

Chapter II

Independence, Conditional Expectation

The notions of statistical independence, conditional expectation, and conditional probability are the cornerstones of probability theory. Since probabilities may be expressed as expected values (integrals) of random variables, i.e., $P(A) = \mathbb{E}\mathbf{1}_A$, $A \in \mathcal{F}$, much can be gained by beginning with a formulation of independence of random maps, and conditional expectations of random variables.

Consider a finite set of random variables (maps) X_1, X_2, \dots, X_n , where each X_i is a measurable map on (Ω, \mathcal{F}, P) into (S_i, \mathcal{S}_i) ($1 \leq i \leq n$). The **product σ -field**, denoted by $S_1 \otimes \dots \otimes S_n$ is defined as the σ -field generated by the collection \mathcal{R} of **measurable rectangles** of the form $R = \{\mathbf{x} \in S_1 \times \dots \times S_n : (x_1, \dots, x_n) \in B_1 \times \dots \times B_n\}$, for $B_i \in \mathcal{S}_i$, $1 \leq i \leq n$. Alternatively, the product σ -field may be viewed, equivalently, as the smallest σ -field of subsets of $S_1 \times \dots \times S_n$ which makes each of the **coordinate projections**, say $\pi_k(\mathbf{x}) = x_k$, $\mathbf{x} \in S_1 \times \dots \times S_n$, a measurable map. In particular, if one gives $S_1 \times \dots \times S_n$ the product σ -field $S_1 \otimes \dots \otimes S_n$, then the vector $X := (X_1, \dots, X_n) : \Omega \rightarrow S_1 \times \dots \times S_n$ is a measurable map. This makes $S_1 \otimes \dots \otimes S_n$ a natural choice for a σ -field on $S_1 \times \dots \times S_n$.

The essential idea of the definition of independence of X_1, \dots, X_n below is embodied in the extension of the formula

$$P(\cap_{j=1}^n [X_j \in A_j]) = \prod_{j=1}^n P(X_j \in A_j), \quad A_j \in \mathcal{S}_j, 1 \leq j \leq n,$$

or equivalently

$$P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n),$$

for $A_j \in \mathcal{S}_j$, $1 \leq j \leq n$, to the full distribution of (X_1, \dots, X_n) . This is readily obtained via the notion of **product measure** (see Appendix A).

Definition 2.1 Finitely many random variables (maps) X_1, X_2, \dots, X_n , with X_i a measurable map on (Ω, \mathcal{F}, P) into (S_j, \mathcal{S}_j) ($1 \leq j \leq k$), are said to be **independent** if the distribution Q of (X_1, X_2, \dots, X_n) on the product space $(S = S_1 \times S_2 \times \dots \times S_n, \mathcal{S} = \mathcal{S}_1 \otimes \mathcal{S}_2 \otimes \dots \otimes \mathcal{S}_n)$ is a product measure $Q_1 \times Q_2 \times \dots \times Q_n$, where Q_j is a probability measure on (S_j, \mathcal{S}_j) , $1 \leq i \leq n$. Events $A_j \in \mathcal{S}_j$, $1 \leq j \leq n$ are said to be **independent events** if the corresponding indicator random variables $\mathbf{1}_{A_j}$, $1 \leq j \leq n$, are independent.

Notice that if X_1, \dots, X_n are independent then

$$\begin{aligned} P(X_i \in B_i) &= P((X_1, \dots, X_n) \in S_1 \times \dots \times B_i \times \dots \times S_n) \\ &= Q_1(S_1) \cdots Q_i(B_i) \cdots Q_n(S_n) \\ &= Q_i(B_i), \quad B_i \in \mathcal{S}_i, 1 \leq i \leq n. \end{aligned}$$

Moreover since, by the π - λ theorem, product measure is uniquely determined by its values on the π -system \mathcal{R} of measurable rectangles, X_1, X_2, \dots, X_n are independent if and only if $Q(B_1 \times B_2 \times \dots \times B_n) = \prod_{i=1}^n Q_i(B_i)$, $\forall B_i \in \mathcal{S}_i$, $1 \leq i \leq n$, or equivalently

$$P(X_i \in B_i, 1 \leq i \leq n) = \prod_{i=1}^n P(X_i \in B_i), \quad B_i \in \mathcal{S}_i, 1 \leq i \leq n. \quad (2.1)$$

Observe that any subcollection of finitely many independent random variables will be independent. In particular, pairs of random variables will be independent. The converse is not true (Exercise 15).

The following important application of Fubini–Tonelli readily extends to any finite sum of independent random variables, Exercise 2. Also see Exercise 8 for applications to sums of independent exponentially distributed random variables and Gaussian random variables.

Theorem 2.1 Suppose $\mathbf{X}_1, \mathbf{X}_2$ are independent k -dimensional random vectors having distributions Q_1, Q_2 , respectively. The distribution of $\mathbf{X}_1 + \mathbf{X}_2$ is given by the **convolution** of Q_1 and Q_2 :

$$Q_1 * Q_2(B) = \int_{\mathbb{R}^k} Q_1(B - y) Q_2(dy), \quad B \in \mathcal{B}^k,$$

where $B - y := \{x - y : x \in B\}$.

Proof Since sums of measurable functions are measurable, $\mathbf{S} = \mathbf{X}_1 + \mathbf{X}_2$ is a random vector, and for $B \in \mathcal{B}^k$

$$P_S(B) = P(S \in B) = P((\mathbf{X}_1, \mathbf{X}_2) \in C) = P_{X_1} \times P_{X_2}(C),$$

where $C = \{(x, y) : x + y \in B\}$. Now simply observe that $C^y = B - y$ and apply Fubini–Tonelli to $P_{X_1} \times P_{X_2}(C)$. \blacksquare

Corollary 2.2 Suppose $\mathbf{X}_1, \mathbf{X}_2$ are independent k -dimensional random vectors having distributions Q_1, Q_2 , respectively. Assume that at least one of Q_1 or Q_2 is absolutely continuous with respect to k -dimensional Lebesgue measure with pdf g . Then the distribution of $S = \mathbf{X}_1 + \mathbf{X}_2$ is absolutely continuous with density

$$f_S(s) = \int_{\mathbb{R}^k} g(s - y) Q_2(dy).$$

Proof Without loss of generality, assume Q_1 has pdf g . Then, using Theorem 2.1, change of variable, and Fubini-Tonelli, one has for any Borel set B

$$\begin{aligned} P(S \in B) &= \int_{\mathbb{R}^k} Q_1(B - y) Q_2(dy) = \int_{\mathbb{R}^k} \int_{B-y} g(z) dz Q_2(dy) \\ &= \int_B \left\{ \int_{\mathbb{R}^k} g(s - y) Q_2(dy) \right\} ds. \end{aligned} \quad (2.2)$$

This establishes the assertion. ■

From the Corollary 2.2, one sees that if both Q_1 and Q_2 have pdf's g_1, g_2 , say, then the pdf of $Q_1 * Q_2$ may be expressed as a convolution of densities $g_1 * g_2$ as given by

$$f_S(s) = \int_{\mathbb{R}^k} g_1(s - y) g_2(y) dy. \quad (2.3)$$

As given in Appendix A, the notion of product measure $\mu_1 \times \cdots \times \mu_n$ can be established for a finite number of σ -finite measure spaces $(S_1, \mathcal{S}_1, \mu_1), \dots, (S_n, \mathcal{S}_n, \mu_n)$. The σ -finiteness is essential for such important properties as **associativity** of the product, i.e., $(\mu_1 \times \mu_2) \times \mu_3 = \mu_1 \times (\mu_2 \times \mu_3)$, the Fubini-Tonelli theorem, and other such useful properties. In practice, to determine integrability of a measurable function $f : S_1 \times \cdots \times S_n \rightarrow \mathbb{R}$, one typically applies the Tonelli part (a) of the Fubini-Tonelli theorem (requiring nonnegativity) to $|f|$ in order to determine whether the Fubini part (b) is applicable to f ; cf. Appendix A.

The following result is an often used consequence of independence.

Theorem 2.3 If X_1, \dots, X_n are independent random variables on (Ω, \mathcal{F}, P) such that $\mathbb{E}|X_j| < \infty$, $1 \leq j \leq n$, then $\mathbb{E}|X_1 \cdots X_n| < \infty$ and

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n).$$

Proof Let $Q_j = P \circ X_j^{-1}$, $j \geq 1$. Since by independence, (X_1, \dots, X_n) has product measure as joint distribution, one may apply a change of variables and the Tonelli part to obtain

$$\begin{aligned}
\mathbb{E}|X_1 \cdots X_n| &= \int_{\Omega} |X_1 \cdots X_n| dP \\
&= \int_{\mathbb{R}^n} |x_1 \cdots x_n| Q_1 \times \cdots \times Q_n(dx_1 \times \cdots \times dx_n) \\
&= \prod_{j=1}^n \int_{\mathbb{R}} |x_j| Q_j(dx_j) = \prod_{j=1}^n \mathbb{E}|X_j| < \infty.
\end{aligned}$$

With the integrability established one may apply the Fubini part to do the same thing for $\mathbb{E}(X_1 \cdots X_n)$ and the product measure distribution $P \circ X_1^{-1} \times \cdots \times P \circ X_n^{-1}$ of (X_1, \dots, X_n) . ■

The role of independence in modeling occurs either as an assumption about various random variables defining a particular model or, alternatively, as a property that one may check within a specific model.

Example 1 (*Finitely Many Repeated Coin Tosses*) As a model of n -repeated tosses of a fair coin, one might assume that the successive binary outcomes are represented by a sequence of independent of 0 – 1 valued random variables X_1, X_2, \dots, X_n such that $P(X_j = 0) = P(X_j = 1) = 1/2, j = 1, \dots, n$, defined on a probability space (Ω, \mathcal{F}, P) . Alternatively, as described in Example 1 in Chapter I, one may define a probability space $\Omega = \{0, 1\}^n$, with sigma-field $\mathcal{F} = 2^\Omega$, and probability $P(\{\omega\}) = 2^{-n}$, for all $\omega \in \Omega$. Within the framework of this model one may then check that the Bernoulli variables $X_j(\omega) = \omega_j, \omega = (\omega_1, \dots, \omega_n) \in \Omega, 1 \leq j \leq n$, define a sequence of independent random variables with $P(X_j = 0) = P(X_j = 1) = 1/2, 1 \leq j \leq n$. For a parameter $p \in [0, 1]$, the model of n independent repeated tosses of a (possibly biased) coin is naturally defined as a sequence of independent Bernoulli 0 – 1 valued random variables X_1, \dots, X_n with $P(X_j = 1) = p = 1 - P(X_j = 0), 1 \leq j \leq n$.

Example 2 (*Percolation on Binary Trees*) The set $T_n = \cup_{j=0}^n \{1, 2\}^j$ may be viewed as a *rooted binary tree graph* in which, for $1 \leq j \leq n, v = (v_1, \dots, v_j) \in T_n$ is a *vertex* of length $|v| = j$, with the added convention that $\{1, 2\}^0 = \{\theta\}$ and $v = \theta$ has length $|\theta| = 0$. The special vertex θ is also designated the root. The *parent vertex* of $v = (v_1, \dots, v_j), j \geq 2$, is defined by $\overleftarrow{v} = (v_1, \dots, v_{j-1})$, with $\theta = \overleftarrow{(1)} = \overleftarrow{(2)}$. A pair of vertices v, w are connected by an *edge* if either $\overleftarrow{v} = w$ or $\overleftarrow{w} = v$. The infinite tree graph is defined by $T = \cup_{n=0}^\infty T_n = \cup_{j=0}^\infty \{1, 2\}^j$, with the corresponding definitions of vertices and edges. For $v = (v_1, v_2, \dots, v_n) \in \partial T_n := \{1, 2\}^n$, or $v = (v_1, v_2, \dots) \in \partial T := \{1, 2\}^\infty$, denote the restriction to the first j generations by $v|j = (v_1, \dots, v_j)$, with $v|0 = \theta$. Then $\theta = v|0, v|1, v|2, \dots$ may be viewed as a *path* of nearest neighboring vertices to the root. Now, let $\{X_v : v \in T\}$ be a family of independent and identically distributed (i.i.d.) Bernoulli 0 – 1-valued random variables with $p = P(X_v = 1)$. Define $X_\theta \equiv 1$ with probability one. A path $v \in \partial T$ is said to be *open* if $X_{v|j} = 1 \forall j = 0, 1, 2, \dots$. The tree graph is said to *percolate* if there is at least one open nearest neighbor path from θ to ∂T . Let B denote the event that the graph percolates. The problem is to find p such that $P(B) > 0$. Let N_n be

the number of open nearest neighbor paths from the root to ∂T_n , and N the number of open nearest neighbor paths to ∂T . Then

$$B = [N > 0] = \cap_{n=1}^{\infty} [N_n > 0], \quad [N_{n+1} > 0] \subset [N_n > 0], \quad n = 1, 2, \dots$$

So it suffices to investigate p for which there is a positive lower bound on $\lim_{n \rightarrow \infty} P(N_n > 0)$. First observe that $\mathbb{E}N_n = \mathbb{E} \sum_{|v|=n} \prod_{j=0}^n X_{v|j} = (2p)^n$. Since $P(N_n > 0) \leq \mathbb{E}N_n$, it follows that $P(N = 0) = 1$ for $p < \frac{1}{2}$, i.e., $P(B) = 0$ and the graph does not percolate. For larger p we use the second moment bound $(\mathbb{E}N_n)^2 \leq P(N_n > 0)\mathbb{E}N_n^2$ by Cauchy-Schwarz. In particular,

$$\begin{aligned} \mathbb{E}N_n^2 &= \sum_{|u|=n, |v|=n} \mathbb{E} \prod_{j=1}^n X_{u|j} X_{v|j} \\ &= \sum_{k=1}^n \sum_{|w|=k, |u|=n-k, |v|=n-k, u_1 \neq v_1} \prod_{i=1}^k \mathbb{E}X_{w|i}^2 \\ &\quad \prod_{j=k+1}^n \mathbb{E}X_{w*u|j} \mathbb{E}X_{w*v|j}, \end{aligned} \tag{2.4}$$

where $w * v = (w_1, \dots, w_k, v_1, \dots, v_{n-k})$ denotes concatenation of the paths. It follows for $2p > 1$ that $\mathbb{E}N_n^2 \leq \frac{1}{2p-1}(2p)^{2n}$. In particular, from the second moment bound one finds that $P(B) \geq \inf_n (\mathbb{E}N_n)^2 / \mathbb{E}N_n^2 \geq 2p - 1 \geq 0$ for $p > \frac{1}{2}$. The parameter value $p_c = \frac{1}{2}$ is thus the *critical* probability for percolation on the tree graph.¹ It should be clear that the methods used for the binary tree carry over to the b -ary tree for $b = 3, 4, \dots$ by precisely the same method, Exercise 13.

