 f_1 + \xi_2 f_2)}) = E(e^{-i\xi_1 f_1}) E(e^{-i\xi_2 f_2}), \quad \forall\, \xi_j \in \mathbb{R}. \tag{15.60}$$

Extend this to a criterion for independence of $f_1, \dots, f_k$.
Hint. Write the left side of (15.60) as

$$\iint e^{-i(\xi_1 x_1 + \xi_2 x_2)} \, d\nu_{f_1,f_2}(x_1, x_2)$$

and the right side as a similar Fourier transform, using $d\nu_{f_1} \times d\nu_{f_2}$.

4. Demonstrate the following partial converse to Lemma 15.1.

Lemma 15.7. *Let f_1 and f_2 be random variables on (X, μ) such that $\xi_1 f_1 + \xi_2 f_2$ is Gaussian, of mean zero, for each $(\xi_1, \xi_2) \in \mathbb{R}^2$. Then*

$$E(f_1 f_2) = 0 \Longrightarrow f_1 \quad \text{and} \quad f_2 \quad \text{are independent.}$$

More generally, if $\sum_1^k \xi_j f_j$ are all Gaussian and if $f_1, \dots, f_k$ are mutually orthogonal in $L^2(X, \mu)$, then $f_1, \dots, f_k$ are independent.

Hint. Use

$$E(e^{-i(\xi_1 f_1 + \xi_2 f_2)}) = e^{-\|\xi_1 f_1 + \xi_2 f_2\|^2/2},$$

which follows from (15.41).

Exercises 5–6 deal with results known as the Borel-Cantelli Lemmas. If $(X, \mathfrak{F}, \mu)$ is a probability space and $A_k \in \mathfrak{F}$, we set

$$A = \limsup_{k \to \infty} A_k = \bigcap_{\ell=1}^{\infty} \bigcup_{k=\ell}^{\infty} A_k,$$

the set of points $x \in X$ contained in infinitely many of the sets A_k. Equivalently,

$$\chi_A = \limsup \chi_{A_k}.$$

5. (First Borel-Cantelli Lemma) Show that

$$\sum_{k \geq 1} \mu(A_k) < \infty \Longrightarrow \mu(A) = 0.$$

Hint. $\mu(A) \leq \mu\big(\bigcup_{k \geq \ell} A_k\big) \leq \sum_{k \geq \ell} \mu(A_k)$.

6. (Second Borel-Cantelli Lemma) Assume $\{A_k : k \geq 1\}$ are independent events. Show that

$$\sum_{k \geq 1} \mu(A_k) = \infty \Longrightarrow \mu(A) = 1.$$

Hint. $X \setminus A = \bigcup_{\ell \ge 1} \bigcap_{k \ge \ell} A_k^c$, so to prove $\mu(A) = 1$, we need to show $\mu\big(\bigcap_{k \ge \ell} A_k^c\big) = 0$, for each ℓ. Now independence implies

$$\mu\Big(\bigcap_{k=\ell}^{L} A_k^c\Big) = \prod_{k=\ell}^{L} \big(1 - \mu(A_k)\big),$$

which tends to 0 as $L \to \infty$ provided $\sum \mu(A_k) = \infty$.

7. If x and y are chosen at random on $[0,1]$ (with Lebesgue measure), compute the probability distribution of $x - y$ and of $(x-y)^2$. Equivalently, compute the probability distribution of f and f^2, where $Q = [0,1] \times [0,1]$, with Lebesgue measure, and $f : Q \to \mathbb{R}$ is given by $f(x,y) = x - y$.

8. As in Exercise 7, set $Q = [0,1] \times [0,1]$. If $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are chosen at random on Q, compute the probability distribution of $|x-y|^2$ and of $|x-y|$.
Hint. The random variables $(x_1 - y_1)^2$ and $(x_2 - y_2)^2$ are independent.

9. Suppose $\{f_j : j \in \mathbb{N}\}$ are independent random variables on $(X, \mathfrak{F}, \mu)$, satisfying $E(f_j) = a$ and

$$\|f_j - a\|_{L^2}^2 = \sigma_j, \quad \lim_{k \to \infty} \frac{1}{k^2} \sum_{j=1}^{k} \sigma_j = 0. \tag{15.61}$$

Show that the conclusion (15.22) of Proposition 15.2 holds.

10. Suppose $\{f_j : j \in \mathbb{N}\}$ is a stationary process on $(X, \mathfrak{F}, \mu)$. Let $G : \mathbb{R}^\infty \to \mathbb{R}$ be $\mathfrak{B}(\mathbb{R}^\infty)$-measurable, and set

$$g_j = G(f_j, f_{j+1}, f_{j+2}, \dots).$$

Show that $\{g_j : j \in \mathbb{N}\}$ is a stationary process on $(X, \mathfrak{F}, \mu)$.

In Exercises 11–12, X is a compact metric space, $\mathfrak{F}$ the σ-algebra of Borel sets, μ a probability measure on $\mathfrak{F}$, and $\varphi : X \to X$ a continuous, measure-preserving map with the property that for each $p \in X$, $\varphi^{-1}(p)$ consists of exactly d points, where $d \ge 2$ is some fixed integer. (An example is $X = S^1$, $\varphi(z) = z^d$.) Set

$$\Omega = \prod_{k \ge 1} X, \quad Z = \{(x_k) \in \Omega : \varphi(x_{k+1}) = x_k\}.$$

Note that Z is a closed subset of Ω.

