

Chapter 2

Essentials of Probability Theory

2.1 Introduction

Many properties of physical systems and/or actions of these systems are uncertain, so that the behavior of these systems cannot be forecasted in a precise deterministic manner; it can only be described probabilistically. For example, the weather person tells us the chance of rain tomorrow. Engineers calculate the likelihood that a particular mechanical system will perform according to specified standards.

Suppose a relevant output of a physical system depends on a finite number of uncertain parameters. For weather forecasting, these parameters relate to the current meteorologic conditions and atmospheric processes. For aircraft design, these parameters include material properties, state of electronic components, and flight patterns. Our objective is to calculate the probability that an output of interest has specified features, such as the chance of rain tomorrow for the weather person or the adequate aircraft performance for the aircraft engineer. This chapter provides the framework for calculating these types of probabilities.

2.2 Probability Space

A probability space (Ω, \mathcal{F}, P) consists of a sample space Ω , a σ -field \mathcal{F} , and a probability measure P . Sample spaces, σ -fields, and probability measures are defined and illustrated by examples.

2.2.1 Sample Space

Definition 2.1 A set Ω collecting the outcomes of a particular experiment is called sample space.

For example, $\Omega = \{\text{head, tail}\}$, $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $\Omega = [a, b] \subset [0, \infty)$ are sample spaces for the experiments of tossing a coin, rolling a dice, and measuring daily rainfall amounts in Ithaca, NY. The first two sample spaces and the last sample space have, respectively, a finite and an uncountable number of elements. Sample spaces can be finite, countable, or uncountable.

2.2.2 σ -Field

Consider a game in which one wins \$10 and loses \$5 for outcomes of a die rolling experiment in $\{1, 2\}$ and $\{3, 4, 5, 6\}$, respectively. The particular outcome $\{\omega\}$ is not relevant. The relevant information is contained in $\mathcal{A} = \{\{1, 2\}, \{3, 4, 5, 6\}\}$ since one wins \$10 if $\omega \in \{1, 2\}$ and loses \$5 if $\omega \in \{3, 4, 5, 6\}$. σ -fields are subsets of Ω that are relevant for a particular objective. The σ -field for the game considered here includes \mathcal{A} .

Definition 2.2 A collection \mathcal{F} of subsets of Ω is said to be a σ -field on a sample space Ω if (1) $\emptyset \in \mathcal{F}$, (2) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$, and (3) $A_i \in \mathcal{F}$, $i \in I$, implies $\bigcup_{i \in I} A_i \in \mathcal{F}$, where I is a countable set. The members of \mathcal{F} are called events, or \mathcal{F} -measurable subsets of Ω , or just measurable subsets of Ω . The pair (Ω, \mathcal{F}) is said to be a measurable space.

The first and the third conditions in the previous definition can be replaced with $\Omega \in \mathcal{F}$ and $\bigcap_{i \in I} A_i^c \in \mathcal{F}$ by using condition (2) and De Morgan's formula. Also note that the last two conditions in the definition of \mathcal{F} are consistent with our intuition, which suggests that A^c is observed if A is not and that $\bigcup_{i \in I} A_i$ is observed if a subset of $\{A_i, i \in I\}$ is.

Example 2.1 The σ -field \mathcal{F} associated with the game in which one wins \$10 and loses \$5 for outcomes of a die rolling experiment in $\{1, 2\}$ and $\{3, 4, 5, 6\}$ is $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \Omega\}$, where $\Omega = \{1, 2, 3, 4, 5, 6\}$. If the game is modified such that one wins \$10 if $\omega \in \{1, 2\}$ and loses \$3, \$4, \$5, and \$6 for outcomes $\omega = 3, 4, 5$, and 6, respectively, \mathcal{F} is too coarse to capture all relevant events; it needs to be refined to include the events $\{1, 2\}$, $\{3\}$, $\{4\}$, $\{5\}$, and $\{6\}$, and complements and unions of these events. \diamond

Example 2.2 Atoms are the finest members of a σ -field \mathcal{F} , that is, $A \in \mathcal{F}$ is an atom of \mathcal{F} if any event $B \in \mathcal{F}$ included in A is either \emptyset or A . The sets $\{1, 2\}$ and $\{3, 4, 5, 6\}$ are atoms of the first σ -field and the sets $\{1, 2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$ are atoms of the second σ -field (Example 2.1). \diamond

Definition 2.3 The σ -field generated by a collection of subsets \mathcal{A} of Ω is

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{G} \supseteq \mathcal{A}} \mathcal{G}, \quad \text{where } \{\mathcal{G}\} \text{ are } \sigma\text{-fields on } \Omega. \quad (2.1)$$

There is no σ -field smaller than $\sigma(\mathcal{A})$ that includes \mathcal{A} .

Example 2.3 Let S be a metric space and let \mathcal{T} be the topology induced on S by its metric (Sect. B.1.1). The Borel σ -field on S is the σ -field $\mathcal{S} = \sigma(\mathcal{T})$ generated by \mathcal{T} . The members of $\sigma(\mathcal{T})$ are called Borel sets. The Borel σ -fields on $S = \mathbb{R}^d$, $d > 1$, and $S = \mathbb{R}$ are generated by the intervals in these spaces and are denoted by $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$ and $\mathcal{B}(\mathbb{R}) = \mathcal{B}^1 = \mathcal{B}$, respectively. The Borel sets on \mathbb{R} can be generated by open, semi-open, finite, semi-finite, and other intervals of the real line ([11], Sect. 1.7). \diamond

2.2.3 Probability Measure

We define measures and probability measures, review properties of probability measures useful for applications, and use conditional probabilities to introduce the law of total probability and Bayes theorem.

Definition 2.4 Let (Ω, \mathcal{F}) be a measurable space. A set function $\mu : \mathcal{F} \rightarrow [0, \infty]$ such that (1) $\mu(\emptyset) = 0$ and (2) $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for mutually disjoint sets $A_n \in \mathcal{F}$ is called a measure on (Ω, \mathcal{F}) . If $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}^d)$, then μ is said to be a Borel measure. The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

Definition 2.5 If μ has finite total mass, that is, $\mu(\Omega) < \infty$, the measure μ is said to be finite. In this case, it can be normalized to take values in $[0, 1]$. Any measure with unit total mass is called a probability measure and is denoted by P . The triple (Ω, \mathcal{F}, P) is called a probability space.

Definition 2.6 A measure μ on a measurable space (Ω, \mathcal{F}) is said to be σ -finite if there exists a countable, pairwise disjoint sequence of measurable sets $\{A_n\}$, that is, $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, such that $\mu(A_n) < \infty$, $n = 1, 2, \dots$, and $\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap A_n)$ for every $A \in \mathcal{F}$ ([5], Proposition 13).

The Lebesgue measure λ on $(\mathbb{R}, \mathcal{B})$ is σ -finite since there exists a pairwise disjoint sequence $A_n = [n, n+1)$, $n \in \mathbb{Z}$, such that $\lambda(A_n) < \infty$ and $\lambda(A) = \sum_{n=1}^{\infty} \lambda(A \cap A_n)$ for every $A \in \mathcal{B}$. Similar arguments show that the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is σ -finite. However, the Lebesgue measure is not finite.

Definition 2.7 Let (Ω, \mathcal{F}, P) be a probability space. A set $N \in \mathcal{F}$ such that $P(N)=0$ is called a null set. A property valid on $\Omega \setminus N$ is said to hold almost everywhere (a.e.), almost surely (a.s.), for almost every ω , or with probability one (w.p.1.).

Definition 2.8 A probability space (Ω, \mathcal{F}, P) is complete if $A \subset B$ such that $B \in \mathcal{F}$ and $P(B) = 0$, then $A \in \mathcal{F}$, which implies $P(A) = 0$, by the second condition in Definition 2.2.

It is assumed throughout the book that the probability spaces are complete. The assumption is not restrictive since for any probability space (Ω, \mathcal{F}, P) there exists a complete space $(\Omega, \overline{\mathcal{F}}, \overline{P})$ such that $\mathcal{F} \subseteq \overline{\mathcal{F}}$ and $P = \overline{P}$ on \mathcal{F} ([4], Theorem

2.2.5). The completion of measures is rather simple ([1], p. 4). For example, set $\mathcal{N} = \{B \subseteq \Omega : \exists N \in \mathcal{F} \text{ with } P(N) = 0 \text{ and } B \subset N\}$, $\overline{\mathcal{F}} = \{A \cup B : A \in \mathcal{F}, B \in \mathcal{N}\}$, and $\overline{P}(A \cup B) = P(A)$, where $A \in \mathcal{F}$ and $B \in \mathcal{N}$. Then \overline{P} is a probability measure on $(\Omega, \overline{\mathcal{F}})$ and $\overline{\mathcal{F}}$ is a σ -field.