Two random variables X_1, X_2 in $L^2 = L^2(\Omega, \mathcal{F}, P)$ are said to be **uncorrelated** if their **covariance** $\text{Cov}(X_1, X_2)$ is zero, where

$$\text{Cov}(X_1, X_2) := \mathbb{E}[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))] = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2). \tag{2.5}$$

The **variance** $\text{Var}(Y)$ of a random variable $Y \in L^2$ is defined by the average squared deviation of Y from its mean $\mathbb{E}Y$. That is,

$$\text{Var}(Y) = \text{cov}(Y, Y) = \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}Y^2 - (\mathbb{E}Y)^2.$$

The covariance term naturally appears in consideration of the variance of sums of random variables $X_j \in L^2(\Omega, \mathcal{F}, P)$, $1 \leq j \leq n$, i.e.,

¹Criteria for percolation on the d -dimensional integer lattice is a much deeper and technically challenging problem. In the case $d = 2$ the precise identification of the critical probability for (bond) percolation as $p_c = \frac{1}{2}$ is a highly regarded mathematical achievement of Harry Kesten, see Kesten, H. (1982). For $d \geq 3$ the best known results for p_c are expressed in terms of bounds.

$$\text{Var} \left(\sum_{j=1}^n X_j \right) = \sum_{j=1}^n \text{Var}(X_j) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

Note that if X_1 and X_2 are independent, then it follows from Theorem 2.3 that they are uncorrelated; but the converse is easily shown to be false. For the record one has the following important corollary to Theorem 2.3.

Corollary 2.4 If X_1, \dots, X_n are independent random variables with finite second moment, then

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n).$$

Example 3 (Chebyshev Sample Size) Suppose that X_1, \dots, X_n is a sequence of independent and identically distributed (i.i.d.) Bernoulli 0 – 1 valued random variables with $P(X_j = 1) = p$, $1 \leq j \leq n$. As often happens in random polls, for example, the parameter p is unknown and one seeks an estimate based on the sample proportion $\hat{p}_n := \frac{X_1 + \dots + X_n}{n}$. Observe that $n\hat{p}_n$ has the Binomial distribution with parameters n, p obtained by

$$\begin{aligned} P(n\hat{p}_n = k) &= \sum_{(\varepsilon_1, \dots, \varepsilon_n) \in \{0,1\}^n: \sum_{j=1}^n \varepsilon_j = k} P(X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n) \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned} \quad (2.6)$$

for $k = 0, 1, \dots, n$. In a typical polling application one seeks a sample size n such that

$$P(|\hat{p}_n - p| > .03) \leq .05.$$

Since $\mathbb{E}|\hat{p}_n - p|^2 = \text{var}(\hat{p}_n) = np(1-p)/n^2 \leq 1/4n$, the second moment Chebyshev bound yields $n = 5,556$. This is obviously much larger than that used in a standard poll ! To improve on this one may consider a fourth moment Chebyshev bound. Rather tedious calculation yields (Exercise 3)

$$\begin{aligned} \mathbb{E}|\hat{p}_n - p|^4 &= n(p(1-p)^4 + p^4(1-p) + 3(n-1)p^2(1-p)^2)/n^4 \\ &\leq \frac{(3n-2)}{16n^3} \leq \frac{3}{16n^2}. \end{aligned}$$

Thus the fourth moment Chebyshev bound yields a reduction to see that $n = 2,154$ is a sufficient sample size. This example will be used in subsequent chapters to explore various other inequalities involving deviations from the mean.

Definition 2.2 Let $\{X_t : t \in \Lambda\}$ be a possibly infinite family of random maps on (Ω, \mathcal{F}, P) , with X_t a measurable map into (S_t, \mathcal{S}_t) , $t \in \Lambda$. We will say that $\{X_t : t \in \Lambda\}$ is a **family of independent maps** if every finite subfamily is a family

of independent maps. That is, for all $n \geq 1$ and for every n -tuple (t_1, t_2, \dots, t_n) of distinct points in Λ , the maps $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ are independent (in the sense of (2.1)).

Definition 2.3 A sequence of independent random maps X_1, X_2, \dots is said to be **independent and identically distributed**, denoted **i.i.d.**, if the distribution of X_n does not depend on n , i.e., is the same for each $n = 1, 2, \dots$

Remark 2.1 The general problem of establishing the existence of infinite families $\{X_t : t \in \Lambda\}$ of random maps, including that of infinite sequences, defined on a common probability space (Ω, \mathcal{F}, P) and having consistently specified distributions of finitely many variables $(X_{t_1}, \dots, X_{t_n})$, say, for $t_j \in \Lambda, j = 1, \dots, n$, is treated in Chapter VIII under the guise of **Kolmogorov's extension theorem**. This is a non-trivial problem of constructing probability measures with prescribed properties on an infinite product space, also elaborated upon at the close of this chapter. Kolmogorov provided a frequently useful solution, especially for countable Λ .

Let us consider the notions of uncorrelated and independent random variables a bit more fully. Before stating the main result in this regard the following proposition provides a very useful perspective on the meaning of measurability with respect to a σ -field generated by random variables.

Proposition 2.5 Let Z, Y_1, \dots, Y_k be real-valued random variables on a measurable space (Ω, \mathcal{F}) . A random variable $Z : \Omega \rightarrow \mathbb{R}$ is $\sigma(Y_1, \dots, Y_k)$ -measurable iff there is a Borel measurable function $g : \mathbb{R}^k \rightarrow \mathbb{R}$ such that $Z = g(Y_1, \dots, Y_k)$.

Proof If $Z = g(Y_1, \dots, Y_k)$, then $\sigma(Y_1, \dots, Y_k)$ -measurability is clear, since for $B \in \mathcal{B}(\mathbb{R})$, $[Z \in B] = [(Y_1, \dots, Y_k) \in g^{-1}(B)]$ and $g^{-1}(B) \in \mathcal{B}(\mathbb{R}^k)$ for Borel measurable g .

For the converse, suppose that Z is a simple $\sigma(Y_1, \dots, Y_k)$ -measurable random variable with *distinct* values z_1, \dots, z_m . Then $[Z = z_j] \in \sigma(Y_1, \dots, Y_k)$ implies that there is a $B_j \in \mathcal{B}(\mathbb{R}^k)$ such that $[Z = z_j] = [(Y_1, \dots, Y_k) \in B_j]$, since the class of all sets of the form $[(Y_1, \dots, Y_k) \in B], B \in \mathcal{B}(\mathbb{R}^k)$, is $\sigma(Y_1, \dots, Y_k)$. $Z = \sum_{j=1}^k f_j(Y_1, \dots, Y_k)$, where $f_j(y_1, \dots, y_k) = z_j \mathbf{1}_{B_j}(y_1, \dots, y_k)$, so that $Z = g(Y_1, \dots, Y_k)$ with $g = \sum_{j=1}^k f_j$. More generally, one may use approximation by simple functions to write $Z(\omega) = \lim_{n \rightarrow \infty} Z_n(\omega)$, for each $\omega \in \Omega$, where Z_n is a $\sigma(Y_1, \dots, Y_k)$ -measurable simple function, $Z_n(\omega) = g_n(Y_1(\omega), \dots, Y_k(\omega)), n \geq 1, \omega \in \Omega$. Let $\tilde{B} = \{(y_1, \dots, y_k) \in \mathbb{R}^k : \lim_{n \rightarrow \infty} g_n(y_1, \dots, y_k) \text{ exists}\}$. Then $\tilde{B} \in \mathcal{B}(\mathbb{R}^k)$ (Exercise 38). Denoting this limit by g on \tilde{B} and letting $g = 0$ on $(\tilde{B})^c$, one has $Z(\omega) = g(Y_1(\omega), \dots, Y_k(\omega))$. ■

Corollary 2.6 Suppose that X_1, X_2 are two independent random maps with values in $(S_1, \mathcal{S}_1), (S_2, \mathcal{S}_2)$, respectively. Then for Borel measurable functions $g_i : S_i \rightarrow \mathbb{R}, i = 1, 2$, the random variables $Z_1 = g_1(X_1)$ and $Z_2 = g_2(X_2)$ are independent.

Proof For Borel sets B_1, B_2 one has $g_i^{-1}(B_i) \in \mathcal{S}_i, i = 1, 2$, and

$$\begin{aligned}
P(Z_1 \in B_1, Z_2 \in B_2) &= P(X_1 \in g_1^{-1}(B_1), X_2 \in g_2^{-1}(B_2)) \\
&= P(X_1 \in g_1^{-1}(B_1))P(X_2 \in g_2^{-1}(B_2)). \quad (2.7)
\end{aligned}$$

Thus (2.1) follows. ■

Now let us return to the formulation of independence in terms of correlations. Although zero correlation is a weaker notion than statistical independence, if a sufficiently large class of functions of disjoint finite sets of random variables are uncorrelated then independence will follow. More specifically

Proposition 2.7 A family of random maps $\{X_t : t \in \Lambda\}$ (with X_t a measurable map into (S_t, \mathcal{S}_t)) is an independent family if and only if for every pair of disjoint finite subsets Λ_1, Λ_2 of Λ , any random variable $V_1 \in L^2(\sigma\{X_t : t \in \Lambda_1\})$ is uncorrelated with any random variable $V_2 \in L^2(\sigma\{X_t : t \in \Lambda_2\})$

Proof Observe that the content of the uncorrelation condition may be expressed multiplicatively as

$$\mathbb{E}V_1V_2 = \mathbb{E}V_1\mathbb{E}V_2, \quad V_i \in L^2(\sigma\{X_t : t \in \Lambda_i\}), \quad i = 1, 2.$$

Suppose the uncorrelation condition holds as stated. Let $\{t_1, \dots, t_n\} \subset \Lambda$ be an arbitrary finite subset. To establish (2.1) proceed inductively by first selecting $\Lambda_1 = \{t_1\}$, $\Lambda_2 = \{t_2, \dots, t_n\}$, with $V_1 = \mathbf{1}_{B_1}(X_{t_1})$, $V_2 = \prod_{j=2}^n \mathbf{1}_{B_j}(X_{t_j})$, for arbitrary $B_j \in \mathcal{S}_{t_j}$, $1 \leq j \leq n$. Then

$$\begin{aligned}
P(X_{t_j} \in B_j, 1 \leq j \leq n) &= \mathbb{E}V_1V_2 = \mathbb{E}V_1\mathbb{E}V_2 \\
&= P(X_{t_1} \in B_1)P(X_{t_j} \in B_j, 2 \leq j \leq n). \quad (2.8)
\end{aligned}$$

Iterating this process establishes (2.1). For the converse one may simply apply Proposition 2.5 and its Corollary to see V_1, V_2 are independent and therefore uncorrelated. ■

Definition 2.4 A collection \mathcal{C} of events $A \in \mathcal{F}$ is defined to be a set of **independent events** if the set of indicator random variables $\{\mathbf{1}_A : A \in \mathcal{C}\}$ is an independent collection.

The notion of independence of events may also be equivalently defined in terms of sub- σ -fields of \mathcal{F} .

Definition 2.5 Given (Ω, \mathcal{F}, P) , a family $\{\mathcal{F}_t : t \in \Lambda\}$ of σ -fields (contained in \mathcal{F}) is a **family of independent σ -fields** if for every n -tuple of distinct indices (t_1, t_2, \dots, t_n) in Λ one has $P(F_{t_1} \cap F_{t_2} \cap \dots \cap F_{t_n}) = P(F_{t_1})P(F_{t_2}) \dots P(F_{t_n})$ for all $F_{t_i} \in \mathcal{F}_{t_i}$ ($1 \leq i \leq n$); here n is an arbitrary finite integer, $n \leq \text{cardinality of } \Lambda$.