11. Show that there is a unique probability measure ν on Ω with the property that for $x = (x_1, x_2, x_3, \dots) \in \Omega$, $A \in \mathfrak{F}$, the probability that $x_1 \in A$ is $\mu(A)$, and given $x_1 = p_1, \dots, x_k = p_k$, then $x_{k+1} \in \varphi^{-1}(p_k)$, and x_{k+1} has probability $1/d$ of being any one of these pre-image points.

    More formally, construct a probability measure ν on Ω such that if $A_j \in \mathfrak{F}$,

$$\nu(A_1 \times \cdots \times A_k) = \int\limits_{A_1} \cdots \int\limits_{A_k} d\gamma_{x_{k-1}}(x_k) \cdots d\gamma_{x_1}(x_2)\, d\mu(x_1),$$

    where, given $p \in X$, we set

$$\gamma_p = \frac{1}{d} \sum_{q \in \varphi^{-1}(p)} \delta_q,$$

    δ_q denoting the point mass concentrated at q. Equivalently, if $f \in C(\Omega)$ has the form $f = f(x_1, \dots, x_k)$, then

$$\int\limits_{\Omega} f\, d\nu = \int \cdots \int f(x_1, \dots, x_k)\, d\gamma_{x_{k-1}}(x_k) \cdots d\gamma_{x_1}(x_2)\, d\mu(x_1).$$

    Such ν is supported on Z.
    Hint. The construction of ν can be done in a fashion parallel to, but simpler than, the construction of Wiener measure made at the beginning of Chapter 16. One might read down to (16.13) and return to this problem.

12. Define $f_j : Z \to \mathbb{R}$ by $f_j(x) = x_j$. Show that $\{f_j : j \in \mathbb{N}\}$ is a stationary process on (Z, ν), as constructed in Exercise 11.

13. In the course of proving Proposition 15.5, it was shown that if ν_k and γ are probability measures on $\mathbb{R}$ and $\hat{\nu}_k(\xi) \to \hat{\gamma}(\xi)$ for each $\xi \in \mathbb{R}$, then $\nu_k \to \gamma$, weak* in $C_*(\mathbb{R})'$. Prove the converse:

    Assertion. *If ν_k and γ are probability measures and $\nu_k \to \gamma$ weak* in $C_*(\mathbb{R})'$, then $\hat{\nu}_k(\xi) \to \hat{\gamma}(\xi)$ for each $\xi \in \mathbb{R}$.*

    Hint. Show that for each $\varepsilon > 0$ there exist $R, N \in (0, \infty)$ such that $\nu_k(\mathbb{R} \setminus [-R, R]) < \varepsilon$ for all $k \geq N$.

14. Produce a counterexample to the assertion in Exercise 13 when ν_k are probability measures but γ is not.

15. Establish the following counterpart to Proposition 15.2 and Theorem 15.4. Let $\{f_j : j \in \mathbb{N}\}$ be independent, identically distributed random variables on a probability space $(X, \mathfrak{F}, \mu)$. Assume $f_j \geq 0$ and $\int f_j \, d\mu = +\infty$. Show that, as $k \to \infty$,

$$\frac{1}{k} \sum_{j=1}^{k} f_j \longrightarrow +\infty, \quad \mu\text{-a.e.}$$

16. Given $y \in \mathbb{R},\ t > 0$, show that

$$\chi_{t,y}(\xi) = e^{-t(1-e^{-iy\xi})}$$

is the characteristic function of the probability distribution

$$\nu_{t,y} = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{-t} \, \delta_{ky}.$$

That is, $\chi_{t,y}(\xi) = \int e^{-ix\xi} \, d\nu_{t,y}(x)$. These probability distributions are called Poisson distributions. Recalling how (9.64) leads to (15.37), show that $\nu_{s,y} * \nu_{t,y} = \nu_{s+t,y}$.

17. Suppose $\psi(\xi)$ has the property that for each $t > 0$, $e^{-t\psi(\xi)}$ is the characteristic function of a probability distribution ν_t. (One says ν_t is infinitely divisible and that ψ generates a Lévy process.) Show that if $\varphi(\xi)$ also has this property, so does $a\psi(\xi) + b\varphi(\xi)$, given $a, b \in \mathbb{R}^+$.

18. Show that whenever μ is a positive Borel measure on $\mathbb{R} \setminus 0$ such that $\int_{\mathbb{R}} (|y^2| \wedge 1) \, d\mu < \infty$, then, given $A \geq 0,\ b \in \mathbb{R}$,

$$\psi(\xi) = A\xi^2 + ib\xi + \int \bigl(1 - e^{-iy\xi} - iy\xi\chi_I(y)\bigr) \, d\mu(y)$$

has the property exposed in Exercise 17, i.e., ψ generates a Lévy process. Here, $\chi_I = 1$ on $I = [-1, 1]$, 0 elsewhere.
Hint. Apply Exercise 17 and a limiting argument. For material on Lévy processes, see [**Sat**].