The following properties of the probability measure P are useful for calculations, and result directly from its definition.

$$\begin{aligned}
 P(A) &= 1 - P(A^c), \quad \text{for } A \in \mathcal{F}, \\
 P(A) &\leq P(B), \quad \text{for } A \subseteq B, \quad A, B \in \mathcal{F}, \\
 P(\cup_{i=1}^{\infty} A_i) &\leq \sum_{i=1}^{\infty} P(A_i), \quad \text{for } A_i \in \mathcal{F}, \\
 P(A \cup B) &= P(A) + P(B) - P(A \cap B), \quad \text{for } A, B \in \mathcal{F}, \text{ and} \\
 P(B) &= \sum_{i=1}^n P(B \cap A_i), \quad \text{for } A_i, B \in \mathcal{F} \text{ and } A_1, \dots, A_n \text{ a partition of } \Omega.
 \end{aligned} \tag{2.2}$$

The probability measure also satisfies the inclusion–exclusion formula

$$\begin{aligned}
 P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{i=2}^n \sum_{j=1}^{i-1} P(A_i \cap A_j) \\
 &\quad + \sum_{i=3}^n \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(\cap_{q=1}^n A_q).
 \end{aligned} \tag{2.3}$$

Theorem 2.1 *If (Ω, \mathcal{F}, P) is a probability space and $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, are events on this space, then*

$$P(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n). \tag{2.4}$$

Proof Set $B_1 = A_1$ and $B_n = A_n \setminus (\cup_{i=1}^{n-1} A_i)$, $n > 1$. The sets $\{B_n\}$ are disjoint events with the properties $B_n \subseteq A_n$, and $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$. Hence, $P(\cup_{n=1}^{\infty} A_n) = P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$. \blacktriangle

Definition 2.9 Let (Ω, \mathcal{F}, P) be a probability space and $B \in \mathcal{F}$ an event such that $P(B) > 0$. The probability of $A \in \mathcal{F}$ conditional on B is

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{F}. \tag{2.5}$$

The definition is meaningful in the sense that $P(\cdot | B)$ is a probability measure.

Suppose an experiment is performed in which B either occurs or does not occur, and that $P(B), P(B^c) > 0$. If B or B^c is observed, the conditional probability of A

is $P(A | B)$ or $P(A | B^c)$, respectively. Hence, the conditional probability of A is equal to $P(A \cap B)/P(B)$ and $P(A \cap B^c)/P(B^c)$ with probabilities $P(B)$ and $P(B^c)$. This remark will be revisited later in this chapter (Example 2.56).

The following two formulas involving conditional probabilities are very useful for applications. Let (Ω, \mathcal{F}, P) be a probability space, $A_i \in \mathcal{F}$, $i = 1, \dots, n$, a partition of Ω , that is, $\Omega = \cup_{i=1}^n A_i$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. If $P(A_i) > 0$, and $B \in \mathcal{F}$, then

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B | A_i)P(A_i) \quad (\text{Law of total probability})$$

$$P(A_j | B) = \frac{P(A_j)P(B | A_j)}{P(B)} = \frac{P(A_j)P(B | A_j)}{\sum_{i=1}^n P(A_i)P(B | A_i)} \quad (\text{Bayes' formula}). \quad (2.6)$$

In the Bayesian framework, the probabilities $P(A_j)$ and $P(A_j | B)$ are referred to as the prior and the posterior probabilities of A_j .

Example 2.4 Let B denote the event that a system performs satisfactorily, and let $\{A_i, i = 1, 2\}$ be events partitioning Ω . Suppose $P(A_1) = 0.8$, $P(A_2) = 0.2$, $P(B | A_1) = 0.9$, and $P(B | A_2) = 0.7$. The probability that the system performs satisfactorily is $P(B) = (0.9)(0.8) + (0.7)(0.2) = 0.86$. \diamond

Theorem 2.2 *The inequalities*

$$P\left(\cup_{i=1}^n A_i\right) \leq P_u = \sum_{i=1}^n P(A_i) - \sum_{i=2}^n \max_{j=1, \dots, i-1} P(A_j \cap A_i) \quad \text{and}$$

$$P\left(\cup_{i=1}^n A_i\right) \geq P_l = P(A_1) + \sum_{i=2}^n \max\left(0, P(A_i) - \sum_{j=1}^{i-1} P(A_j \cap A_i)\right), \quad (2.7)$$

hold for $A_i \in \mathcal{F}$ arbitrary.

Proof Set $B_i = A_i^c$, $i = 1, \dots, n$, and $G = \cap_{i=1}^{n-1} B_i$, and note that $P\left(\cap_{i=1}^n B_i\right) = P(G \cap B_n) = P(G) - P(G \cap A_n)$. Repeated application of this formula gives $P\left(\cap_{i=1}^n B_i\right) = P(B_1) - \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} \cap A_k)$ so that

$$\begin{aligned} P\left(\cup_{i=1}^n A_i\right) &= 1 - P\left(\cap_{i=1}^n B_i\right) = P(A_1) + \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} \cap A_k) \\ &= P(A_1) + \sum_{k=2}^n P(B_1 \cap \dots \cap B_{k-1} | A_k)P(A_k). \end{aligned}$$

Since $P(B_1 \cap \dots \cap B_{k-1} | A_k) \leq P(B_j | A_k) = 1 - P(A_j | A_k)$, $j = 1, \dots, k-1$, and $P(B_1 \cap \dots \cap B_{k-1} | A_k) = 1 - P(A_1 \cup \dots \cup A_{k-1} | A_k) \geq 1 - \sum_{j=1}^{k-1} P(A_j | A_k)$ hold, we have the bounds in (2.7). \blacktriangle

The bounds in (2.7) are relatively simple to calculate since they involve the probabilities of the events A_i and $A_i \cap A_j$, rather than probabilities of intersections of multiple events as in (2.3). Note that $P(\cup_{i=1}^n A_i)$ can be interpreted as the probability of failure for a physical system with failure modes A_1, \dots, A_n occurring with probabilities $P(A_i)$, $i = 1, \dots, n$.

2.2.4 Construction of Probability Spaces

In most applications a random experiment rather than a probability space is specified, so that we need to construct a probability space based on the available information. We present three methods for constructing probability spaces and illustrate them by examples.

2.2.4.1 Countable Sample Space

Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a countable sample space and let $\{p_i \geq 0\}$ such that $\sum_{i=1}^{\infty} p_i = 1$. Take \mathcal{F} to be the collection of all subsets of Ω , that is, the power set of Ω . The set function $P: \mathcal{F} \rightarrow [0, 1]$ defined by

$$P(A) = \sum_{\omega_i \in A} p_i, \quad A \in \mathcal{F}, \quad (2.8)$$

is a probability measure on (Ω, \mathcal{F}) since it is countably additive and has the properties $P(\emptyset) = 0$ and $P(\Omega) = 1$.

Example 2.5 Suppose service life has been recorded for n nominally identical devices. Let $0 = \tau_0 < \tau_1 < \dots < \tau_{m-1}$, let $[0, \tau_1), [\tau_1, \tau_2), \dots, [\tau_{m-1}, \infty)$ be a partition of $[0, \infty)$, and denoted by $n_i \leq n$ the number of devices with measured service life in the range $[\tau_{i-1}, \tau_i)$, $i = 1, \dots, m$, with the notation $\tau_m = \infty$. The members of the probability space (Ω, \mathcal{F}, P) describing this experiment are $\Omega = \{\omega_1, \dots, \omega_m\}$, \mathcal{F} is the power set of Ω , and $P(\{\omega_i\}) \simeq n_i/n$. \diamond

2.2.4.2 Product Probability Space

Let $(\Omega_k, \mathcal{F}_k, P_k)$, $k = 1, 2$, be probability spaces. The elements of the product probability space (Ω, \mathcal{F}, P) are

$$\begin{aligned} \Omega &= \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_k \in \Omega_k, k = 1, 2\} \\ \mathcal{F} &= \mathcal{F}_1 \times \mathcal{F}_2 = \sigma(\mathcal{R}), \quad \text{where } \mathcal{R} = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\} \\ P &= P_1 \times P_2, \quad \text{where } P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \text{ for } A_1 \in \mathcal{F}_1 \text{ and } A_2 \in \mathcal{F}_2, \end{aligned} \quad (2.9)$$

and are called the product sample space, σ -field, and probability measure.

The product sample space Ω contains the joint outcomes of, for example, two experiments. The σ -field \mathcal{F} can also be obtained from $\mathcal{F} = \sigma(\mathcal{R}) = \sigma(\mathcal{G}_1, \mathcal{G}_2)$

where $\mathcal{G}_1 = \{A_1 \times \Omega_2 : A_1 \in \mathcal{F}_1\}$ and $\mathcal{G}_2 = \{\Omega_1 \times A_2 : A_2 \in \mathcal{F}_2\}$ since \mathcal{G}_1 and \mathcal{G}_2 are σ -fields on Ω that are included in \mathcal{F} and every member of \mathcal{R} is the intersection of sets from \mathcal{G}_1 and \mathcal{G}_2 . The construction of the product probability measure is less simple. It can be shown that there exists a unique probability P on (Ω, \mathcal{F}) that satisfies (2.9) ([4], Theorem 3.3.5).

The formulas in (2.9) can be generalized to define product probability spaces of $n \geq 3$ probability spaces. The definition extends directly to the case in which n is infinity. If the probability spaces $(\Omega_k, \mathcal{F}_k, P_k)$ are identical, the product sample space, σ -field, and probability are denoted by Ω_1^n , \mathcal{F}_1^n , and P_1^n .

Example 2.6 Let $(\Omega_k, \mathcal{F}_k, P_k)$, $k = 1, 2$, be probability spaces associated with the experiment of rolling two dice, so that $\Omega_k = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F}_k =$ all subsets of Ω_k , and $P_k(\{i\}) = 1/6$, $i = 1, \dots, 6$. The product sample space $\Omega = \Omega_1 \times \Omega_2 = \{\omega = (i, j), i, j = 1, \dots, 6\}$ includes all outcomes of the experiment. The product σ -field \mathcal{F} consists of all subsets of Ω since the members of \mathcal{R} are (i, j) , $\cup_{i \in I_1}(i, j)$, $\cup_{j \in I_2}(i, j)$, $\cup_{i \in I_1, j \in I_2}(i, j)$, where $I_1, I_2 \subseteq \{1, 2, 3, 4, 5, 6\}$. The product probability measure is $P(\{\omega\}) = 1/36$, $\omega \in \Omega$, since the outcomes of the experiment are equally likely. \diamond

Example 2.7 Consider a facility that fails under wind speeds exceeding a critical value v_{cr} . Let $p = P(A)$, where $A = \{V > v_{cr}\}$ and $V \geq 0$ denotes the maximum yearly wind speed at the facility site. The members of the probability space (Ω, \mathcal{F}, P) describing this problem are $\Omega = \{A, A^c\}$, $\mathcal{F} =$ all parts of Ω , and $P(A) = p$. To evaluate the probability that the facility performs satisfactorily during its design life of n years, we need to construct a new probability space corresponding to an “experiment” consisting of n independent repetitions of the yearly experiment. The construction resembles the experiment of tossing a loaded coin n times with sides $\{1\}$ and $\{0\}$ corresponding to events A and A^c ([11], p. 41).