It is straightforward to check that the independence of a family $\{X_t : t \in \Lambda\}$ of random maps is equivalent to the independence of the family $\{\sigma(X_t) : t \in \Lambda\}$ of

σ -fields $\sigma(X_t) \equiv \{[X_t \in B] : B \in \mathcal{S}_t\}$ generated by $X_t (t \in \Lambda)$, where (S_t, \mathcal{S}_t) is the image space of X_t . One may also check that $\sigma(X_t)$ is the smallest sub- σ -field of measurable subsets of Ω that makes $X_t : \Omega \rightarrow S$ measurable. More generally

Definition 2.6 Suppose $\{X_t : t \in \Lambda\}$ is a collection of random maps defined on (Ω, \mathcal{F}) . The smallest sub- σ -field of \mathcal{F} such that every $X_t, t \in \Lambda$, is measurable, denoted $\sigma(X_t : t \in \Lambda)$, is referred to as the **σ -field generated by $\{X_t : t \in \Lambda\}$** .

Proposition 2.8 Let X_1, X_2, \dots be a sequence of independent random maps with values in measurable spaces $(S_1, \mathcal{S}_1), (S_2, \mathcal{S}_2), \dots$, respectively, and let $n_1 < n_2 < \dots$ be a nondecreasing sequence of positive integers. Suppose that $Y_1 = f_1(X_1, \dots, X_{n_1}), Y_2 = f_2(X_{n_1+1}, \dots, X_{n_2}), \dots$, where f_1, f_2, \dots are Borel-measurable functions on the respective product measure spaces $S_1 \times \dots \times S_{n_1}, S_{n_1+1} \times \dots \times S_{n_2}, \dots$. Then Y_1, Y_2, \dots is a sequence of independent random variables.

Proof It suffices to check that the distribution of an arbitrary finite subset of random variables $(Y_{n_1}, \dots, Y_{n_m}), 1 \leq n_1 < n_2 < \dots < n_m$, is product measure. But this follows readily from the distribution of (X_1, \dots, X_{n_m}) being product measure, by observing for any $k \geq 2$,

$$\begin{aligned} & P(Y_1 \in B_1, \dots, Y_k \in B_k) \\ &= P((X_1, \dots, X_{n_k}) \in f_1^{-1}(B_1), \dots, (X_{n_{k-1}+1}, \dots, X_{n_k}) \in f_k^{-1}(B_k)) \\ &= \prod_{j=1}^k P(X_{n_{j-1}+1}, \dots, X_{n_j}) \in f_j^{-1}(B_j). \end{aligned}$$

Taking $k = n_m$ and $B_i = S_i$ for $n_{j-1} < i < n_j$, the assertion follows. ■

The σ -field formulation of independence can be especially helpful in tracking independence, as illustrated by the following consequence of the $\pi - \lambda$ theorem.

Proposition 2.9 If $\{\mathcal{C}_t\}_{t \in \Lambda}$ is a family of π -systems such that $P(C_{t_1} \cap \dots \cap C_{t_n}) = \prod_{i=1}^n P(C_{t_i}), C_{t_i} \in \mathcal{C}_{t_i}$, for any distinct $t_i \in \Lambda, n \geq 2$, then $\{\sigma(\mathcal{C}_t)\}_{t \in \Lambda}$ is a family of independent σ -fields.

Proof Fix a finite set $\{t_1, \dots, t_n\} \subset \Lambda$. Fix arbitrary $C_{t_i} \in \mathcal{C}_{t_i}, 2 \leq i \leq n$. Then $\{F_1 \in \sigma(\mathcal{C}_{t_1}) : P(F_1 \cap C_{t_2} \cap \dots \cap C_{t_n}) = P(F_1)P(C_{t_2}) \dots P(C_{t_n})\}$ is a λ -system containing the π -system \mathcal{C}_{t_1} . Thus, by the $\pi - \lambda$ theorem, one has

$$P(F_1 \cap C_{t_2} \cap \dots \cap C_{t_n}) = P(F_1)P(C_{t_2}) \dots P(C_{t_n}), \forall F_1 \in \sigma(\mathcal{C}_{t_1}).$$

Now proceed inductively to obtain

$$P(F_1 \cap F_2 \cap \dots \cap F_n) = P(F_1)P(F_2) \dots P(F_n), \forall F_i \in \sigma(\mathcal{C}_{t_i}) \ 1 \leq i \leq n.$$

■

The simple example in which $\Omega = \{a, b, c, d\}$ consists of four equally probable outcomes and $\mathcal{C}_1 = \{\{a, b\}\}$, $\mathcal{C}_2 = \{\{a, c\}, \{a, d\}\}$, shows that the π -system requirement is indispensable. For a more positive perspective, note that if $A, B \in \mathcal{F}$ are independent events then it follows immediately that A, B^c and A^c, B^c are respective pairs of independent events, since $\sigma(\{A\}) = \{A, A^c, \emptyset, \Omega\}$ and similarly for $\sigma(\{B\})$, and each of the singletons $\{A\}$ and $\{B\}$ is a π -system. More generally, a collection of events $\mathcal{C} \subset \mathcal{F}$ is a collection of independent events if and only if $\{\sigma(\{C\}) : C \in \mathcal{C}\}$ is a collection of independent σ -fields. As a result, for example, C_1, \dots, C_n are independent events in \mathcal{F} if and only if the 2^n equations

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \prod_{i=1}^n P(A_i),$$

where $A_i \in \{C_i, C_i^c\}$, $1 \leq i \leq n$, are satisfied; also see Exercise 15.

Often one also needs the notion of independence of (among) several families of σ -fields or random maps.

Definition 2.7 Let $A_i, i \in \mathcal{I}$, be a family of index sets and, for each $i \in \mathcal{I}$, $\{\mathcal{F}_t : t \in A_i\}$ a collection of (sub) σ -fields of \mathcal{F} . The **families** $\{\mathcal{F}_t : t \in A_i\}_{i \in \mathcal{I}}$ are said to be **independent** (of each other) if the σ -fields $\mathcal{G}_i := \sigma(\{\mathcal{F}_t : t \in A_i\})$ generated by $\{\mathcal{F}_t : t \in A_i\}$ (i.e., \mathcal{G}_i is the smallest σ -field containing $\cup_{t \in A_i} \mathcal{F}_t, i \in \mathcal{I}$),² are independent in the sense of Definition 2.5.

The corresponding definition of **independence of (among) families of random maps** $\{X_t : t \in A_i\}_{i \in \mathcal{I}}$ can now be expressed in terms of independence of the σ -fields $\mathcal{F}_t := \sigma(X_t), t \in A_i, i \in \mathcal{I}$.

We will conclude the discussion of independence with a return to considerations of a converse to the Borel–Cantelli lemma I. Clearly, by taking $A_n = A_1, \forall n$, then $P(A_n \text{ i.o.}) = P(A_1) \in [0, 1]$. So there is no general theorem without some restriction on how much dependence exists among the events in the sequence. Write A_n **eventually for all** n to denote the event $[A_n^c \text{ i.o.}]^c$, i.e., “ A_n occurs for all but finitely many n .”

Lemma 1 (Borel–Cantelli II) Let $\{A_n\}_{n=1}^\infty$ be a sequence of independent events in a probability space (Ω, \mathcal{F}, P) . If $\sum_{n=1}^\infty P(A_n) = \infty$ then $P(A_n \text{ i.o.}) = 1$.

Proof Consider the complementary event to get from continuity properties of P , independence of complements, and the simple bound $1 - x \leq e^{-x}, x \geq 0$, that $1 \geq P(A_n \text{ i.o.}) = 1 - P(A_n^c \text{ eventually for all } n) = 1 - P(\cup_{n=1}^\infty \cap_{m=n}^\infty A_m^c) = 1 - \lim_{n \rightarrow \infty} \prod_{m=n}^\infty P(A_m^c) \geq 1 - \lim_{n \rightarrow \infty} \exp\{-\sum_{m=n}^\infty P(A_m)\} = 1$. ■

Example 4 Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent and identically distributed (i.i.d.) Bernoulli 0 or 1-valued random variables with $P(X_1 = 1) = p > 0$. Then $P(X_n = 1 \text{ i.o.}) = 1$ is a quick and easy consequence of Borel–Cantelli II.

²Recall that the σ -field \mathcal{G}_i generated by $\cup_{t \in A_i} \mathcal{F}_t$ is referred to as the *join* σ -field and denoted $\bigvee_{t \in A_i} \mathcal{F}_t$.

Example 5 (Random Power Series) Consider the random formal power series

$$S(x) = \sum_{n=0}^{\infty} \varepsilon_n x^n, \quad (2.9)$$

where $\varepsilon_0, \varepsilon_1, \dots$, is an i.i.d. sequence of random variables, *not* a.s. zero, with $\mathbb{E} \log^+ |\varepsilon_1| < \infty$. Let us see how the Borel–Cantelli lemmas can be used to determine those values of x for which $S(x)$ is actually the almost sure limit of the sequence of partial sums, i.e., almost surely a power series in x . As a warm-up notice that if for some positive constants b, c , one has $|\varepsilon_n| \leq cb^n$ a.s. eventually for all n then a.s. convergence holds for $|x| < \frac{1}{b}$. More generally, if $|x| < 1$, then there is a number $1 < b < \frac{1}{|x|}$,

$$\begin{aligned} \sum_{n=0}^{\infty} P(|\varepsilon_n| \geq b^n) &= \sum_{n=0}^{\infty} P(|\varepsilon_1| > b^n) \\ &= \sum_{n=0}^{\infty} P(\log^+ |\varepsilon_1| > n \log^+ b) \\ &\leq \frac{1}{\log^+ b} \int_0^{\infty} P(\log^+ |\varepsilon_1| > x) dx < \infty. \end{aligned} \quad (2.10)$$

It follows from Borel–Cantelli I that almost surely $|\varepsilon_n| \leq b^n$ eventually for all n . In particular the series is almost surely convergent for $|x| < 1$. Conversely, if $|x| > 1$ then there is a $1 < b < |x|$ such that

$$\sum_{n=0}^{\infty} P(|\varepsilon_n x^n| \geq b^n) = \sum_{n=0}^{\infty} P(|\varepsilon_1| > (\frac{b}{|x|})^n) = \infty, \quad (2.11)$$

since $\lim_{n \rightarrow \infty} P(|\varepsilon_1| > (\frac{b}{|x|})^n) = P(|\varepsilon_1| > 0) > 0$. Thus, using Borel–Cantelli II, one sees that $P(|\varepsilon_n x^n| > b^n \text{ i.o.}) = 1$. In particular, the series is almost surely divergent for $|x| > 1$. In the case $|x| = 1$, one may choose $\delta > 0$ such that $P(|\varepsilon_1| > \delta) > 0$. Again it follows from Borel–Cantelli II that $|\varepsilon_n x^n| \equiv |\varepsilon_n| > \delta$ i.o. with probability one. Thus the series is a.s. divergent for $|x| = 1$. Analysis of the case $\mathbb{E} \log^+ \varepsilon_1 = \infty$ is left to Exercise 20.

We now come to another basic notion of fundamental importance in probability—the notion of **conditional probability** and **conditional expectation**. It is useful to consider the spaces $L^p(\Omega, \mathcal{G}, P)$, where \mathcal{G} is a sub- σ -field of \mathcal{F} . A little thought reveals that an element of this last (Banach) space is not in general an element of $L^p(\Omega, \mathcal{F}, P)$. For if Z is \mathcal{G} -measurable, then the set (equivalence class) \tilde{Z} of all \mathcal{F} -measurable random variables each of which differs from Z on at most a P -null set may contain random variables that are not \mathcal{G} -measurable. However, if we denote by $L^p(\mathcal{G})$ the set of all elements of $L^p(\Omega, \mathcal{F}, P)$, each equivalent to some

\mathcal{G} -measurable Z with $\mathbb{E}|Z|^p < \infty$, then $L^p(\mathcal{G})$ becomes a closed linear subspace of $L^p(\Omega, \mathcal{F}, P)$. In particular, under this convention, $L^2(\mathcal{G})$ is a closed linear subspace of $L^2 \equiv L^2(\mathcal{F}) \equiv L^2(\Omega, \mathcal{F}, P)$, for every σ -field $\mathcal{G} \subset \mathcal{F}$. The first definition below exploits the Hilbert space structure of L^2 through the projection theorem (see Appendix C) to obtain the conditional expectation of X , given \mathcal{G} , as the orthogonal projection of X onto $L^2(\mathcal{G})$. Since this projection, as an element of $L^2(\Omega, \mathcal{F}, P)$, is an equivalence class which contains in general \mathcal{F} -measurable functions which may not be \mathcal{G} -measurable, one needs to select from it a \mathcal{G} -measurable version, i.e., an equivalent element of $L^2(\Omega, \mathcal{G}, P)$. If \mathcal{F} is P -complete, and so are all its sub-sigma fields, then such a modification is not necessary.

Definition 2.8 (*First Definition of Conditional Expectation (on L^2)*). Let $X \in L^2$ and \mathcal{G} be a sub- σ -field of \mathcal{F} . Then a **conditional expectation of X given \mathcal{G}** , denoted by $\mathbb{E}(X|\mathcal{G})$, is a \mathcal{G} -measurable version of the orthogonal projection of X onto $L^2(\mathcal{G})$.

Intuitively, $\mathbb{E}(X|\mathcal{G})$ is the best prediction of X (in the sense of least mean square error), given information about the experiment coded by events that constitute \mathcal{G} . In the case $\mathcal{G} = \sigma\{Y\}$ is a random map with values in a measurable space (S, \mathcal{S}) , this makes $\mathbb{E}(X|\mathcal{G})$ a version of a Borel measurable function of Y . This is because of the Proposition 2.5.

As simple examples, consider the sub- σ -fields $\mathcal{G}_0 = \{\Omega, \emptyset\}$, $\sigma(X)$, and \mathcal{F} . (The σ -field \mathcal{G}_0 , or the one comprising only P -null sets and their complements, is called the **trivial σ -field**). One has for all $X \in L^2$,

$$\mathbb{E}(X|\mathcal{G}_0) = \mathbb{E}(X), \quad \mathbb{E}(X|\sigma(X)) = X, \quad \mathbb{E}(X|\mathcal{F}) = X. \quad (2.12)$$

The first of these follows from the facts that (i) the only \mathcal{G}_0 -measurable functions are constants, and (ii) $\mathbb{E}(X - c)^2$ is minimized, uniquely, by the constant $c = \mathbb{E}X$. The other two relations in (2.12) are obvious from the definition.

For another perspective, if $X \in L^2$, then the orthogonal projection of X onto $1^\perp \equiv \{Y \in L^2 : Y \perp 1\} = \{Y \in L^2 : \mathbb{E}Y = 0\}$ is given by $X - \mathbb{E}(X)$, or equivalently, the projection of X onto the space of (equivalence classes of) constants is $\mathbb{E}(X)$.

In addition to the intuitive interpretation of $\mathbb{E}(X|\mathcal{G})$ as a best predictor of X , there is also an interpretation based on *smoothing* in the sense of averages that extends beyond L^2 . For example, as noted above, $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}X = \int_\Omega X(\omega)P(d\omega)$. In particular, this may be viewed as a *smoothing* of the function X over all sample points $\omega \in \Omega$. Similarly, and $B \in \mathcal{F}$, $0 < P(B) < 1$, and $X \in L^2$, one may check that (Exercise 24)

$$\mathbb{E}(X|\{\emptyset, B, B^c, \Omega\}) = \left(\frac{1}{P(B)} \int_B X dP \right) \mathbf{1}_B + \left(\frac{1}{P(B^c)} \int_{B^c} X dP \right) \mathbf{1}_{B^c}. \quad (2.13)$$

It is worth repeating that the conditional expectation is well defined only up to a \mathcal{G} -measurable P -null set. That is, if X is a version of $\mathbb{E}(X|\mathcal{G})$, then so is any

\mathcal{G} -measurable Y such that $P(Y \neq X) = 0$. Thus the conditional expectation $\mathbb{E}(X|\mathcal{G})$ is uniquely defined only as an element of $L^2(\Omega, \mathcal{G}, P)$. We will, however, continue to regard $\mathbb{E}(X|\mathcal{G})$ as a \mathcal{G} -measurable version of the orthogonal projection of X onto $L^2(\mathcal{G})$. The orthogonality condition is expressed by

$$\int_{\Omega} (X - \mathbb{E}(X|\mathcal{G}))Z dP = 0 \quad \forall Z \in L^2(\Omega, \mathcal{G}, P), \quad (2.14)$$

or

$$\int_{\Omega} XZ dP = \int_{\Omega} \mathbb{E}(X|\mathcal{G})Z dP \quad \forall Z \in L^2(\Omega, \mathcal{G}, P). \quad (2.15)$$

In particular, with $Z = \mathbf{1}_G$ for $G \in \mathcal{G}$ in (2.15), one has

$$\int_G X dP = \int_G \mathbb{E}(X|\mathcal{G}) dP \quad \forall G \in \mathcal{G}. \quad (2.16)$$

It is simple to check that for $X \in L^2(\Omega, \mathcal{F}, P)$, (2.16) is equivalent to (2.14) or (2.15). But (2.16) makes sense for all $X \in L^1 \equiv L^1(\Omega, \mathcal{F}, P)$, which leads to the second, more general, definition.

Definition 2.9 (*Second Definition of Conditional Expectation (on L^1)*). Let $X \in L^1(\Omega, \mathcal{F}, P)$, and let \mathcal{G} be a sub- σ -field of \mathcal{F} . A \mathcal{G} -measurable random variable is said to be a **conditional expectation of X given \mathcal{G}** , denoted by $\mathbb{E}(X|\mathcal{G})$, if (2.16) holds.

That $\mathbb{E}(X|\mathcal{G})$ exists for $X \in L^1$, and is well defined a.e., may be proved by letting $X_n \in L^2$ converge to X in L^1 (i.e., $\|X_n - X\|_1 \rightarrow 0$ as $n \rightarrow \infty$), applying (2.16) to X_n , and letting $n \rightarrow \infty$. Note that L^2 is dense in L^1 (Exercise 39). Alternatively, one may apply the Radon–Nikodym theorem to the finite (signed) measure $\nu(G) := \int_G X dP$ on (Ω, \mathcal{G}) , which is absolutely continuous with respect to P (restricted to (ω, \mathcal{G})), i.e., if $P(G) = 0$, then $\nu(G) = 0$. Hence there exists a \mathcal{G} -measurable function, denoted $\mathbb{E}(X|\mathcal{G})$, such that (2.16) holds. Viewed as an element of $L^1(\Omega, \mathcal{G}, P)$, $\mathbb{E}(X|\mathcal{G})$ is unique.

There are variations on the requirement (2.16) in the definition of conditional expectation that may be noted. In particular, a version of $\mathbb{E}(X|\mathcal{G})$ is uniquely determined by the condition that (1) it be a \mathcal{G} -measurable random variable on Ω and, (2) it satisfy the equivalent version

$$\mathbb{E}\{XZ\} = \mathbb{E}\{\mathbb{E}(X|\mathcal{G})Z\} \quad \forall Z \in \Gamma, \quad (2.17)$$

of (2.16), where, in view of the Radon–Nikodym theorem, Γ may be taken as the set of indicator random variables $\{\mathbf{1}_G : G \in \mathcal{G}\}$. There is some flexibility in the choice of Γ as a set of *test functions* when verifying and/or using the definition. Depending on the context, Γ may more generally be selected as (i) the collection of all bounded nonnegative \mathcal{G} -measurable random variables Z on Ω or (ii) the collection

of all bounded \mathcal{G} measurable random variables Z on Ω , or even the collection of all nonnegative, bounded \mathcal{G} measurable random variables Z on Ω , for example, as convenient. Certainly all of these would include the indicator random variables and therefore more than sufficient. Moreover, by simple function approximation, the defining condition extends to such choices for Γ and may be applied accordingly in computations involving conditional expectations.

The following properties of $\mathbb{E}(X|\mathcal{G})$ are important and, for the most part, immediate consequences of the definitions. As illustrated by the proofs, the basic approach to the determination of $\mathbb{E}(X|\mathcal{G})$ may be by “guessing” a \mathcal{G} -measurable random variable and then checking that it satisfies (2.17), or by starting from the left side of (2.17) and making calculations/deductions that lead to the right side, with an explicit \mathcal{G} -measurable random variable that reveals $\mathbb{E}(X|\mathcal{G})$. To check almost sure properties, on the other hand, an approach is to show that the event G , say, for which the desired property fails, has probability zero. These alternative approaches are illustrated in the proofs of the properties given in the following theorem.

Theorem 2.10 Let (Ω, \mathcal{F}, P) be a probability space, $L^1 = L^1(\Omega, \mathcal{F}, P)$, \mathcal{G}, \mathcal{D} sub- σ -fields of \mathcal{F} , $X, Y \in L^1$. Then the following holds, almost surely (P):

- (a) $\mathbb{E}(X|\{\Omega, \phi\}) = \mathbb{E}(X)$.
- (b) $\mathbb{E}[\mathbb{E}(X|\mathcal{G})] = \mathbb{E}(X)$.
- (c) If X is \mathcal{G} -measurable, then $\mathbb{E}(X|\mathcal{G}) = X$.
- (d) (*Linearity*). $\mathbb{E}(cX + dY|\mathcal{G}) = c\mathbb{E}(X|\mathcal{G}) + d\mathbb{E}(Y|\mathcal{G})$ for all constants c, d .
- (e) (*Order*). If $X \leq Y$ a.s., then $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$.
- (f) (*Smoothing*). If $\mathcal{D} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{D}] = \mathbb{E}(X|\mathcal{D})$.
- (g) If $XY \in L^1$ and X is \mathcal{G} -measurable, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.
- (h) If $\sigma(X)$ and \mathcal{G} are independent then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.
- (i) (*Conditional Jensen's Inequality*). Let ψ be a convex function on an interval J such that ψ has finite right- (or left-)hand derivative(s) at left (or right) end-point(s) of J if J is not open. If $P(X \in J) = 1$, and if $\psi(X) \in L^1$, then

$$\psi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\psi(X)|\mathcal{G}).$$

- (j) (*Contraction*). For $X \in L^p(\Omega, \mathcal{F}, P)$, $p \geq 1$, $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p \forall p \geq 1$.
- (k) (*Convergences*).
 - (k1) If $X_n \rightarrow X$ in L^p then $\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G})$ in L^p ($p \geq 1$).
 - (k2) (*Conditional Monotone Convergence*) If $0 \leq X_n \uparrow X$ a.s., X_n and $X \in L^1$ ($n \geq 1$), then $\mathbb{E}(X_n|\mathcal{G}) \uparrow \mathbb{E}(X|\mathcal{G})$ a.s. and $\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G})$ in L^1 .
 - (k3) (*Conditional Dominated Convergence*) If $X_n \rightarrow X$ a.s. and $|X_n| \leq Y \in L^1$, then $\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G})$ a.s.
- (l) (*Substitution Property*) Let U, V be random maps into (S_1, \mathcal{S}_1) and (S_2, \mathcal{S}_2) , respectively. Let ψ be a measurable real-valued function on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$. If U is \mathcal{G} -measurable, $\sigma(V)$ and \mathcal{G} are independent, and $\mathbb{E}|\psi(U, V)| < \infty$, then one has that $\mathbb{E}[\psi(U, V)|\mathcal{G}] = h(U)$, where $h(u) := E\psi(u, V)$.

(m) $\mathbb{E}(X|\sigma(Y, Z)) = \mathbb{E}(X|\sigma(Y))$ if (X, Y) and Z are independent.

Proof (a–h) follow easily from the definition. In the case of (e) take $G = [\mathbb{E}(X|\mathcal{G}) > \mathbb{E}(Y|\mathcal{G})] \in \mathcal{G}$ in the definition (2.16) of conditional expectation with X replaced by $Y - X$. To prove (g), let $Z \in \Gamma$, the set of bounded, \mathcal{G} -measurable random variables. Then, since X and XY are both integrable, $XZ \in L^1(\Omega, \mathcal{G}, P)$ and $XYZ \in L^1(\Omega, \mathcal{F}, P)$. Thus, $X\mathbb{E}(Y|\mathcal{G})$ is \mathcal{G} -measurable, and $\mathbb{E}(ZX\mathbb{E}(Y|\mathcal{G})) = \mathbb{E}(ZXY) = \mathbb{E}(Z\mathbb{E}(XY|\mathcal{G}))$. To prove (h), again let $Z \in \Gamma$ be a bounded, \mathcal{G} -measurable random variable. By independence of $\sigma(X)$ and \mathcal{G} , one has that X and Z are independent random variables; namely since $[Z \in B] \in \mathcal{G}$ for all Borel sets B , one has $P(X \in A, Z \in B) = P(X \in A)P(Z \in B)$ for all Borel sets A, B . Thus one has, using Theorem 2.3, $\mathbb{E}(ZX) = \mathbb{E}(Z)\mathbb{E}(X) = \mathbb{E}(Z\mathbb{E}(X))$. Since the constant $\mathbb{E}(X)$ is \mathcal{G} -measurable, indeed, constant random variables are measurable with respect to any σ -field, (h) follows by the defining property (2.17).

For (i) use the line of support Lemma 3 from Chapter I. If J does not have a right endpoint, take $x_0 = \mathbb{E}(X|\mathcal{G})$, and $m = \psi^+(\mathbb{E}(X|\mathcal{G}))$, where ψ^+ is the right-hand derivative of ψ , to get $\psi(X) \geq \psi(\mathbb{E}(X|\mathcal{G})) + \psi^+(\mathbb{E}(X|\mathcal{G}))(X - \mathbb{E}(X|\mathcal{G}))$. Now take the conditional expectation, given \mathcal{G} , and use (e) to get (i). Similarly, if J does not have a left endpoint, take $m = \psi^-(\mathbb{E}(X|\mathcal{G}))$. If J has both right and left endpoints, say $a < b$, let $m = \psi^+(\mathbb{E}(X|\mathcal{G}))$ on $[\mathbb{E}(X|\mathcal{G}) \neq b]$ and $m = \psi^-(\mathbb{E}(X|\mathcal{G}))$ on $[\mathbb{E}(X|\mathcal{G}) = a]$.

The contraction property (j) follows from this by taking $\psi(x) = |x|^p$ in the conditional Jensen inequality (i), and then taking expectations on both sides. The first convergence in (k) follows from (j) applied to $X_n - X$. The second convergence in (k) follows from the order property (e), and the monotone convergence theorem. The L^1 convergence in (k3) follows from (k1). For the a.s. convergence in (k3), let $Z_n := \sup\{|X_m - X| : m \geq n\}$. Then $Z_n \leq |X| + |Y|$, $|X| + |Y| - Z_n \uparrow |X| + |Y|$ a.s., so that by (k2), $\mathbb{E}(|X| + |Y| - Z_n|\mathcal{G}) \uparrow \mathbb{E}(|X| + |Y||\mathcal{G})$ a.s. Hence $\mathbb{E}(Z_n|\mathcal{G}) \downarrow 0$ a.s., and by (e), $|\mathbb{E}(X_n|\mathcal{G}) - \mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X_n - X||\mathcal{G}) \leq \mathbb{E}(Z_n|\mathcal{G}) \rightarrow 0$ a.s.