The members of the product probability space corresponding to a n year time horizon are $\Omega^n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i = 0 \text{ or } 1\}$, $\mathcal{F}^n =$ all subsets of Ω^n , and $P^n(B) = \sum_{\omega \in B} p^{n_\omega} q^{n-n_\omega}$, where $B \in \mathcal{F}^n$, $q = 1 - p$, and $n_\omega = \sum_{i=1}^n \omega_i$ gives the numbers of 1's in $\omega = (\omega_1, \dots, \omega_n)$. \diamond

2.2.4.3 Extension of Probability Measure

Let Ω be a sample space associated with an experiment and let \mathcal{C} be a collection of subsets of Ω . If \mathcal{C} is a field on Ω and R is a real-valued, positive, and countably additive function defined on \mathcal{C} such that $R(\Omega) = 1$, then there exists a unique probability P on $\mathcal{F} = \sigma(\mathcal{C})$ such that $P(A) = R(A)$ for each $A \in \mathcal{C}$, that is, the restriction of P to \mathcal{C} is equal to R ([5], Theorem 14, p. 94). A field is a collection of sets satisfying the conditions of Definition 2.2 with I finite.

Example 2.8 Let $\Omega = \mathbb{R}$ and let \mathcal{C} consist of the empty set and the collection of all finite unions of intervals of the type $(a, b]$ for $a < b$, $(-\infty, a]$, (a, ∞) , and $(-\infty, \infty)$. Let $F: \mathbb{R} \rightarrow [0, 1]$ be a continuous increasing function such that

$\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$, and define $R: \mathcal{C} \rightarrow [0, 1]$ by $R((a, b]) = F(b) - F(a)$, $R((-\infty, a]) = F(a)$, $R((a, \infty)) = 1 - F(a)$, and $R((-\infty, \infty)) = 1$. For example, $R((a, \infty))$ may represent the probability that the strength of a particular material exceeds a , so that it can be estimated from experiments. Since \mathcal{C} is a field, the set function R can be extended uniquely to a probability measure on $(\mathbb{R}, \mathcal{B} = \sigma(\mathcal{C}))$ ([5], Proposition 9, p. 90, and Theorem 14, p. 94). \diamond

2.3 Measurable Functions and Random Elements

Consider two measurable spaces (Ω, \mathcal{F}) and (Ψ, \mathcal{G}) and a mapping $h: \Omega \rightarrow \Psi$.

Definition 2.10 The mapping $h: \Omega \rightarrow \Psi$ is said to be measurable from (Ω, \mathcal{F}) to (Ψ, \mathcal{G}) or $(\mathcal{F}, \mathcal{G})$ -measurable if $h^{-1}(\mathcal{G}) \subseteq \mathcal{F}$, that is, $h^{-1}(B) = \{\omega \in \Omega : h(\omega) \in B\} \in \mathcal{F}$ for all $B \in \mathcal{G}$. This property of h is denoted by $h \in \mathcal{F}/\mathcal{G}$ or just $h \in \mathcal{F}$ if there is no confusion about \mathcal{G} .

Example 2.9 Let $A \in \mathcal{F}$, $(\Psi, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, and 1_A the indicator function of A , that is, $1_A(\omega) = 1$ for $\omega \in A$ and $1_A(\omega) = 0$ for $\omega \notin A$. The function $1_A: \Omega \rightarrow \mathbb{R}$ is \mathcal{F}/\mathcal{B} -measurable since $1_A^{-1}(\{0\}) = A^c$ and $1_A^{-1}(\{1\}) = A$, so that $1_A^{-1}(\mathcal{B}) = \{\emptyset, \Omega, A, A^c\} \subseteq \mathcal{F}$. \diamond

Example 2.10 Let $\{A_i\}$, $i = 1, \dots, n$, be events partitioning Ω and $\{c_i\}$, $i = 1, \dots, n$, real constants. The function $h = \sum_{i=1}^n c_i 1_{A_i}$ is $(\mathcal{F}, \mathcal{G})$ -measurable, where $\Psi = \{c_1, \dots, c_n\}$ and \mathcal{G} is the power set of Ψ . The image of the members of Ψ in Ω can be obtained simply, for example, $h^{-1}(\{c_i, c_j\}) = A_i \cup A_j \in \mathcal{F}$. \diamond

Theorem 2.3 If $h: \mathbb{R}^d \rightarrow \mathbb{R}^q$ is continuous, it is also $(\mathcal{B}^d, \mathcal{B}^q)$ -measurable.

Proof Function h is measurable if and only if $h^{-1}(\mathcal{T}^q) \subseteq \mathcal{B}^d$ ([11], Proposition 3.2.1), where \mathcal{T}^r denotes the topology generated by the open balls in \mathbb{R}^r . Since h is continuous, we have $h^{-1}(\mathcal{T}^q) \subseteq \mathcal{T}^d \subseteq \mathcal{B}^d$ (Sect. B.1.1), where the latter inequality holds since \mathcal{B}^d is generated by \mathcal{T}^d . \blacktriangle

Definition 2.11 Let (Ω, \mathcal{F}, P) be a probability space and (S, \mathcal{S}) a measurable space, where S is a metric space and $\mathcal{S} = \sigma(\mathcal{T})$ denotes the σ -field generated by the topology \mathcal{T} induced on S by its metric (Example 2.3 and Sect. B.1). The mapping $X: \Omega \rightarrow S$ is a random element if it is $(\mathcal{F}, \mathcal{S})$ -measurable.

The element X in the above definition is a real-valued random variable, complex-valued random variable, random vector or an \mathbb{R}^d -valued random variable, \mathbb{C}^d -valued random variable, real-valued stochastic process with continuous samples defined on $[0, 1]$, or a real-valued random field with continuous samples defined on a subset D of $\mathbb{R}^{d'}$ if $S = \mathbb{R}$, $S = \mathbb{C}$, $S = \mathbb{R}^d$, $S = \mathbb{C}^d$, $S = C[0, 1]$, or $S = C(D)$, respectively. The image $X(\omega)$, $\omega \in \Omega$, of X is a number, vector in \mathbb{R}^d , vector in \mathbb{C}^d , real-valued continuous function defined on $[0, 1]$, or a real-valued continuous function defined on D if $S = \mathbb{R}$, $S = \mathbb{R}^d$, $S = \mathbb{C}^d$, $S = C[0, 1]$, or $S = C(D)$, respectively. For

example, $X(\omega) = (X_1(\omega), \dots, X_d(\omega)) \in \mathbb{R}^d$ is a d -dimensional vector for $S = \mathbb{R}^d$ and $X(\omega)(t)$, $t \in [0, 1]$, is a sample of a real-valued stochastic process defined on $[0, 1]$ for $S = C[0, 1]$. It is common to denote $X(\omega)(t)$ by $X(t, \omega)$, as will be seen in a subsequent section. For a fixed time $t \in [0, 1]$, $X(t, \omega)$ is a random variable.

Definition 2.12 Let (Ω, \mathcal{F}, P) be a probability space, (S, \mathcal{S}) a measurable space, and $X : \Omega \rightarrow S$ an $(\mathcal{F}, \mathcal{S})$ -measurable mapping. The probability measure induced by X or the distribution of X is the probability measure

$$Q(B) = P(X^{-1}(B)) = P \circ X^{-1}(B), \quad B \in \mathcal{S}, \quad (2.10)$$

on the measurable space (S, \mathcal{S}) .

Example 2.11 If $S = \mathbb{R}$ and $B = (-\infty, x]$, $x \in \mathbb{R}$, then $Q(B) = P(\{\omega : X(\omega) \leq x\}) = P(X \leq x)$ is called the probability distribution function or just the distribution of random variable X , and is typically denoted by $F(x) = P(X \leq x)$. If $S = \mathbb{R}^d$ and $B = \times_{i=1}^d (-\infty, x_i]$, $x_i \in \mathbb{R}$, then $Q(B) = P(\{\omega : X_i(\omega) \leq x_i, i = 1, \dots, d\}) = P(\cap_{i=1}^d \{X_i \leq x_i\})$ is called the joint distribution of random vector X . If $S = C[0, 1]$, $\{t_1, \dots, t_n\} \subset [0, 1]$, and $B = \times_{k=1}^n (-\infty, x_k]$, $x_k \in \mathbb{R}$, then $Q(B) = P(\{\omega : X(t_k, \omega) \leq x_k, k = 1, \dots, n\}) = P(\cap_{k=1}^n \{X(t_k) \leq x_k\})$ is called the finite dimensional distribution of order n of X . The finite dimensional distributions of a stochastic process X are the joint distributions of random vectors $(X(t_1), \dots, X(t_n))$ consisting of values of the process at times t_1, \dots, t_n . \diamond

Example 2.12 Let (Ω, \mathcal{F}, P) be a probability space, $A \in \mathcal{F}$, $S = \{0, 1\}$, and $1_A : \Omega \rightarrow S$ an indicator function. The function 1_A is $(\mathcal{F}, \mathcal{S})$ -measurable, where $\mathcal{S} = \{\emptyset, S, \{0\}, \{1\}\}$. The probability measure Q induced by 1_A on (S, \mathcal{S}) is $Q(\{0\}) = P(1_A^{-1}(\{0\})) = P(A^c)$ and $Q(\{1\}) = P(1_A^{-1}(\{1\})) = P(A)$. The distribution of random variable 1_A is $F(x) = P(A^c)1(x \geq 0) + P(A)1(x \geq 1)$. \diamond

Example 2.13 Let (Ω, \mathcal{F}, P) be a probability space and let $X : \Omega \rightarrow \mathbb{R}$ be $(\mathcal{F}, \mathcal{B})$ -measurable. Suppose the probability measure Q induced by P on $(\mathbb{R}, \mathcal{B})$ has the expression $Q((-\infty, x]) = P(X^{-1}((-\infty, x])) = \Phi((x - \mu)/\sigma)$, where x, μ , and $\sigma > 0$ are reals, $\Phi(u) = \int_{-\infty}^u \phi(\xi) d\xi$, and $\phi(\xi) = \exp(-\xi^2/2)/\sqrt{2\pi}$. Then X is said to be Gaussian random variable with mean μ , standard deviation σ , and variance σ^2 , a property denoted by $X \sim N(\mu, \sigma^2)$. \diamond

Theorem 2.4 Let (Ω, \mathcal{F}, P) be a probability space. A function $X : \Omega \rightarrow \mathbb{R}^d$ is an \mathbb{R}^d -valued random variable on (Ω, \mathcal{F}, P) if its coordinates are real-valued random variables on this probability space.

Proof If X is a random vector, its coordinates $X_i = \pi_i \circ X$, $i = 1, \dots, d$, are random variables because the projection map $\pi_i(x) = x_i$ is continuous.

Suppose now that X_i are random variables. Let \mathcal{R} be the collection of open rectangles in \mathbb{R}^d with members $R = I_1 \times \dots \times I_d$, where I_i are open intervals in \mathbb{R} . The σ -field \mathcal{B}^d is generated by these rectangles, that is, $\mathcal{B}^d = \sigma(\mathcal{R})$ ([11], Proposition 3.2.4), so that it is sufficient to show that $X^{-1}(\mathcal{R})$ is in \mathcal{F} . We have

$X^{-1}(R) = \cap_{i=1}^d X_i^{-1}(I_i) \in \mathcal{F}$ since X_i are random variables. Hence, X is a random vector if and only if $\{\omega \in \Omega : X_i(\omega) \leq x_i\} \in \mathcal{F}$ for all $x_i \in \mathbb{R}$, $i = 1, \dots, d$. \blacktriangle

Example 2.14 Consider a series $X = (X_1, X_2, \dots)$, where X_i are measurable functions from (Ω, \mathcal{F}) to (Ψ, \mathcal{G}) . Let \mathcal{K} be the collection of all subsets of $\mathbb{Z}^+ = \{1, 2, \dots\}$. The function $(m, \omega) \mapsto X_m(\omega)$ depending on the arguments m and ω is measurable from $(\mathbb{Z}^+ \times \Omega, \mathcal{K} \times \mathcal{F})$ to (Ψ, \mathcal{G}) . Generally, this property does not hold if the discrete index m in this example is allowed to take values in an uncountable set. \diamond

Proof Let $A = \{(m, \omega) : X_m(\omega) \in B\}$ be the inverse image of the function $(m, \omega) \mapsto X_m(\omega)$ in $\mathbb{Z}^+ \times \Omega$ corresponding to an arbitrary member B of \mathcal{G} . Because X_m is measurable, the set A can be expressed as the countable union $\cup_m \{\omega : X_m(\omega) \in B\}$ of sets $\{\omega : X_m(\omega) \in B\}$ that are in \mathcal{F} for each $m \geq 1$. Hence, A is in $\mathcal{K} \times \mathcal{F}$. Note also that the function $m \mapsto X_m(\omega)$ is \mathcal{K} -measurable for a fixed $\omega \in \Omega$ since $\{m : X_m(\omega) \in B\}$ is a subset of \mathbb{Z}^+ so that it is in \mathcal{K} . \blacktriangle

Definition 2.13 Let (Ω, \mathcal{F}) and (S, \mathcal{S}) be measurable spaces and let $X : \Omega \rightarrow S$ be a random element, that is, an $(\mathcal{F}, \mathcal{S})$ -measurable mapping. The σ -field generated by X , denoted by $\sigma(X)$ or \mathcal{F}^X , is $\sigma(X) = X^{-1}(\mathcal{S}) = \{X^{-1}(B) : \forall B \in \mathcal{S}\}$.

Example 2.15 Let (Ω, \mathcal{F}, P) be a probability space and $1_A : \Omega \rightarrow \mathbb{R}$, $A \in \mathcal{F}$, an indicator function. The σ -field generated by 1_A is $\sigma(1_A) = \{\emptyset, \Omega, A, A^c\}$. There is no smaller field with respect to which 1_A is measurable. \diamond

It is common in applications to be interested in functions of random variables. For example, we may have to find the distribution of the deformation or any other response Y of a physical system with random properties X_1 that is subjected to a random action X_2 . Suppose X_1 and X_2 are random variables on a probability space (Ω, \mathcal{F}, P) . To achieve this objective, we first need to show that Y , which is a function of (X_1, X_2) , is a random variable on (Ω, \mathcal{F}, P) . The following theorem shows that Y has this property if the mapping $(X_1, X_2) \mapsto Y$ is measurable.

Theorem 2.5 Let $h : (\Omega, \mathcal{F}) \rightarrow (\Psi, \mathcal{G})$ and $g : (\Psi, \mathcal{G}) \rightarrow (\Phi, \mathcal{H})$ be measurable functions, where (Ω, \mathcal{F}) , (Ψ, \mathcal{G}) , (Φ, \mathcal{H}) are measurable spaces. Then the function $g \circ h : (\Omega, \mathcal{F}) \rightarrow (\Phi, \mathcal{H})$ is measurable.

Proof We have $g^{-1}(\mathcal{H}) \subset \mathcal{G}$ and $h^{-1}(g^{-1}(\mathcal{H})) \subset h^{-1}(\mathcal{G}) \subset \mathcal{F}$ since g and h are measurable functions. \blacktriangle

2.4 Independence

We define independent σ -fields and show that this concept applies directly to characterize independent events and random elements.

Definition 2.14 Let (Ω, \mathcal{F}, P) be a probability space and \mathcal{F}_i , $i \in I$, a collection of sub- σ -fields of \mathcal{F} . If I is finite, the σ -fields \mathcal{F}_i , $i \in I$, are independent if

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i), \quad \forall A_i \in \mathcal{F}_i. \quad (2.11)$$

If I is infinite and (2.11) holds for all finite subsets of I , then the σ -fields \mathcal{F}_i , $i \in I$, are said to be independent.

The definition implies that (2.11) must hold for any subset of $\{A_i \in \mathcal{F}_i, i \in I\}$ since some A_i can be selected to coincide with Ω . If the σ -fields \mathcal{F}_i , $i \in I$, are on different sample spaces, the above independence condition needs to be applied on the corresponding product probability space.

Definition 2.15 A collection of events $A_i \in \mathcal{F}$, $i \in I$, is said to be independent if the σ -fields $\sigma(A_i) = \{\emptyset, \Omega, A_i, A_i^c\}$ are independent.

Example 2.16 Two events A and B on a probability space (Ω, \mathcal{F}, P) are independent if $P(A \cap B) = P(A)P(B)$. This is the classical definition for the independence between two events. \diamond

Proof By definition, A and B are independent if the fields $\sigma(A)$ and $\sigma(B)$ are independent, which implies $P(A' \cap B') = P(A')P(B')$ for all $A' \in \sigma(A)$ and $B' \in \sigma(B)$. The stated condition of independence follows by considering all possible pairs of events (A', B') . The converse also holds, that is, $P(A \cap B) = P(A)P(B)$ implies the independence of the fields $\sigma(A)$ and $\sigma(B)$. For example, $P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(A^c)P(B)$.

Note also that, the classical definition of independence between two events follows from that of the conditional probability. If A and B are independent and $P(B) > 0$, the conditional probability $P(A | B) = P(A \cap B)/P(B)$ is unaffected by the occurrence of B so that $P(A | B) = P(A)$ implying $P(A \cap B) = P(A)P(B)$. \blacktriangle

Example 2.17 Consider two events A and B as in Example 2.16. If $A \cap B = \emptyset$ and $P(A), P(B) > 0$, then A and B are not independent since $P(A \cap B) = P(\emptyset) = 0$ and $P(A), P(B) > 0$. \diamond

Definition 2.16 The events A_i , $i = 1, \dots, n$, on a probability space (Ω, \mathcal{F}, P) are independent if

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = \prod_{k=1}^m P(A_{i_k}) \quad (2.12)$$

holds for any subset $\{i_1, \dots, i_m\}$ of $\{1, \dots, n\}$.