If one takes $\mathcal{G} = \sigma(U)$, then (l) follows by the Fubini–Tonelli theorem (if one uses the change of variables formula to do integrations on the product space $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2, Q_1 \times Q_2)$, where Q_1, Q_2 are the distributions of U and V , respectively). For the general case, first consider ψ of the form $\psi(u, v) = \sum_{i=1}^n f_i(u)g_i(v)$ with f_i and g_i bounded and measurable (on (S_1, \mathcal{S}_1) and (S_2, \mathcal{S}_2) , respectively), $1 \leq i \leq n$. In this case, for every $G \in \mathcal{G}$, one has $h(U) = \sum_{i=1}^n f_i(U)\mathbb{E}g_i(V)$, and

$$\begin{aligned} \int_G \psi(U, V) dP &\equiv \mathbb{E}(\mathbf{1}_G \sum_{i=1}^n f_i(U)g_i(V)) \\ &= \sum_{i=1}^n \mathbb{E}(\mathbf{1}_G f_i(U) \cdot g_i(V)) = \sum_{i=1}^n \mathbb{E}(\mathbf{1}_G f_i(U)) \cdot \mathbb{E}g_i(V) \\ &= \mathbb{E}\left(\mathbf{1}_G \left\{ \sum_{i=1}^n f_i(U) \cdot \mathbb{E}g_i(V) \right\}\right) = \mathbb{E}(\mathbf{1}_G h(U)) \\ &\equiv \int_G h(U) dP. \end{aligned}$$

The case of arbitrary $\psi(U, V) \in L^1(\Omega, \mathcal{F}, P)$ follows by the convergence result (i), noting that functions of the form $\sum_{i=1}^n f_i(u)g_i(v)$ are dense in $L^1(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2, Q_1 \times Q_2)$ (Exercise 2).

For the proof of (m) observe that for bounded, measurable g , one has using the substitution property that $\mathbb{E}(Xg(Y, Z)) = \mathbb{E}(\mathbb{E}[Xg(Y, Z)|\sigma(Z)]) = \mathbb{E}\varphi(Z)$, where $\varphi(z) = \mathbb{E}(Xg(Y, z)) = \mathbb{E}(\mathbb{E}[Xg(Y, z)|\sigma(Y)]) = \mathbb{E}(g(Y, z)\mathbb{E}[X|\sigma(Y)])$. In particular, $\mathbb{E}(Xg(Y, Z)) = \mathbb{E}(\mathbb{E}[X|\sigma(Y)]g(Y, Z))$ completes the proof of (m). ■

The following inequality illustrates a clever application of these properties, including conditional Jensen's inequality, for its proof.³ First, let $X \in L^p$, $p > 1$, be a nonnegative random variable and define

$$v_p(X) = \mathbb{E}X^p - (\mathbb{E}X)^p.$$

In particular $v_2(X)$ is the machine formula for variance of X , and the Neveu–Chauvin inequality is the all-important additive equality for variance of a sum of independent random variables; i.e., Corollary 2.4.

Proposition 2.11 (*Neveu–Chauvin Inequality*) Let X_1, X_2, \dots, X_n be nonnegative, independent random variables in L^p for some $p > 1$, and let c_1, \dots, c_n be nonnegative constants. Then, for $1 < p \leq 2$,

$$v_p\left(\sum_{j=1}^n c_j X_j\right) \leq \sum_{j=1}^n c_j^p v_p(X_j).$$

Proof By induction it is sufficient to establish $v_p(X + Y) \leq v_p(X) + v_p(Y)$ for nonnegative independent random variables X, Y . That is, one must show

$$\mathbb{E}(X + Y)^p - (\mathbb{E}X + \mathbb{E}Y)^p \leq \mathbb{E}X^p + \mathbb{E}Y^p - (\mathbb{E}X)^p - (\mathbb{E}Y)^p.$$

Noting the concavity of $x \rightarrow (x + y)^p - x^p$ on $[0, \infty)$ for fixed $y \geq 0$ and $1 < p \leq 2$, it follows from the substitution property and Jensen's inequality that

$$\mathbb{E}\{(X + Y)^p - X^p | \sigma(Y)\} \leq [\mathbb{E}(X) + Y]^p - [\mathbb{E}X]^p.$$

Thus, taking expected values, using independence and properties of conditional expectation,

$$\mathbb{E}(X + Y)^p - \mathbb{E}X^p \leq \mathbb{E}(\mathbb{E}X + Y)^p - (\mathbb{E}X)^p.$$

Applying this formula to Y and $\mathbb{E}X$ in place of X and Y , respectively, one has

$$\mathbb{E}(Y + \mathbb{E}X)^p - \mathbb{E}Y^p \leq (\mathbb{E}Y + \mathbb{E}X)^p - (\mathbb{E}Y)^p.$$

³This inequality appears in J. Neveu (1988): Multiplicative martingales for spatial branching processes, *Seminar on Stochastic Processes*, 223–242, with attribution to joint work with Brigitte Chauvin.

Thus

$$\begin{aligned}\mathbb{E}(X + Y)^p &\leq \mathbb{E}(\mathbb{E}X + Y)^p + \mathbb{E}X^p - (\mathbb{E}X)^p \\ &\leq \mathbb{E}Y^p + (\mathbb{E}Y + \mathbb{E}X)^p - (\mathbb{E}X)^p - (\mathbb{E}Y)^p + \mathbb{E}X^p \\ &= (\mathbb{E}Y + \mathbb{E}X)^p + v_p(X) + v_p(Y).\end{aligned}$$

As noted at the outset, this inequality is sufficient for the proof. ■

Specializing the notion of conditional expectation to indicator functions $\mathbf{1}_A$ of sets A in \mathcal{F} , one defines the **conditional probability of A given \mathcal{G}** , denoted by $P(A|\mathcal{G})$, by

$$P(A|\mathcal{G}) := \mathbb{E}(\mathbf{1}_A|\mathcal{G}), \quad A \in \mathcal{F}. \quad (2.18)$$

As before, $P(A|\mathcal{G})$ is a (unique) element of $L^1(\Omega, \mathcal{G}, P)$, and thus defined only up to “equivalence” by the (second) definition (2.16). That is, there are in general different versions of (2.18) differing from one another only on P -null sets in \mathcal{G} . In particular, the orthogonality condition may be expressed as follows:

$$P(A \cap G) = \int_G P(A|\mathcal{G})(\omega) P(d\omega), \quad \forall G \in \mathcal{G}. \quad (2.19)$$

It follows from properties (d), (e), (h) (linearity, order, and monotone convergence) in Theorem 2.10 that (outside \mathcal{G} -measurable P -null sets)

$$0 \leq P(A|\mathcal{G}) \leq 1, \quad P(\phi|\mathcal{G}) = 0, \quad P(\Omega|\mathcal{G}) = 1, \quad (2.20)$$

and that for every countable disjoint sequence $\{A_n\}_{n=1}^\infty$ in \mathcal{F} ,

$$P(\cup_n A_n|\mathcal{G}) = \sum_n P(A_n|\mathcal{G}). \quad (2.21)$$

In other words, conditional probability, given \mathcal{G} , has properties like those of a probability measure. Indeed, under certain conditions one may choose for each $A \in \mathcal{F}$ a version of $P(A|\mathcal{G})$ such that $A \rightarrow P(A|\mathcal{G})(\omega)$ is a probability measure on (Ω, \mathcal{F}) for every $\omega \in \Omega$. However, such a probability measure may not exist in the full generality in which conditional expectation is defined.⁴ The technical difficulty in constructing the conditional probability measure (for each $\omega \in \Omega$) is that each one of the relations in (2.20) and (2.21) holds outside a P -null set, and individual P -null sets may pile up to a nonnull set. Such a probability measure, when it exists, is called a **regular conditional probability measure given \mathcal{G}** , and denoted by $P^\mathcal{G}(A)(\omega)$. It is more generally available as a probability measure (for each ω outside a P -null set) on appropriate sub- σ -fields of \mathcal{F} (even if it is not a probability measure on all

⁴Counterexamples have been constructed, see for example, Halmos (1950), p. 210.

of \mathcal{F}). An important case occurs under the terminology of a **regular conditional distribution** of a random map Z (on (Ω, \mathcal{F}, P) into some measurable space (S, \mathcal{S})).

Definition 2.10 Let Y be a random map on (Ω, \mathcal{F}, P) into (S, \mathcal{S}) . Let \mathcal{G} be a sub- σ -field of \mathcal{F} . A **regular conditional distribution of Y given \mathcal{G}** is a function $(\omega, C) \rightarrow Q^{\mathcal{G}}(\omega, C) \equiv P^{\mathcal{G}}([Y \in C]|\mathcal{G})(\omega)$ on $\Omega \times S$ such that

- (i) $\forall C \in \mathcal{S}, Q^{\mathcal{G}}(\cdot, C) = P([Y \in C]|\mathcal{G})$ a.s. (and $Q^{\mathcal{G}}(\cdot, C)$ is \mathcal{G} -measurable),
- (ii) $\forall \omega \in \Omega, C \rightarrow Q^{\mathcal{G}}(\omega, C)$ is a probability measure on (S, \mathcal{S}) .

The following definition provides a topological framework in which one can be assured existence of a regular version of the conditional distribution of a random map.

Definition 2.11 A topological space S whose topology can be induced by a metric is said to be **metrizable**. If S is metrizable as a complete and separable metric space then S is referred to as a **Polish space**.

For our purposes a general existence theorem as the Doob–Blackwell theorem⁵ stated in the footnote will be unnecessary for the present text since we will have an explicit expression of $Q^{\mathcal{G}}$ given directly when needed. Once $Q^{\mathcal{G}}$ is given, one can calculate $\mathbb{E}(f(Y)|\mathcal{G})$ (for arbitrary functions f on (S, \mathcal{S}) such that $f(Y) \in L^1$) as

$$\mathbb{E}(f(Y)|\mathcal{G}) = \int f(y) Q^{\mathcal{G}}(\cdot, dy). \quad (2.22)$$

This formula holds for $f(y) = \mathbf{1}_C(y) \forall C \in \mathcal{S}$ by definition. The general result follows by approximation of f by simple functions, using linearity and convergence properties of conditional expectation (and of corresponding properties of integrals with respect to a probability measure $Q^{\mathcal{G}}(\omega, \cdot)$). Combining (2.22) with Theorem 2.10(b) yields the so-called **disintegration formula**

$$\mathbb{E}(f(Y)) = \int_{\Omega} \int f(y) Q^{\mathcal{G}}(\omega, dy) P(d\omega). \quad (2.23)$$

The conditional Jensen inequality (i) of Theorem 2.10 follows from the existence of a regular conditional distribution of X , given \mathcal{G} . The following simple examples tie up the classical concepts of conditional probability with the more modern general framework presented above.

Example 6 Let $B \in \mathcal{F}$ be such that $P(B) > 0, P(B^c) > 0$. Let $\mathcal{G} = \sigma(B) \equiv \{\Omega, B, B^c, \emptyset\}$. Then for every $A \in \mathcal{F}$ one has

⁵The Doob–Blackwell theorem provides the existence of a regular conditional distribution of a random map Y , given a σ -field \mathcal{G} , taking values in a Polish space equipped with its Borel σ -field $\mathcal{B}(S)$. For a proof, see Breiman (1968), pp. 77–80.

$$P(A|\mathcal{G})(\omega) = \begin{cases} P(A|B) := \frac{P(A \cap B)}{P(B)}, & \text{if } \omega \in B \\ P(A|B^c) := \frac{P(A \cap B^c)}{P(B^c)}, & \text{if } \omega \in B^c. \end{cases}$$

More generally, let $\{B_n : n = 1, 2, \dots\}$ be a countable disjoint sequence in \mathcal{F} such that $\cup_n B_n = \Omega$, called a **partition** of Ω . Let $\mathcal{G} = \sigma(\{B_n : n \geq 1\})$ (\mathcal{G} is the class of all unions of sets in this countable collection). Then for every A in \mathcal{F} , assuming $P(B_n) > 0$, one has

$$P(A|\mathcal{G})(\omega) = \frac{P(A \cap B_n)}{P(B_n)} \quad \text{if } \omega \in B_n. \quad (2.24)$$

If $P(B_n) = 0$ then for $\omega \in B_n$, define $P(A|\mathcal{G})(\omega)$ to be some constant, say c , chosen arbitrarily (Exercise 2).

Remark 2.2 Let $Y \in L^1(\Omega, \mathcal{F}, P)$ and suppose X is a random map on (Ω, \mathcal{F}, P) with values in (S, \mathcal{S}) . In view of Proposition 2.5, $\mathbb{E}(Y|\sigma(X))$ is a function of X , say $f(X)$, and thus constant on each event $[X = x]$, $x \in S$; i.e., $\mathbb{E}(Y|\sigma(X))(\omega) = f(X(\omega)) = f(x)$, $\omega \in [X = x] = \{\omega \in \Omega : X(\omega) = x\}$. In particular, the notation $\mathbb{E}(Y|X = x)$ may be made precise by defining $\mathbb{E}(Y|X = x) := f(x)$, $x \in S$.