Example 2.18 The requirement in (2.12) is essential for three or more events. It is not sufficient to satisfy (2.12) for the entire collection of events. For example, set $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} =$ all parts of Ω , and $P(\{1\}) = \sqrt{2}/2 - 1/4$, $P(\{2\}) = 1/4$, $P(\{3\}) = 3/4 - \sqrt{2}/2$, and $P(\{4\}) = 1/4$. Let $A_1 = \{1, 3\}$, $A_2 = \{2, 3\}$, and $A_3 = \{3, 4\}$ be events in (Ω, \mathcal{F}) . The probability of $A_1 \cap A_2 \cap A_3 = \{3\}$ is $P(\{3\}) = 3/4 - \sqrt{2}/2$ and is equal to $P(A_1)P(A_2)P(A_3)$. However, $P(A_1 \cap A_2) \neq P(A_1)P(A_2)$ ([9], p. 2). \diamond

Definition 2.17 Let (Ω, \mathcal{F}, P) be a probability space and let $\mathcal{C}_i \subset \mathcal{F}$, $i \in I$, be families of subsets of \mathcal{F} . If $I = \{1, \dots, n\}$ is finite, \mathcal{C}_i are said to be independent if any $A_1 \in \mathcal{C}_1, \dots, A_n \in \mathcal{C}_n$ are independent events. If I is not finite, \mathcal{C}_i are independent if \mathcal{C}_i , $i \in J$, are independent families for each finite $J \subseteq I$.

This definition and the following criterion can be used to show that two or more σ -fields are independent. The criterion uses classes of events forming a π -system. A collection \mathcal{C} of subsets of Ω is said to be a π -system if it is closed to finite intersection, that is, $A, B \in \mathcal{C}$ implies $A \cap B \in \mathcal{C}$. If \mathcal{C}_i is a nonempty class of events in \mathcal{F} for each $i = 1, \dots, n$, such that (1) \mathcal{C}_i is a π -system and (2) \mathcal{C}_i , $i = 1, \dots, n$, are independent, then the σ -fields $\sigma(\mathcal{C}_i)$, $i = 1, \dots, n$, are independent ([11], Theorem 4.1.1).

Definition 2.18 Let (Ω, \mathcal{F}, P) be a probability space, S a metric space, \mathcal{S} the σ -field induced on S by its metric, and $X_i : \Omega \rightarrow S$, $i \in I$, a collection of random elements, that is, a collection of $(\mathcal{F}, \mathcal{S})$ -measurable functions, where I may or may not be finite. The random elements X_i are independent if the σ -fields $\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{S}))$, $i \in I$ generated by these elements are independent.

Example 2.19 Let X and Y be real-valued random variables defined on (Ω, \mathcal{F}, P) . The independence of $\sigma(X)$ and $\sigma(Y)$ implies $P(X^{-1}((-\infty, x]) \cap Y^{-1}((-\infty, y])) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$, $x, y \in \mathbb{R}$, where F_X and F_Y denote the distributions of X and Y . The converse also holds, that is, $P(X \leq x, Y \leq y) = F_X(x)F_Y(y)$ implies the independence of $\sigma(X)$ and $\sigma(Y)$. For example, $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = P(X \leq a_2, Y \leq b_2) - P(X \leq a_1, Y \leq b_2) - P(X \leq a_2, Y \leq b_1) + P(X \leq a_1, Y \leq b_1)$ by properties of the probability measure, or $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = F_X(a_2)F_Y(b_2) - F_X(a_1)F_Y(b_2) - F_X(a_2)F_Y(b_1) + F_X(a_1)F_Y(b_1)$, so that $P(\{a_1 < X \leq a_2\} \cap \{b_1 < Y \leq b_2\}) = (F_X(a_2) - F_X(a_1))(F_Y(b_2) - F_Y(b_1)) = P(a_1 < X \leq a_2)P(b_1 < Y \leq b_2)$. The latter relationship implies the independence of the σ -fields $\sigma(X)$ and $\sigma(Y)$ since the intervals $(a_1, a_2]$ and $(b_1, b_2]$ are arbitrary. \diamond

Example 2.20 Let X_k , $k = 1, 2, \dots$, be independent, real-valued random variables on a probability space (Ω, \mathcal{F}, P) and $\varphi_k : \mathbb{R} \rightarrow \mathbb{R}$ Borel measurable functions. The random variables $\varphi_k \circ X_k$, $k = 1, 2, \dots$, are independent.

Proof We have $\varphi_k^{-1}(\mathcal{B}) \subseteq \mathcal{B}$ and $X_k^{-1}(\varphi_k^{-1}(\mathcal{B})) \subseteq \mathcal{F}$ because φ_k is a Borel measurable function and X_k is a random variable. Since $X_k^{-1}(\varphi_k^{-1}(\mathcal{B})) \subseteq X_k^{-1}(\mathcal{B})$ and the σ -fields $X_k^{-1}(\mathcal{B})$, $k = 1, 2, \dots$, are independent by assumption, the random variables $\varphi_k \circ X_k$ are independent. \diamond

Example 2.21 Let $X(t)$ and $Y(t)$, $t \in [0, 1]$, be simple real-valued processes with continuous samples $(x_1(t), \dots, x_m(t))$ and $(y_1(t), \dots, y_n(t))$ occurring with probabilities (p_1, \dots, p_m) and (q_1, \dots, q_n) , respectively. It is assumed that both processes are defined on the same probability space (Ω, \mathcal{F}, P) , so that they are measurable functions from (Ω, \mathcal{F}) to $(C[0, 1], \mathcal{C})$, and $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$, $i = 1, \dots, m$, and $B_j = \{\omega \in \Omega : Y(\omega) = y_j\}$, $j = 1, \dots, n$, are measurable partitions of Ω . The

processes X and Y are independent if the σ -fields generated by $\{A_i, i = 1, \dots, m\}$ and $\{B_j, j = 1, \dots, n\}$ are independent. \diamond

2.5 Sequence of Events

Let $\{A_n, n = 1, 2, \dots\}$ be a sequence of events in a probability space (Ω, \mathcal{F}, P) . Properties of probability measures that involve increasing/decreasing, convergent, and arbitrary sequences of events are discussed.

Definition 2.19 The sequence $\{A_n, n = 1, 2, \dots\}$ is said to be increasing if $A_n \subseteq A_{n+1}$ for all n . If $A_n \supseteq A_{n+1}$ for all n , the sequence is decreasing. The sequence is convergent if $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ and $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$ coincide, and we use the notation $\lim_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$ for the limit of $\{A_n\}$. Note that $\limsup_{n \rightarrow \infty} A_n$, $\liminf_{n \rightarrow \infty} A_n$, and $\lim_{n \rightarrow \infty} A_n$ are events in (Ω, \mathcal{F}, P) .

Theorem 2.6 (Continuity of probability measure) *Let $\{A_n, n = 1, 2, \dots\}$ be an increasing or decreasing sequence of events. The numerical sequence $\{P(A_n), n = 1, 2, \dots\}$ is increasing or decreasing and converges to $P(A)$, where $A = \lim_{n \rightarrow \infty} A_n$.*

Proof Suppose $\{A_n\}$ is increasing, so that it converges to $A = \lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Set $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$, $n = 2, 3, \dots$, so that $A_n = \bigcup_{k=1}^n B_k$, $A = \bigcup_{n=1}^{\infty} B_n$, $P(A) = P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(B_k) = \lim_{n \rightarrow \infty} P(\bigcup_{k=1}^n B_k) = \lim_{n \rightarrow \infty} P(A_n)$ since $\{B_n\}$ are disjoint events. Similar arguments hold for decreasing sequences. \blacktriangle

A direct consequence of Theorem 2.6 is that, for a sequence $\{A_n, n = 1, 2, \dots\}$ of convergent events, probability and limit can be interchanged, that is,

$$\lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n) = P(A), \quad (2.13)$$

where $A = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$ (Exercise 2.9 and [11], Sect. 2.1).

Let $\{A_n, n = 1, 2, \dots\}$ be a sequence of events in a probability space (Ω, \mathcal{F}, P) , and let $A = \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ be an event in this space. It can be seen that $\omega \in A$ if and only if $\omega \in A_n$ for infinitely many indices n . We use the notation

$$\{A_n \text{ i.o.}\} = \{A_n \text{ infinitely often}\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \limsup_{n \rightarrow \infty} A_n \quad (2.14)$$

to indicate this property.

Theorem 2.7 (Borel–Cantelli lemma) *If $\{A_n, n = 1, 2, \dots\}$ is a sequence of events such that $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ i.o.}) = 0$.*

If $\{A_n, n = 1, 2, \dots\}$ is a sequence of independent events, then $P(A_n \text{ i.o.}) = 0$ and $P(A_n \text{ i.o.}) = 1$ if and only if $\sum_{n=1}^{\infty} P(A_n) < \infty$ and $\sum_{n=1}^{\infty} P(A_n) = \infty$, respectively.

Proof We have $P(A_n \text{ i.o.}) = \lim_{n \rightarrow \infty} P(\cup_{k=n}^{\infty} A_k) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0$, where the first and the last equalities hold since $\{\cup_{k=n}^{\infty} A_k\}$ is an increasing sequence of events and $\sum_{n=1}^{\infty} P(A_n)$ is convergent by assumption.

The proof of the second part of the Borel–Cantelli lemma can be found in [11] (Proposition 4.5.2). ▲

Example 2.22 Let $\{X_n, n = 1, 2, \dots\}$ be a sequence of Bernoulli random variables taking the values 1 and 0 with probabilities $P(X_n = 1) = p_n = 1 - P(X_n = 0)$, $n \geq 1$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $P(\{X_n = 1\} \text{ i.o.}) = 0$ by the Borel–Cantelli lemma, so that X_n is equal to 1, a finite number of times. Other illustrations of the use of the Borel–Cantelli lemma can be found in [11] (Sect. 4.5). ◇

Example 2.23 Let $p_0 = P(V \leq v_{\text{cr}})$ denote the probability that yearly wind speed maximum does not exceed a critical value v_{cr} at a site. The event $\{V \leq v_{\text{cr}}\}$ occurs infinitely often with probability one since $\sum_{n=1}^{\infty} p_0 = \infty$ for $p_0 > 0$ and the events $\{V \leq v_{\text{cr}}\}$ in distinct years are independent. The statement follows from the second part of the Borel–Cantelli lemma. ◇

2.6 Expectation

The expectation operator is first defined for simple random variables. The definition is then extended to positive and arbitrary random variables, random vectors, and random functions. The expectation operator is used to define moments of random elements. Fubini’s theorem and applications of this theorem conclude the section.

Definition 2.20 Let (Ω, \mathcal{F}, P) be a probability space, $\{A_i \in \mathcal{F}, i \in I\}$ a partition of Ω , I a finite index set, and $a_i \in \mathbb{R}^d$ such that $\|a_i\| < \infty$. Then

$$X = \sum_{i \in I} a_i 1_{A_i}, \quad a_i \in \mathbb{R}^d, \quad (2.15)$$

is called a finite, simple \mathbb{R}^d -valued random variable. The collection of simple random variable is a vector space (Exercise 2.11).

Definition 2.21 The expectation of X in (2.15), denoted by $E[X]$, $\int_{\Omega} X dP$, or $\int_{\Omega} X(\omega) dP(\omega)$, is

$$E[X] = \int_{\Omega} X dP = \int_{\Omega} X(\omega) dP(\omega) = \sum_{i \in I} a_i P(A_i). \quad (2.16)$$

Consider a mapping $g : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ such that $\|g(a_i)\| < \infty$, $i \in I$. Then, $g(X)$ is a simple $\mathbb{R}^{d'}$ -valued random variable with expectation

$$E[g(X)] = \sum_{i \in I} g(a_i) P(A_i). \quad (2.17)$$

The definition is meaningful since $\|g(a_i)\| < \infty$, $i \in I$.

Following are the properties of the expectation for simple random variables. Except for Jensen's inequality (Exercise 2.12), all the other properties result directly from (2.16) and (2.17). The random variables in the first three properties given by (2.18) are real-valued.

If $a_i \geq 0$ in (2.15), then $E[X] \geq 0$,

If $X \leq Y$ a.s., then $E[X] \leq E[Y]$,

If $g: \mathbb{R} \rightarrow \mathbb{R}$ is convex, then $g(E[X]) \leq E[g(X)]$ (Jensen's inequality),

$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$, $\alpha, \beta \in \mathbb{R}$, and

$$\int_{\cup_i B_i} X dP = \sum_i \int_{B_i} X dP, \quad \text{where } \{B_i \in \mathcal{F}\} \text{ is a measurable partition of } \Omega. \quad (2.18)$$

The second and the fourth properties show that $E[\cdot]$ is a monotone and a linear operator. Jensen's inequality implies $|E[X]| \leq E[|X|]$ since the absolute value is a convex function. The inequality also follows from the definition of the expectation, which gives $|E[X]| = |\sum_{i \in I} a_i P(A_i)| \leq \sum_{i \in I} |a_i| P(A_i) = E[|X|]$.

We now extend the definition of expectation in (2.16) to positive real-valued random variables and, subsequently, arbitrary random variables.

Definition 2.22 Let X be a real-valued, positive random variable defined on a probability space (Ω, \mathcal{F}, P) , that is, $P(X \geq 0) = 1$. If $P(X = \infty) > 0$, set $E[X] = \infty$. Otherwise,

$$E[X] = \lim_{n \rightarrow \infty} E[X_n], \quad (2.19)$$

where X_n is an approximating sequence of finite-valued simple random variables such that $X_n \uparrow X$ and $E[X_n]$ is finite for each n .

The definition is meaningful since (1) there exists an increasing sequence of simple random variables $X_n \geq 0$, $n = 1, 2, \dots$, referred to as an approximating sequence of X , such that $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ for almost all ω 's ([11], Theorem 5.1.1) and (2) the expectation in (2.19) is well-defined since the value of $E[X]$ does not change if $\{X_n\}$ is replaced with another approximating sequence $Y_m \uparrow X$ ([11], Proposition 5.2.1).

Definition 2.23 Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) , and set $X^+ = X \vee 0 = \max(X, 0)$ and $X^- = (-X) \vee 0 = \max(-X, 0)$. The expectation of X is

$$E[X] = \begin{cases} E[X] = E[X^+] - E[X^-], & \text{if } E[X^+] < \infty \text{ and/or } E[X^-] < \infty \\ \text{Does not exist,} & \text{if } E[X^+] = E[X^-] = \infty. \end{cases} \quad (2.20)$$

If both expectations $E[X^+]$ and $E[X^-]$ are finite, then $E[X]$ exists, is finite, and has the expression $E[X] = E[X^+] - E[X^-]$. If only one of the expectations $E[X^+]$ and

$E[X^-]$ is finite, then $E[X]$ exists but is unbounded. As for simple random variables, we have

$$E[X1_B] = \int_{\Omega} 1_B(\omega)X(\omega) dP(\omega) = \int_B X dP, \quad B \in \mathcal{F}, \quad (2.21)$$

and, if $E[X1_B]$ is finite, we say that X is integrable with respect to P over B or just P -integrable over B . The definitions in (2.20) and (2.21) extend directly to random vectors and complex-valued random variables by applying them to the coordinates and real/imaginary parts of these variables, respectively.

Example 2.24 Let $X(\omega) = i^2 + j^2$ be a random variable defined on (Ω, \mathcal{F}, P) , where $\Omega = \{\omega = (i, j) : i, j = 1, 2, \dots, 6\}$ is the sample space for the experiment of rolling two dice, \mathcal{F} consists of all subsets of Ω , and $P(\{\omega\}) = 1/36$. Then X is a positive, simple random variable and $E[X1_B] = \sum_{\omega \in B} (i^2 + j^2)(1/36) = [(2)(40) + (2)(34) + 32]/36 = 5$ for $B = \{\omega = (i, j) : i + j = 8\}$. \diamond

The second and the third properties in (2.18) are also valid for the expectation in (2.20) (Exercises 2.12–2.14 and [4], Sect. 3.2). Following are the properties specific to the expectation in (2.20).

Theorem 2.8 *Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) . Then ([4], Sect. 3.2)*

$$E[X1_A] = \int_A X dP, \quad A \in \mathcal{F}, \quad \text{is finite if and only if } \int_A |X| dP < \infty,$$

$$\text{If } X \geq 0 \text{ a.s., then } \int_A X dP \geq 0, \text{ and}$$

If X is P -integrable, then X is finite a.s., that is,

$$N = \{\omega : X(\omega) = \pm\infty\} \in \mathcal{F} \text{ and } P(N) = 0. \quad (2.22)$$

Definition 2.24 Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) and let $q \geq 1$ be an integer. If $\mu(q) = E[X^q]$ exists and is finite, it is called the moment of order q of X . The mean of X is $\mu = \mu(1)$. If $E[(X - \mu)^q]$ exists and is finite, it is called the central moment of order q of X . The central moment $\sigma^2 = E[(X - \mu)^2]$ is the variance of X . The square root σ of the variance σ^2 is the standard deviation of X . The scaled versions of the third and fourth central moments $\gamma_3 = E[(X - \mu)^3]/\sigma^3$ and $\gamma_4 = E[(X - \mu)^4]/\sigma^4$ are the skewness and kurtosis coefficients of X . The ratio $v = \sigma/\mu$ defined for $\mu \neq 0$ is called coefficient of variation.

The definition of moments of X is meaningful since the mappings $X \mapsto X^q$, $(X - \mu)^q$ are continuous and, therefore, measurable. Hence, X^q , $(X - \mu)^q$ are random variables on (Ω, \mathcal{F}, P) .

Definition 2.25 Let $X = (X_1, \dots, X_d)$ be an \mathbb{R}^d -valued random variable defined on a probability space (Ω, \mathcal{F}, P) and let $q \geq 1$ be an integer. The moments of order q of X are

$$\mu(q_1, \dots, q_d) = E \left[\prod_{i=1}^d X_i^{q_i} \right] \quad (2.23)$$

provided the expectations in (2.23) exist and are finite, where $q_i \geq 0$ are integers such that $\sum_{i=1}^d q_i = q$. The mean or the expectation of X is the vector $E[X] = (E[X_1] = \mu(1, 0, \dots, 0), \dots, E[X_d] = \mu(0, \dots, 0, 1))$. The (d, d) -matrix $r = \{r_{ij} = E[X_i X_j], i, j = 1, \dots, d\}$ whose entries are moments of order $q=2$ is called the correlation matrix of X . The formula in (2.23) with $X_i - E[X_i]$ in place of X_i , $i = 1, \dots, d$, give the central moments of order q of X . The (d, d) -matrix $c = \{c_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])], i, j = 1, \dots, d\}$ is called the covariance matrix of X . The scaled version of the covariance matrix $\rho = \{\rho_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]/(\sigma_i \sigma_j)], i, j = 1, \dots, d\}$ is the correlation coefficient matrix, where $\sigma_i^2 = c_{ii}$, $i = 1, \dots, d$.

The definition in (2.23) is meaningful since the mapping $X \mapsto \prod_{i=1}^d X_i^{q_i}$ is continuous so that it is measurable implying that $\prod_{i=1}^d X_i^{q_i}$ is a random variable defined on the same probability space as X . Similar considerations hold for the mapping $X \rightarrow \prod_{i=1}^d (X_i - E[X_i])^{q_i}$.

Definition 2.26 Let X be an \mathbb{R}^d -valued random variable with finite moments of order 2. Two distinct coordinates X_i and X_j of X are said to be orthogonal if $r_{ij} = E[X_i X_j] = 0$. If $c_{ij} = (X_i - E[X_i])(X_j - E[X_j]) = 0$, then X_i and X_j are said to be uncorrelated. The correlations and covariances of X coincide if $E[X_i] = 0$ for all $i = 1, \dots, d$.