Example 7 (*A Canonical Probability Space*) Let $(S_i, \mathcal{S}_i, \mu_i)$, $i = 1, 2$ be two σ -finite measure spaces, that may serve as the image spaces of a pair of random maps. The *canonical probability model* is constructed on the image space as follows. Let $\Omega = S_1 \times S_2$, $\mathcal{F} = \mathcal{S}_1 \otimes \mathcal{S}_2$. Assume that the probability P is absolutely continuous with respect to $\mu = \mu_1 \times \mu_2$ with density f i.e., f is a nonnegative \mathcal{F} -measurable function such that $\int_{\Omega} f d\mu = 1$, and $P(A) = \int_A f d\mu$, $A \in \mathcal{F}$. One may view P as the distribution of the joint coordinate maps (X, Y) , where $X(\omega) = x$, $Y(\omega) = y$, for $\omega = (x, y) \in S_1 \times S_2$. The σ -field $\mathcal{G} = \{B \times S_2 : B \in \mathcal{S}_1\}$ is the σ -field generated by the first coordinate map X . A little thought leads naturally to a reasonable guess for a (regular) conditional distribution of Y given $\sigma(X)$. Namely, for every event $A = [Y \in C] = S_1 \times C$ ($C \in \mathcal{S}_2$), one has

$$P(A|\mathcal{G})(\omega) = \frac{\int_C f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y') \mu_2(dy')} \quad \text{if } \omega = (x, v) \in \Omega. \quad (2.25)$$

To check this, first note that by the Fubini–Tonelli theorem, the function D defined by the right-hand side of (2.25) is \mathcal{G} -measurable. Second, for every nonnegative bounded Borel measurable g on S_1 , i.e., \mathcal{G} -measurable test random variables $Z = g(X)$, one has

$$\begin{aligned}
\mathbb{E}(ZD) &= \int_{S_1 \times S_2} g(x) \frac{\int_C f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y) \mu_2(dy)} f(x, y') \mu_1(dx) \mu_2(dy') \\
&= \int_{S_1} g(x) \left\{ \int_{S_2} \frac{\int_C f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y') \mu_2(dy')} f(x, y') \mu_2(dy') \right\} \mu_1(dx) \\
&= \int_{S_1} g(x) \left\{ \frac{\int_C f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y') \mu_2(dy')} \cdot \int_{S_2} f(x, y') \mu_2(dy') \right\} \mu_1(dx) \\
&= \int_{S_1} g(x) \left\{ \int_C f(x, y) \mu_2(dy) \right\} \mu_1(dx) = \mathbb{E}(\mathbf{1}_{S_1 \times C} g(X)) \\
&= \mathbb{E}(Z \mathbf{1}_{\{Y \in C\}}).
\end{aligned}$$

In particular, $P(A|\mathcal{G}) = \mathbb{E}(\mathbf{1}_A|\mathcal{G}) = D$. The function $f(x, y)/\int_{S_2} f(x, y) \mu_2(dy)$ is called the **conditional pdf of Y given $X = x$** , and denoted by $f(y|x)$; i.e., the conditional pdf is simply the joint density (section) $y \rightarrow f(x, y)$ normalized to a probability density by dividing by the (marginal) pdf $f_X(x) = \int_{S_2} f(x, y) \mu_2(dy)$ of X . Let $A \in \mathcal{F} = S_1 \otimes S_2$. By the same calculations using Fubini–Tonelli one more generally obtains

$$P(A|\mathcal{G})(\omega) = \int_{A_x} f(y|x) \mu_2(dy) \equiv \frac{\int_{A_x} f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y) \mu_2(dy)} \quad \text{if } \omega \equiv (x, y'), \quad (2.26)$$

where $A_x = \{y \in S_2 : (x, y) \in A\}$.

One may change the perspective here a little and let (Ω, \mathcal{F}, P) be any probability space on which are defined two maps X and Y with values in (S_1, \mathcal{S}_1) and (S_2, \mathcal{S}_2) , respectively. If the (joint) distribution of (X, Y) on $(S_1 \times S_2, S_1 \otimes S_2)$ has a pdf f with respect to a product measure $\mu_1 \times \mu_2$, where μ_i is a σ -finite measure on (S, \mathcal{S}_i) , $i = 1, 2$, then for $\mathcal{G} = \sigma(X)$, after using the change of variable formula mapping $\Omega \rightarrow S_1 \times S_2$, precisely the same calculations show that the (regular) conditional distribution of Y given \mathcal{G} (or “given X ”) is given by

$$P([Y \in C]|\mathcal{G})(\omega) = \frac{\int_C f(x, y) \mu_2(dy)}{\int_{S_2} f(x, y) \mu_2(dy)} \quad \text{if } X(\omega) = x, \quad (2.27)$$

i.e., if $\omega \in [X = x] \equiv X^{-1}\{x\}$, $x \in S_1$. Note that the conditional probability is constant on $[X = x]$ as required for $\sigma(X)$ -measurability; cf Proposition 2.5.

Two particular frameworks in which conditional probability and conditional expectation are very prominent are those of (i) Markov processes and (ii) martingales. The former is most naturally expressed as the property that the conditional distribution of future states of a process, given the past and present states coincides with the conditional distribution given the present, and the latter is the property that the conditional expectation of a future state given past and present states, is simply the present. The Markov property is illustrated in the next example, and martingales are the topic of the next chapter.

Example 8 (Markov Property for General Random Walks on \mathbb{R}^k) Let $\{Z_n : n \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) k -dimensional random vectors defined on a probability space (Ω, \mathcal{F}, P) . Let μ denote the distribution of Z_1 (hence of each Z_n). For arbitrary $x \in \mathbb{R}^k$, a **random walk starting at x with step-size distribution μ** is defined by the sequence $S_n^x := x + Z_1 + \cdots + Z_n$ ($n \geq 1$), $S_0^x = x$.

For notational simplicity we will restrict to the case of $k = 1$ dimensional random walks, however precisely the same calculations are easily seen to hold for arbitrary $k \geq 1$ (Exercise 2). Let Q_x denote the distribution of $\{S_n^x : n \geq 0\}$ on the product space $(\mathbb{R}^\infty, \mathcal{B}^\infty)$. Here \mathcal{B}^∞ is the σ -field generated by cylinder sets of the form $C = B_m \times \mathbb{R}^\infty := \{\mathbf{y} = (y_0, y_1, \dots) \in \mathbb{R}^\infty; (y_0, y_1, \dots, y_m) \in B_m\}$ with B_m a Borel subset of \mathbb{R}^{m+1} ($m = 0, 1, 2, \dots$). Note that $Q_x(B_m \times \mathbb{R}^\infty) = P((S_0^x, S_1^x, \dots, S_m^x) \in B_m)$, so that $Q_x(B_m \times \mathbb{R}^\infty)$ may be expressed in terms of the m -fold product measure $\mu \times \mu \times \cdots \times \mu$, which is the distribution of (Z_1, Z_2, \dots, Z_m) . For our illustration, let $\mathcal{G}_n = \sigma(\{S_j^x : 0 \leq j \leq n\}) = \sigma(\{Z_1, Z_2, \dots, Z_n\})$ ($n \geq 1$). We would like to establish the following property: **The conditional distribution of the “after- n process”** $S_n^{x+} := \{S_{n+m}^x : m = 0, 1, 2, \dots\}$ on $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ **given \mathcal{G}_n is $Q_y|_{y=S_n^x} \equiv Q_{S_n^x}$** . In other words, for the random walk $\{S_n^x : n \geq 0\}$, the conditional distribution of the *future evolution* defined by S_n^{x+} , given the *past states* S_0^x, \dots, S_{n-1}^x and *present state* S_n^x , depends solely on the present state S_n^x , namely, $Q_{S_n^x}$ i.e., it is given by the regular conditional distribution $Q^{\mathcal{G}_n}(\omega, \cdot) = Q_{S_n^x(\omega)}(\cdot)$.

Theorem 2.12 (Markov Property) For every $n \geq 1$, the conditional distribution of S_n^{x+} given $\sigma(S_0^x, \dots, S_n^x)$ is a function only of S_n^x .

Proof To prove the theorem choose a cylinder set $C \in \mathcal{B}^\infty$. That is, $C = B_m \times \mathbb{R}^\infty$ for some $m \geq 0$. We want to show that

$$P([S_n^{x+} \in C] | \mathcal{G}_n) \equiv \mathbb{E}(\mathbf{1}_{[S_n^{x+} \in C]} | \mathcal{G}_n) = Q_{S_n^x}(C). \quad (2.28)$$

Now $[S_n^{x+} \in C] = [(S_n^x, S_n^x + Z_{n+1}, \dots, S_n^x + Z_{n+1} + \cdots + Z_{n+m}) \in B_m]$, so that one may write

$$\mathbb{E}(\mathbf{1}_{[S_n^{x+} \in C]} | \mathcal{G}_n) = \mathbb{E}(\psi(U, V) | \mathcal{G}_n),$$

where $U = S_n^x$, $V = (Z_{n+1}, Z_{n+2}, \dots, Z_{n+m})$ and, for $u \in \mathbb{R}$ and $v \in \mathbb{R}^m$, $\psi(u, v) = \mathbf{1}_{B_m}(u, u + v_1, u + v_1 + v_2, \dots, u + v_1 + \cdots + v_m)$. Since S_n^x is \mathcal{G}_n -measurable and V is independent of \mathcal{G}_n , it follows from property (ℓ) of Theorem 2.10 that $\mathbb{E}(\psi(U, V) | \mathcal{G}_n) = h(S_n^x)$, where $h(u) = \mathbb{E}\psi(u, V)$. But

$$\begin{aligned} \mathbb{E}\psi(u, V) &= P((u, u + Z_{n+1}, \dots, u + Z_{n+1} + \cdots + Z_{n+m}) \in B_m) \\ &= P((u, u + Z_1, u + Z_1 + Z_2, \dots, u + Z_1 + \cdots + Z_m) \in B_m) \\ &= P((S_0^u, S_1^u, \dots, S_m^u) \in B_m) = Q_u(C). \end{aligned}$$

Therefore, $P([S_n^{x+} \in C] | \mathcal{G}_n) = (Q_u(C))_{u=S_n^x} = Q_{S_n^x}(C)$. We have now shown that the class \mathcal{L} of sets $C \in \mathcal{B}^\infty$ for which “ $P([S_n^{x+} \in C] | \mathcal{G}_n) = Q_{S_n^x}(C)$ a.s.”

holds contains the class \mathcal{C} of all cylinder sets. Since this class is a λ -system (see the convergence property (k) of Theorem 2.10) containing the π -system of cylinder sets that generate \mathcal{B}^∞ , it follows by the $\pi - \lambda$ theorem that $\mathcal{L} = \mathcal{B}^\infty$. ■

Example 9 (Recurrence of 1-d Simple Symmetric Random Walk) Let X_0 be an integer-valued random variable, and X_1, X_2, \dots an i.i.d. sequence of symmetric Bernoulli ± 1 -valued random variables, independent of X_0 , defined on a probability space (Ω, \mathcal{F}, P) . For $x \in \mathbb{Z}$, the sequence of random variables $S^x = \{S_0^x, S_1^x, \dots\}$ defined by $S_0^x = x$, $S_n^x = x + X_1 + \dots + X_n$, $n = 1, 2, \dots$, is referred to as the *one-dimensional simple symmetric random walk on \mathbb{Z} started at x* . Let $Q^x = P \circ (S^x)^{-1}$ denote the distribution of S^x . Observe that the conditional distribution of the after-one process $S_{+1}^x = (S_1^x, S_2^x, \dots)$ given $\sigma(X_1)$ is the composition $\omega \rightarrow Q^{S_1(\omega)}$. Thus, letting $R = \bigcup_{n=1}^\infty \{(x_0, x_1, x_2, \dots) \in \mathbb{Z}^\infty : x_n = x_0\}$, $[S^x \in R]$ denotes the event of eventual return to x , one has

$$\begin{aligned} Q^x(R) &= P(S^x \in R) = \mathbb{E}P(S^x \in R | \sigma(X_1)) \\ &= \mathbb{E}Q^{S_1^x}(R) = Q^{x+1}(R)P(S_1^x = x+1) + Q^{x-1}(R)P(S_1^x = x-1) \\ &= Q^{x+1}(R)\frac{1}{2} + Q^{x-1}(R)\frac{1}{2}. \end{aligned} \quad (2.29)$$

Since $Q^x(R) \geq P(X_1 = 1, X_2 = -1) = 1/4 > 0$, it follows that the only solution to (2.29) is $Q^x(R) = 1$. Note that this does not depend on the particular choice of $x \in \mathbb{Z}$, and is therefore true for all $x \in \mathbb{Z}$. Thus the simple symmetric random walk started at any $x \in \mathbb{Z}$ is certain to eventually return to x .

Independence and conditional probability underly most theories of stochastic processes in fundamental ways. From the point of view of ideas developed thus far, the general framework is as follows. A **stochastic process** $\{X_t : t \in \Lambda\}$ on a probability space (Ω, \mathcal{F}, P) is a family of random maps $X_t : \Omega \rightarrow S_t$, $t \in \Lambda$, for measurable spaces (S_t, \mathcal{S}_t) , $t \in \Lambda$. The index set Λ is most often of one of the following types: (i) $\Lambda = \{0, 1, 2, \dots\}$. Then $\{X_t : t = 0, 1, 2, \dots\}$ is referred to as a **discrete-parameter stochastic process**, usually with $S = \mathbb{Z}^k$ or \mathbb{R}^k . (ii) $\Lambda = [0, \infty)$. Then $\{X_t : t \geq 0\}$ is called a **continuous-parameter stochastic process**, usually with $S = \mathbb{Z}^k$ or \mathbb{R}^k .