Example 2.25 Let $X = U + iV$ be a complex-valued random variable, where U and V are real-valued random variables. If $E[|U|] < \infty$ and $E[|V|] < \infty$, then $E[X]$ exists and is equal to $E[X] = E[U] + iE[V]$, that is, the real and imaginary parts of X are viewed as the coordinates of a two-dimensional vector. \diamond

The coordinate by coordinate definition of the expectation for random vectors extends directly to random functions by viewing their values at various arguments as coordinates of a random vector with finite, countable, or uncountable number of coordinates.

Definition 2.27 Let $X(t) = (X_1(t), \dots, X_d(t))$, $t \in D \subseteq \mathbb{R}^{d'}$, be an \mathbb{R}^d -valued random function defined on a probability space (Ω, \mathcal{F}, P) . The expectation of $X(t)$ exists if and only if the expectation of all its coordinates $\{X_i(t), t \in D, i = 1, \dots, d\}$ exist. In this case, we have $E[X(t)] = (E[X_1(t)], \dots, E[X_d(t)])$, $t \in D$.

Example 2.26 Let Y_1 and Y_2 be random variables on a probability space (Ω, \mathcal{F}, P) such that $E[|Y_k|] < \infty$, $k=1, 2$. The \mathbb{R}^2 -valued random function $X(t) = (X_1(t) = Y_1 \cos(t), X_2(t) = Y_2 \sin(t))$, $t \in [0, 2\pi]$, is defined on (Ω, \mathcal{F}, P) and has expectation $(E[X_1(t)] = E[Y_1] \cos(t), E[X_2(t)] = E[Y_2] \sin(t))$, $t \in [0, 2\pi]$. \diamond

We conclude this section with the statement of Fubini's theorem specifying conditions under which integrals of measurable functions defined on product probability

spaces can be performed sequentially. The theorem is used extensively in calculations, as illustrated by two examples.

Theorem 2.9 (Fubini's theorem) *Let $(\Omega_k, \mathcal{F}_k, P_k)$, $k = 1, 2$, be two complete probability spaces and denote by $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, and $P = P_1 \times P_2$ the product sample space, σ -field, and probability measure. If $(\omega_1, \omega_2) \mapsto X(\omega_1, \omega_2)$ is \mathcal{F} -measurable and P -integrable, then*

$$\begin{aligned} X(\omega_1, \cdot) \text{ is } \mathcal{F}_2\text{---measurable and } P_2\text{---integrable for each } \omega_1 \in \Omega_1, \\ \int_{\Omega_2} X(\cdot, \omega_2) P_2(d\omega_2) \text{ is } \mathcal{F}_1\text{---measurable and } P_1\text{---integrable, and} \\ \int_{\Omega} X(\omega) P(d\omega) = \int_{\Omega_1} \left[\int_{\Omega_2} X(\omega_1, \omega_2) P_2(d\omega_2) \right] P_1(d\omega_1). \end{aligned} \quad (2.24)$$

If in addition X is positive and either side of (2.24) exists and is finite or infinite, so is the other side of the equality, and the equality is valid ([4], p. 59).

Example 2.27 Let (Ω, \mathcal{F}, P) and $([0, 1], \mathcal{H}([0, 1]), \lambda)$ be measure spaces, where λ denotes the Lebesgue measure. Let $X : ([0, 1] \times \Omega, \mathcal{B}([0, 1]) \times \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ be a measurable function defined on the product of these spaces endowed with the product measure $\lambda \times P$. It is common to interpret the first argument of X as time. The integral $I(A, \omega) = \int_A 1_B(X(s, \omega)) ds$, $A \in \mathcal{B}([0, 1])$, $B \in \mathcal{B}$, represents the time $X(\cdot, \omega)$, $\omega \in \Omega$, spends in B during a time interval A . The expectation of this occupation time is $E[I(A, \omega)] = \int_A P(X(s) \in B) ds$. \diamond

Proof The measurable mapping $(s, \omega) \mapsto X(s, \omega)$ is said to be a stochastic process, and $s \mapsto X(s, \omega)$ is sample ω of X . For $\mathcal{K} = \{\emptyset, \{0, 1\}, \{0\}, \{1\}\}$, the indicator function, $1_B : (\mathbb{R}, \mathcal{B}) \rightarrow (\{0, 1\}, \mathcal{K})$, $B \in \mathcal{B}$, is measurable so that $1_B \circ X : ([0, 1] \times \Omega, \mathcal{B}([0, 1]) \times \mathcal{F}) \rightarrow (\{0, 1\}, \mathcal{K})$ is also measurable. The expectation of the occupation time is

$$\begin{aligned} E[I(A, \omega)] &= \int_{\Omega \times A} 1_B(X(s, \omega)) ds P(d\omega) = \int_{\Omega} \left[\int_A 1_B(X(s, \omega)) ds \right] P(d\omega) \\ &= \int_A \left[\int_{\Omega} 1_B(X(s, \omega)) P(d\omega) \right] ds = \int_A P(X(s) \in B) ds, \end{aligned}$$

by Fubini's theorem ([11], Example 5.9.1). \blacktriangle

Example 2.28 Consider a cantilever beam with unit length and stiffness that is fixed at its left end and subjected to a distributed random load $X(x)$, $x \in [0, 1]$. It is assumed that the mapping $(x, \omega) \mapsto X(x, \omega)$ is measurable from $([0, 1] \times \Omega, \mathcal{B}[0, 1] \times \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$ and $\lambda \times P$ -integrable, where λ denotes the Lebesgue measure on the real line. The beam displacement $U(1)$ at its free end and its expectations are

$$U(1) = \int_{[0,1]^3} X(x+u)u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz \quad \text{and}$$

$$E[U(1)] = \int_{[0,1]^3} E[X(x+u)]u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz.$$

If $E[X(x)] = q$ is space invariant, then $E[U(1)] = q/8$. \diamond

Proof The beam displacement satisfies the equation $U''(x) = -M(x)$ with boundary conditions $U(0) = 0$ and $U'(0) = 0$, where $M(x) = -\int_0^{1-x} X(x+u)u du$ denotes the bending moment in the beam. The solution of this equation is $U(x) = \int_0^x dz \int_0^z dy \int_0^{1-y} X(x+u)u du$, so that

$$E[U(1)] = \int_{\Omega \times [0,1]^3} X(x+u, \omega)u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz P(d\omega),$$

where $du dy dz$ is the Lebesgue measure on $[0, 1]^3$. Since $(u, y, z, \omega) \mapsto X(u, y, z, \omega)$ is measurable, we have

$$E[U(1)] = \int_{[0,1]^3} \left[\int_{\Omega} X(x+u, y, z, \omega) P(d\omega) \right] u1(0 \leq u \leq 1-y, 0 \leq y \leq z) du dy dz,$$

by Fubini's theorem, which gives the stated formula of $E[U(1)]$. \blacktriangle

Example 2.29 Let $X(t, \omega)$ and $Y(t, \omega)$ be real-valued random variables defined on a probability space $([0, \tau] \times \Omega, \mathcal{B}[0, \tau] \times \mathcal{F}, \lambda \times P)$, $0 < \tau < \infty$. Suppose X is the solution of the differential equation $\ddot{X}(t, \omega) + 2\zeta v_0 \dot{X}(t, \omega) + v_0^2 X(t, \omega) = Y(t, \omega)$ with $X(0, \omega) = 0$, $\dot{X}(0, \omega) = 0$, where $\zeta \in (0, 1)$ and $v_0 > 0$ are constants, and the dots denote time derivatives. We will see that the solution of this equation is a stochastic process, that is, a family of random variables indexed by $t \in [0, \tau]$, defined by $X(t, \omega) = \int_0^t h(t-s)Y(s, \omega) ds$, where $h(u) = \exp(-\zeta v_0 u) \sin(v_d u)/v_d$ and $v_d = v_0 \sqrt{1 - \zeta^2}$. The expectation of this process at an arbitrary time t is the double integral $E[X(t, \omega)] = \int_{\Omega} \left[\int_0^t h(t-s)Y(s, \omega) ds \right] P(d\omega)$. If $h(t-s)Y(s, \omega)$ is $\mathcal{B}[0, \tau] \times \mathcal{F}$ -measurable and $\lambda \times P$ -integrable, Fubini's theorem applies and gives

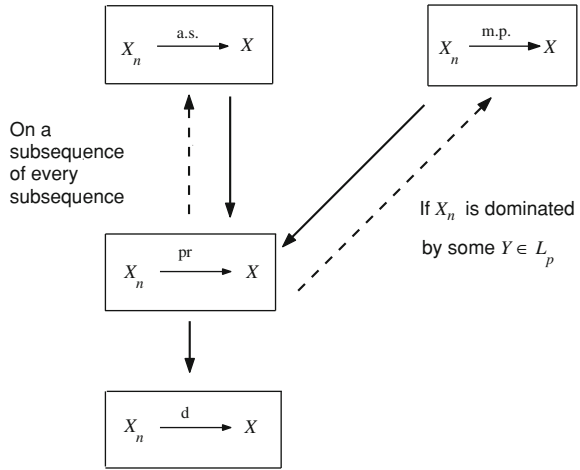
$$E[X(t, \omega)] = \int_0^t h(t-s)E[Y(s, \omega)] ds \quad (2.25)$$

so that $E[X(t, \omega)] = \mu_y \int_0^t h(t-s) ds$ if $E[Y(s, \omega)] = \mu_y$ is time invariant. \diamond

2.7 Convergence of Sequences of Random Variables

Let X and X_n , $n = 1, 2, \dots$, be real-valued random variables defined on a probability space (Ω, \mathcal{F}, P) . There are various definitions for the convergence $X_n \rightarrow X$ depending on the manner in which the discrepancy between X_n and X is measured.