Given an arbitrary collection of sets S_t , $t \in \Lambda$, the **product space**, denoted by $S = \prod_{t \in \Lambda} S_t \equiv \times_{t \in \Lambda} S_t$, is defined as the space of functions $\mathbf{x} = (x_t, t \in \Lambda)$ mapping Λ to $\bigcup_{t \in \Lambda} S_t$ such that $x_t \in S_t$ for each $t \in \Lambda$. This general definition applies to cases in which Λ is finite, countably infinite, or a continuum. In the case that each S_t , $t \in \Lambda$, is also a measurable space with respective σ -fields \mathcal{S}_t , the product σ -field, denoted by $\otimes_{t \in \Lambda} \mathcal{S}_t$, is defined as the σ -field generated by the collection \mathcal{R} of **finite-dimensional rectangles** of the form $R = \{\mathbf{x} \in \prod_{t \in \Lambda} S_t : (x_{t_1}, \dots, x_{t_k}) \in B_1 \times \dots \times B_k\}$, for $k \geq 1$, $B_i \in \mathcal{S}_{t_i}$, $1 \leq i \leq k$. Alternatively, the product σ -field is the smallest σ -field of subsets of $\prod_{t \in \Lambda} S_t$ which makes each of the **coordinate projections**, $\pi_s(\mathbf{x}) = x_s$, $\mathbf{x} \in \prod_{t \in \Lambda} S_t$, $s \in \Lambda$, a measurable map. In this case the pair $(S = \prod_{t \in \Lambda} S_t, \otimes_{t \in \Lambda} \mathcal{S}_t)$ is the (measure-theoretic) **product space**.

As previously noted, it is natural to ask whether, as already known for finite products, given any family of probability measures Q_t (on (S_t, \mathcal{S}_t)), $t \in \Lambda$, can one construct a probability space (Ω, \mathcal{F}, P) on which are defined random maps X_t ($t \in \Lambda$) such that (i) X_t has distribution Q_t ($t \in \Lambda$) and (ii) $\{X_t : t \in \Lambda\}$ is a family of independent maps? Indeed, on the product space $(S \equiv \times_{t \in \Lambda} S_t, \mathcal{S} \equiv \otimes_{t \in \Lambda} \mathcal{S}_t)$ it will be shown in Chapter VIII that there exists such a product probability measure $Q = \prod_{t \in \Lambda} Q_t$; and one may take $\Omega = S$, $\mathcal{F} = \mathcal{S}$, $P = Q$, $X_t(\omega) = x_t$ for $\omega = (x_t, t \in \Lambda) \in S$. The practical utility of such a construction for infinite Λ is somewhat limited to the case when Λ is denumerable. However such a product probability space for a countable sequence of random maps X_1, X_2, \dots is remarkably useful.

The important special case of a sequence X_1, X_2, \dots of **independent** and **identically distributed** random maps is referred to as an **i.i.d.** sequence. An example of the construction of an i.i.d. (coin tossing) sequence $\{X_n\}_{n=1}^\infty$ of Bernoulli-valued random variables with values in $\{0, 1\}$ and defined on a probability space (Ω, \mathcal{F}, P) with prescribed distribution $P(X_1 = 1) = p = 1 - P(X_1 = 0)$, for given $p \in [0, 1]$, is given in Exercise 37. As remarked above, the general existence of infinite product measures will be proved in Chapter VIII. This is a special case of the **Kolmogorov extension theorem** proved in Chapter VIII in the case that (S, \mathcal{S}) has some extra topological structure; see Exercise 37 for a simple special case illustrating how one may exploit topological considerations. Existence of an infinite product probability measure will also be seen to follow in full measure-theoretic generality, i.e., without topological requirements on the image spaces, from the **Tulcea extension theorem** discussed in Chapter VIII.

Exercise Set II

1. Suppose that X_1, \dots, X_n are independent random maps defined on a probability space (Ω, \mathcal{F}, P) . Show that the product measure $Q = P \circ (X_1, \dots, X_n)^{-1}$ is given by $Q_1 \times \dots \times Q_n$, where $Q_i = P \circ X_i^{-1}$. Also show that any subset of $\{X_1, \dots, X_n\}$ comprises independent random maps.
2. Suppose $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are independent k -dimensional random vectors having distributions Q_1, Q_2, \dots, Q_n , respectively. Prove that the distribution of $\mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_n$, $n \geq 2$, is given by the n -fold convolution $Q^{*n} = Q_1 * Q_2 * \dots * Q_n$ inductively defined by $Q^{*n}(B) = \int_{\mathbb{R}^k} Q^{*(n-1)}(B - x) Q_n(dx)$, where $B - x := \{y - x : y \in B\}$ for Borel sets $B \subset \mathbb{R}^k$, $Q^{*(1)} = Q_1$.
3. Let X_1, \dots, X_n be i.i.d. random variables with finite variance $\sigma^2 = \mathbb{E}(X_1 - \mathbb{E}X_1)^2$, and finite central fourth moment $\mu_4 = \mathbb{E}(X_1 - \mathbb{E}X_1)^4$. Let $S_n = \sum_{j=1}^n X_j$. (a) Show $\mathbb{E}(S_n - n\mathbb{E}X_1)^4 = n\mu_4 + 3n(n-1)\sigma^4$. (b) For the case in which X_1, \dots, X_n is the i.i.d. Bernoulli 0–1 valued sequence given in Example 3 show that both the variance and the fourth central moment of $n\hat{p}_n \equiv S_n$ are maximized at $p = 1/2$, and determine the respective maximum values.
4. Let X_1, X_2 be random maps with values in σ -finite measure spaces $(S_1, \mathcal{S}_1, \mu_1)$ and $(S_2, \mathcal{S}_2, \mu_2)$, respectively. Assume that the distribution of (X_1, X_2) has a pdf f with respect to product measure $\mu_1 \times \mu_2$, i.e., f is a nonnegative measurable function such that

$$P((X_1, X_2) \in B) = \int_B f(x_1, x_2) \mu_1 \times \mu_2(dx_1 \times dx_2), B \in \mathcal{S}_1 \otimes \mathcal{S}_2.$$

Show that X_1 and X_2 are independent if and only if $f(x_1, x_2) = f_1(x_1)f_2(x_2)\mu_1 \times \mu_2 - a.e.$ for some $0 \leq f_i, \int_{\mathcal{S}_i} f_i d\mu_i = 1, (i = 1, 2)$.

5. Suppose that X_1, X_2, \dots is a sequence of independent random variables on (Ω, \mathcal{F}, P) . Show that the two families $\{X_1, X_3, X_5, \dots\}$ and $\{X_2, X_4, X_6, \dots\}$ are independent. [Hint: Express $\sigma(X_1, X_3, \dots) = \sigma(\mathcal{C}_1)$, for the π -system $\mathcal{C}_1 = \{(X_1, X_3, \dots, X_{2m-1}) \in A_1 \times \dots \times A_m : A_j \in \mathcal{B}, 1 \leq j \leq m, m \geq 1\}$, and similarly for the even indices.]
6. (a) Show that the coordinate variables $(U_{n,1}, \dots, U_{n,n})$ in Exercise 31 of Chapter I are independent, and each is uniformly distributed over $[0, 2]$. (b) Show that if $\frac{X_n}{n} \rightarrow c$ in probability as $n \rightarrow \infty$ for a constant $c < 0$, then $X_n \rightarrow -\infty$ in probability in the sense that for any $\lambda < 0$, $P(X_n > \lambda) \rightarrow 0$ as $n \rightarrow \infty$. (c) Taking logarithms and using (a) and (b) together with a Chebychev inequality, show that $U_{n,1} \cdots U_{n,n} \rightarrow 0$ in probability as $n \rightarrow \infty$. (d) Use Jensen's inequality to extend (c) to any independent, positive, mean one random variables $(U_{n,1}, \dots, U_{n,n})$ having finite second moment.
7. Show that if (U_1, U_2) is uniformly distributed on the disc $D = \{(x, y) : x^2 + y^2 \leq 1\}$, i.e., distributed as a multiple $(\frac{1}{\pi})$ of Lebesgue measure on D , then X and Y are not independent. Compute $\text{Cov}(X, Y)$.
8. Let X_1, X_2, \dots, X_n be i.i.d. random variables defined on (Ω, \mathcal{F}, P) and having (common) distribution Q .
 - (i) Suppose $Q(dx) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty)}(x) dx$, for some $\lambda > 0$, referred to as the *exponential distribution with parameter λ* . Show that $X_1 + \dots + X_n$ has distribution $Q^{*n}(dx) = \lambda^n \frac{x^{n-1}}{(n-1)!} e^{-\lambda x} \mathbf{1}_{[0, \infty)}(x) dx$. This latter distribution is referred to as a *gamma distribution* with parameters n, λ .
 - (ii) Suppose that $Q(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$; referred to as the *Gaussian* or *normal* distribution with parameters $\mu \in \mathbb{R}, \sigma^2 > 0$. Show that $X_1 + \dots + X_n$ has a normal distribution with parameters $n\mu$ and $n\sigma^2$.
 - (iii) Let X be a standard normal $N(0, 1)$ random variable. Find the distribution Q of X^2 , and compute Q^{*2} . [Hint: By the $\pi - \lambda$ theorem it is sufficient to compute a pdf for $Q(-\infty, x], x \in \mathbb{R}$ to determine $Q(A)$ for all Borel sets A . Also, $\int_0^1 u^{-\frac{1}{2}}(1-u)^{-\frac{1}{2}} du = \pi$.]
9. Let X_1, X_2 be random maps on (Ω, \mathcal{F}, P) taking values in the measurable spaces $(S_1, \mathcal{S}_1), (S_2, \mathcal{S}_2)$, respectively. Show that the joint distribution of (X_1, X_2) on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$ is product measure if and only if $\sigma(X_1)$ and $\sigma(X_2)$ are independent σ -fields.
10. (i) Let V_1 take values ± 1 with probability $1/4$ each, and 0 with probability $1/2$. Let $V_2 = V_1^2$. Show that $\text{Cov}(V_1, V_2) = 0$, though V_1, V_2 are not independent.

- (ii) Show that random maps V_1, V_2 are independent if and only if $f(V_1)$ and $g(V_2)$ are uncorrelated for all pairs of real-valued Borel-measurable functions f, g such that $f(V_1), g(V_2) \in L^2$.
11. Suppose that X_1, X_2, \dots is a sequence of random variables on (Ω, \mathcal{F}, P) each having the same distribution $Q = P \circ X_n^{-1}$. (i) Show that if $\mathbb{E}|X_1| < \infty$ then $P(|X_n| > n \text{ i.o.}) = 0$. [Hint: First use (1.10) to get $\mathbb{E}|X_1| = \int_0^\infty P(|X_1| > x)dx$, and then apply Borel–Cantelli.] (ii) Assume that X_1, X_2, \dots are also independent with $\mathbb{E}|X_n| = \infty$. Show that $P(|X_n| > n \text{ i.o.}) = 1$.
 12. Suppose that Y_1, Y_2, \dots is an i.i.d. sequence of nonconstant random variables. Show that $\liminf_{n \rightarrow \infty} Y_n < \limsup_{n \rightarrow \infty} Y_n$ a.s. In particular $P(\lim_{n \rightarrow \infty} Y_n \text{ exists}) = 0$. [Hint: Use there must be numbers $a < b$ such that $P(Y_1 < a) > 0$ and $P(Y_1 > b) > 0$. Use Borel–Cantelli II.]
 13. [Percolation] Formulate and extend Example 2 by determining the critical probability for percolation on the rooted b -ary tree defined by $T = \cup_{j=0}^\infty \{1, 2, \dots, b\}^j$ for a natural number $b \geq 3$.
 14. Suppose that Y_1, Y_2, \dots is a sequence of i.i.d. positive random variables with $P(Y_1 = 1) < 1$ and $\mathbb{E}Y_1 = 1$. (a) Show that $X_n = Y_1 Y_2 \dots Y_n \rightarrow 0$ a.s. as $n \rightarrow \infty$; [Hint: Consider $\mathbb{E}X_n^t$ for fixed $0 < t < 1$, and apply Chebyshev inequality, Jensen inequality and Borel–Cantelli I.] (b) Is the sequence X_1, X_2, \dots uniformly integrable?
 15. (i) Consider three independent tosses of a balanced coin and let A_i denote the event that the outcomes of the i th and $(i+1)$ st tosses match, for $i = 1, 2$. Let A_3 be the event that the outcomes of the third and first match. Show that A_1, A_2, A_3 are pairwise independent but not independent. (ii) Show that $A_1, \dots, A_n, A_j \in \mathcal{S}_j, 1 \leq j \leq n$, are independent events if and only if the 2^n equations $P(C_1 \cap \dots \cap C_n) = \prod_{j=1}^n P(C_j)$ hold for all choices of $C_j = A_j$ or $C_j = A_j^c, 1 \leq j \leq n$. (iii) Show that $A_1, \dots, A_n, A_j \in \mathcal{S}_j, 1 \leq i \leq n$, are independent events if and only if the $2^n - n - 1$ equations $P(A_{j_1} \cap \dots \cap A_{j_m}) = \prod_{i=1}^m P(A_{j_i}), m = 2, \dots, n$, hold.
 16. Suppose that A_1, A_2, \dots is a sequence of independent events, each having the same probability $p = P(A_n) > 0$ for each $n = 1, 2, \dots$. Show that the event $[A_n \text{ e.o.}] := \cup_{n=1}^\infty A_n$ has probability one, where *e.o.* denotes *eventually occurs*.
 17. Let (Ω, \mathcal{F}, P) be an arbitrary probability space and suppose A_1, A_2, \dots is a sequence of *independent* events in \mathcal{F} with $P(A_n) < 1, \forall n$. Suppose $P(\cup_{n=1}^\infty A_n) = 1$. (i) Show that $P(A_n \text{ i.o.}) = 1$. (ii) Give an example of independent events for which $P(\cup_{n=1}^\infty A_n) = 1$ but $P(A_n \text{ i.o.}) < 1$.
 18. Let (Ω, \mathcal{F}, P) be an arbitrary probability space and suppose $\{A_n\}_{n=1}^\infty$ is a sequence of *independent* events in \mathcal{F} such that $\sum_{n=1}^\infty P(A_n) \geq 2$. Let E denote the event that none of the A_n 's occur for $n \geq 1$. (i) Show that $E \in \mathcal{F}$. (ii) Show that $P(E) \leq \frac{1}{e^2}$. [Hint: $1 - x \leq e^{-x}, x \geq 0$.]
 19. Suppose that X_1, X_2, \dots is an i.i.d. sequence of Bernoulli 0 or 1-valued random variables with $P(X_n = 1) = p, P(X_n = 0) = q = 1 - p$. Fix $r \geq 1$ and let $R_n := [X_n = 1, X_{n+1} = 1, \dots, X_{n+r-1} = 1]$ be the event of a run of 1's of length at least r starting from n .

- (i) Show that $P(R_n \text{ i.o.}) = 1$ if $0 < p \leq 1$.
- (ii) Suppose r is allowed to grow with n , say $r_n = [\theta \log_2 n]$ in defining the event R_n ; here $[x]$ denotes the largest integer not exceeding x . In the case of a balanced coin ($p = 1/2$), show that if $\theta > 1$ then $P(R_n \text{ i.o.}) = 0$, [Hint: Borel–Cantelli Lemma 1], and if $0 < \theta \leq 1$ then $P(R_n \text{ i.o.}) = 1$. [Hint: Consider a subsequence $R_{n_k} = [X_{n_k} = 1, \dots, X_{n_k+r_{n_k}-1} = 1]$ with n_1 sufficiently large that $\theta \log_2 n_1 > 1$, and $n_{k+1} = n_k + r_{n_k}$, $k \geq 1$. Compare $\sum_{k=1}^{\infty} n_k^{-\theta} \equiv \sum_{k=1}^{\infty} \frac{n_k^{-\theta}}{n_{k+1} - n_k} (n_{k+1} - n_k)$ to an integral $\int_{n_1}^{\infty} f(x) dx$ for an appropriately selected function f .]
20. Show that in the case $\mathbb{E} \log^+ \varepsilon_1 = \infty$, the formal power series of Example 5 is almost surely divergent for any $x \neq 0$. [Hint: Consider $0 < |x| < 1$, $|x| \geq 1$, separately.]
21. Let \mathcal{C} denote the collection of functions of the form $\sum_{i=1}^n f_i(u)g_i(v)$, $(u, v) \in S_1 \times S_2$, where f_i, g_i , $1 \leq i \leq n$, are bounded Borel-measurable functions on the probability spaces $(S_1, \mathcal{S}_1, Q_1)$ and $(S_2, \mathcal{S}_2, Q_2)$, respectively. Show that \mathcal{C} is dense in $L^1(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2, Q_1 \times Q_2)$. [Hint: Use the method of approximation by simple functions.]
22. Suppose that X, Y are independent random variables on (Ω, \mathcal{F}, P) . Assume that there is a number $a < 1$ such that $P(X \leq a) = 1$. Also assume that Y is exponentially distributed with mean one. Calculate $\mathbb{E}[e^{XY} | \sigma(X)]$. [Hint: Use the substitution property.]
23. Suppose that (X, Y) is uniformly distributed on the unit disk $D = \{(x, y) : x^2 + y^2 \leq 1\}$, i.e., has constant pdf on D . (i) Calculate the (marginal) distribution of X . (ii) Calculate the conditional distribution of Y given $\sigma(X)$. (iii) Calculate $\mathbb{E}(Y^2 | \sigma(X))$.
24. (i) Give a proof of (2.13) using the second, and therefore the first, definition of conditional expectation. [Hint: The only measurable random variables with respect to $\{\Omega, \emptyset, B, B^c\}$ are those of the form $c\mathbf{1}_B + d\mathbf{1}_{B^c}$, for $c, d \in \mathbb{R}$.] (ii) Prove (2.24), (2.26).
25. Suppose that X, N are independent random variables with standard normal distribution. Let $Y = X + bN$; i.e., X with an independent additive noise term bN . Calculate $\mathbb{E}(X | \sigma(Y))$. [Hint: Compute the conditional pdf of the regular conditional distribution of X given $\sigma(Y)$ by first computing the joint pdf of (X, Y) . For this, either view (X, Y) as a function (linear transformation) of independent pair (X, N) , or notice that the conditional pdf of Y given $\sigma(X)$ follows immediately from the substitution property. Also, the marginal of X is the convolution of two normal distribution and hence, normal. It is sufficient to compute the mean and variance of X . From here obtain the joint density as the product of the conditional density of Y given $\sigma(X)$ and the marginal of X .]
26. Suppose that U is uniformly distributed on $[0, 1]$ and $V = U(1 - U)$. Determine $\mathbb{E}(U | \sigma(V))$. Does the answer depend on the symmetry of the uniform distribution on $[0, 1]$?
27. Suppose X is a real-valued random variable with symmetric distribution Q about zero; i.e., X and $-X$ have the same distribution. (i) Compute $P(X > 0 | \sigma(|X|))$. (ii) Determine the (regular) conditional distribution of X given $\sigma(|X|)$. [Hint:

- Keep in mind that $\int_{\mathbb{R}} f(x)\delta_{\{a\}}(dx) = f(a)$ for any $a \in \mathbb{R}$, and consider $\mathbb{E}f(X)g(|X|)$ for $\mathbb{E}|f(X)| < \infty$ and bounded, Borel measurable g . Partition the integral by $1 = \mathbf{1}_{(0,\infty)}(|X|) + \mathbf{1}_{\{0\}}(|X|)$.
28. Suppose that X_1, X_2, \dots is an i.i.d. sequence of square-integrable random variables with mean μ and variance $\sigma^2 > 0$, and N is a nonnegative integer-valued random variable, independent of X_1, X_2, \dots . Let $S = \sum_{j=1}^N X_j$, with the convention $S = 0$ on the event $[N = 0]$. Compute the mean and variance of S . [Hint: Condition on N .]
29. (a) Let X_1, \dots, X_n be an i.i.d. sequence of random variables on (Ω, \mathcal{F}, P) and let $S_n = X_1 + \dots + X_n$. Assume $\mathbb{E}|X_1| < \infty$. Show that $\mathbb{E}(X_j|\sigma(S_n)) = \mathbb{E}(X_1|\sigma(S_n))$. [Hint: Use Fubini–Tonelli.] Calculate $\mathbb{E}(X_j|\sigma(S_n))$. [Hint: Add up and use properties of conditional expectation.] (b) Generalize (a) to the case in which the distribution of (X_1, \dots, X_n) is invariant under permutations of the indices, i.e., the distribution of $(X_{\pi_1}, \dots, X_{\pi_n})$ is the same for all permutations π of $(1, 2, \dots, n)$.
30. Suppose that (X, Y) is distributed on $[0, 1] \times [0, 1]$ according to the pdf $f(x, y) = 4xy$, $0 \leq x, y \leq 1$. Determine $\mathbb{E}[X|\sigma(X + Y)]$.
31. Suppose that Y_1, \dots, Y_n are i.i.d. exponentially distributed with mean one. Let $S_n = \sum_{j=1}^n Y_j$.
- Calculate $\mathbb{E}(Y_1^2|S_n)$. [Hint: In view of part (iii) of this problem, calculate the joint pdf of $(Y_1, Y_2 + \dots + Y_n)$ and then that of (Y_1, S_n) by a change of variable under the linear transformation $(y, s) \mapsto (y, y + s)$, for an arbitrary distribution of Y_1 having pdf $g(y)$. Namely, $g(y)g^{*(n-1)}(s - y)/g^{*n}(s)$.]
 - Calculate $\mathbb{E}(Y_1 Y_2|S_n)$. [Hint: Consider $S_n^2 = \mathbb{E}(S_n^2|S_n)$ along with the previous exercise.]
 - Make the above calculations in the case that Y_1, Y_2, \dots, Y_n are i.i.d. with standard normal distributions.
32. [Conditional Chebyshev-type] For $X \in L^p$, $p \geq 1$, prove for $\lambda > 0$, $P(|X| > \lambda|\mathcal{G}) \leq \mathbb{E}(|X|^p|\mathcal{G})/\lambda^p$ a.s.
33. [Conditional Cauchy–Schwarz] For $X, Y \in L^2$ show that $|\mathbb{E}(XY|\mathcal{G})|^2 \leq \mathbb{E}(X^2|\mathcal{G})\mathbb{E}(Y^2|\mathcal{G})$.
34. Let Y be an exponentially distributed random variable on (Ω, \mathcal{F}, P) . Fix $a > 0$.
- Calculate $\mathbb{E}(Y|\sigma(Y \wedge a))$, where $Y \wedge a := \min\{Y, a\}$. [Hint: $[Y < a] = [Y \wedge a < a]$. Let g be a bounded Borel-measurable function and either make and verify an intuitive guess for $\mathbb{E}(Y|\sigma(Y \wedge a))$ (based on “lack of memory” of the exponential distribution), or calculate $\mathbb{E}(Yg(Y \wedge a))$ by integration by parts.]
 - Determine the regular conditional distribution of Y given $\sigma(Y \wedge a)$.
 - Repeat this problem for Y having arbitrary distribution Q on $[0, \infty)$, and a such that $0 < P(Y \leq a) < 1$.
35. Let U, V be independent random maps with values in measurable spaces (S_1, \mathcal{S}_1) and (S_2, \mathcal{S}_2) , respectively. Let $\varphi(u, v)$ be a measurable map on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$ into a measurable space (S, \mathcal{S}) . Show that a regular conditional distribution of

$\varphi(U, V)$, given $\sigma(V)$, is given by Q_V , where Q_v is the distribution of $\varphi(U, v)$. [Hint: Use the Fubini–Tonelli theorem or Theorem 2.10(1).]

36. Prove the Markov property for k -dimensional random walks with $k \geq 2$.
37. Let $\Omega = \{0, 1\}^\infty$ be the space of infinite binary 0–1 sequences, and let \mathcal{F}_0 denote the field of finite unions of sets of the form $A_n(\varepsilon_1, \dots, \varepsilon_n) = \{\omega = (\omega_1, \omega_2, \dots) \in \Omega : \omega_1 = \varepsilon_1, \dots, \omega_n = \varepsilon_n\}$ for arbitrary $\varepsilon_i \in \{0, 1\}$, $1 \leq i \leq n$, $n \geq 1$. Fix $p \in [0, 1]$ and define $P_p(A_n(\varepsilon_1, \dots, \varepsilon_n)) = p^{\sum_{i=1}^n \varepsilon_i} (1-p)^{n-\sum_{i=1}^n \varepsilon_i}$. (i) Show that the natural finitely additive extension of P_p to \mathcal{F}_0 defines a measure on the field \mathcal{F}_0 . [Hint: By Tychonoff's theorem from topology, the set Ω is compact for the product topology, see Appendix B. Check that sets $C \in \mathcal{F}_0$ are both open and closed for the product topology, so that by compactness, any countable disjoint union belonging to \mathcal{F}_0 must be a finite union.] (ii) Show that P_p has a unique extension to $\sigma(\mathcal{F}_0)$. This probability P_p defines the infinite product probability, also denoted by $(p\delta_{\{1\}} + (1-p)\delta_{\{0\}})^\infty$. [Hint: Apply the Carathéodory extension theorem.] (iii) Show that the coordinate projections $X_n(\omega) = \omega_n$, $\omega = (\omega_1, \omega_2, \dots) \in \Omega$, $n \geq 1$, define an i.i.d. sequence of (coin tossing) Bernoulli 0 or 1-valued random variables.
38. Prove that the set \tilde{B} in the proof of Proposition 2.5 belongs to $\mathcal{B}(\mathbb{R}^k)$, and the function g , there, is Borel-measurable.
39. (i) Prove that the set of all simple function on (Ω, \mathcal{F}, P) is dense in L^p , $\forall p \geq 1$. (ii) Prove that L^p is dense in L^r (in L^r – norm) for $p > r \geq 1$ (in L^r -norm)



<http://www.springer.com/978-3-319-47972-9>

A Basic Course in Probability Theory

Bhattacharya, R.; Waymire, E.C.

2016, XII, 265 p., Softcover

ISBN: 978-3-319-47972-9