Fig. 2.1 Relations between the convergence of sequences of random variables



Definition 2.28 The sequence $\{X_n\}$ converges to X almost surely ($X_n \xrightarrow{\text{a.s.}} X$), in probability ($X_n \xrightarrow{\text{pr}} X$), in distribution ($X_n \xrightarrow{\text{d}} X$), and in L^p ($X_n \xrightarrow{\text{m.p.}} X$), where $p \geq 1$ is an integer, if

$$\begin{aligned}
 \lim_{n \rightarrow \infty} X_n(\omega) &= X(\omega), \quad \forall \omega \in \Omega \setminus N \text{ with } P(N) = 0, \\
 \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) &= 0, \quad \forall \varepsilon > 0, \\
 \lim_{n \rightarrow \infty} F_n(x) &= F(x) \text{ at continuity points } x \in \mathbb{R}, \text{ and} \\
 \lim_{n \rightarrow \infty} E[|X_n - X|^p] &= 0,
 \end{aligned} \tag{2.26}$$

respectively. The convergence $X_n \xrightarrow{\text{m.p.}} X$ for $p = 2$ is called mean square (m.s.) convergence and is denoted by $X_n \xrightarrow{\text{m.s.}} X$ or l.i.m. $\lim_{n \rightarrow \infty} X_n = X$.

Figure 2.1 adapted from [7] (Sect. 2.13) gives essential relations between various types of convergence. An extensive discussion on relationships between convergence types can be found, for example, in [4] (Chap. 4) and [11] (Sects. 6.3 and 8.5). We only discuss some of the arrows in Fig. 2.1. That m.s. convergence implies convergence in probability follows from the Chebyshev inequality.

The convergence $X_n \xrightarrow{\text{a.s.}} X$ means $P(|X_n - X| > \varepsilon \text{ i.o.}) = 0$ for $\varepsilon > 0$ arbitrary, since it requires that $|X_n(\omega) - X(\omega)|$ gets small and remains small for almost all $\omega \in \Omega$. We have $0 = P(\limsup_{n \rightarrow \infty} |X_n - X| > \varepsilon) = \lim_{n \rightarrow \infty} P(\cup_{m \geq n} |X_m - X| > \varepsilon) \geq \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon)$, that is, $X_n \xrightarrow{\text{pr}} X$.

To show that the convergence in distribution is implied by that in probability, note that for $\varepsilon > 0$ arbitrary, we have

$$\begin{aligned}
 P(X_n \leq x) &= P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(X_n \leq x, |X_n - X| > \varepsilon) \\
 &\leq P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(|X_n - X| > \varepsilon)
 \end{aligned}$$

and $P(X_n \leq x, |X_n - X| \leq \varepsilon) \leq P(X_n \leq x, X \leq X_n + \varepsilon) \leq P(X \leq x + \varepsilon)$, which gives $P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$. This inequality and a similar inequality obtained by interchanging X with X_n yield $P(X \leq x - \varepsilon) - P(|X_n - X| > \varepsilon) \leq P(X_n \leq x) \leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$, so that $\lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x)$ provided the distribution of X is continuous at x .

Example 2.30 Let $\{X_n\}$ be a sequence of random variables converging in probability to X . Then $\{X_n\}$ is a Cauchy sequence in probability, that is, for arbitrary $\varepsilon > 0$ and $\eta > 0$ there exists $n(\varepsilon, \eta)$ such that $P(|X_m - X_n| > \varepsilon) < \eta$ for $m, n \geq n(\varepsilon, \eta)$. \diamond

Proof Since $P(|X_m - X| + |X_n - X| > \varepsilon) \leq P(|X_m - X| > \varepsilon/2) + P(|X_n - X| > \varepsilon/2)$ and $P(|X_m - X| > \varepsilon/2), P(|X_n - X| > \varepsilon/2) \rightarrow 0$ as $m, n \rightarrow \infty$, the sequence $\{X_n\}$ is Cauchy in probability. \blacktriangle

Example 2.31 If $\{X_k\}$ are uncorrelated random variables with finite mean μ and variance σ^2 , then $\sum_{k=1}^n (X_k - \mu)/n \xrightarrow{\text{pr}} 0$ as $n \rightarrow \infty$. The convergence is referred to as the weak law of large numbers ([9], p. 36). \diamond

Example 2.32 If $\{X_k\}$ are independent identically distributed (iid) random variables with finite expectation, then $\sum_{k=1}^n X_k/n \xrightarrow{\text{a.s.}} E[X_1]$ as $n \rightarrow \infty$. The convergence is referred to as the strong law of large numbers [11] (Sects. 7.4 and 7.5). It shows that averages along almost all infinite samples of the sequence (X_1, X_2, \dots) are equal to $E[X_1]$. \diamond

Example 2.33 If $\{X_k\}$ are iid random variables with finite mean $\mu = E[X_1]$ and variance $\sigma^2 = E[(X_1 - \mu)^2]$, then $(1/\sqrt{n}) \sum_{k=1}^n (X_k - \mu)/\sigma \xrightarrow{d} N(0, 1)$ as $n \rightarrow \infty$. The convergence is referred to as the central limit theorem ([11], Sects. 8.2). \diamond

Theorem 2.10 (Dominated, bounded, and monotone convergence theorems) *Let X_n , $n = 1, 2, \dots$, be real-valued random variables defined on a probability space (Ω, \mathcal{F}, P) such that $\lim_{n \rightarrow \infty} X_n = X$ a.s., $A \in \mathcal{F}$, and Y a random variable defined on the same probability space. If*

- (1) $|X_n| \leq Y$ a.s., $Y \geq 0$, and $\int_A Y dP < \infty$,
- (2) $|X_n| \leq c$ a.s. for a positive constant c , or
- (3) $X_n \geq 0$ a.s. is an increasing sequence that can take on the value $+\infty$, then

$$\lim_{n \rightarrow \infty} \int_A X_n dP = \int_A (\lim_{n \rightarrow \infty} X_n) dP = \int_A X dP, \quad (2.27)$$

an equality referred to as the dominated convergence, the bounded convergence, or the monotone convergence under condition (1), condition (2), or condition (3), respectively ([4], Sect. 3.2).

The interchange of limit and integral operators in (2.27) resembles a property of Riemann integrals. For example, if the sequence of real-valued functions $\{h_n(x)\}$

converges uniformly to $h(x)$ on $[a, b]$, we have $\int_a^b h(x) dx = \int_a^b \lim_{n \rightarrow \infty} h_n(x) dx = \lim_{n \rightarrow \infty} \int_a^b h_n(x) dx$, where $\int_a^b (\cdot) dx$ denotes a Riemann integral ([2], Theorem 30.3).

Theorem 2.11 (Integration term by term, Fatou's lemma, Lebesgue's theorem) *If $\{X_n\}$ and Y are random variables on a probability space (Ω, \mathcal{F}, P) and $A \in \mathcal{F}$, the following three statements hold.*

If $\sum_n \int_A |X_n| dP < \infty$, then $\int_A \sum_n X_n dP = \sum_n \int_A X_n dP$;

If $X_n \geq 0$ a.s. on A , then $\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP$; and

If $|X_n| \leq Y$ where $Y \geq 0$ a.s. and P – integrable over A , then

$$\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP \leq \limsup_n \int_A X_n dP \leq \int_A (\limsup_n X_n) dP. \quad (2.28)$$

Proof The proof of the first two statements in (2.28), that is, integration term by term and Fatou's lemma, can be found in [4] (Sect. 3.2). We only prove the last statement, that is, Lebesgue's theorem. Recall that $\wedge_{k \geq n} X_k$ and $\vee_{k \geq n} X_k$ are increasing and decreasing sequences and that $\liminf_{n \rightarrow \infty} X_n = \sup_{n \geq 1} \inf_{k \geq n} X_k = \lim_{n \rightarrow \infty} \wedge_{k \geq n} X_k$ and $\limsup_{n \rightarrow \infty} X_n = \inf_{n \geq 1} \sup_{k \geq n} X_k = \lim_{n \rightarrow \infty} \vee_{k \geq n} X_k$. Note that $\vee_{k \geq n} X_k = -\wedge_{k \geq n} (-X_k)$ so that $\limsup_{n \rightarrow \infty} X_n = -\liminf_{n \rightarrow \infty} (-X_n)$. We also have $\{\vee_n X_n \leq x\} = \cap_n \{X_n \leq x\} \in \mathcal{F}$ and $\{\wedge_n X_n > x\} = \cap_n \{X_n > x\} \in \mathcal{F}$.

The Fatou lemma applied to sequence $\{X_n + Y\}$, $X_n + Y \geq 0$ a.s., gives

$$\int_A (\liminf_n X_n) dP + \int_A Y dP \leq \liminf_n \int_A X_n dP + \int_A Y dP$$

so that $\int_A (\liminf_n X_n) dP \leq \liminf_n \int_A X_n dP$. The last inequality in (2.28) results from $\sup\{X_n, X_{n+1}, \dots\} = -\inf\{-X_n, -X_{n+1}, \dots\}$. The middle inequality in Lebesgue's theorem is valid since $\int_A X_n dP$ is a numerical sequence. ▲

Example 2.34 If $\{X_n\}$ is a sequence of real-valued random variables converging a.s. to X and there exists a random variable $Z \geq 0$ a.s. such that $|X_n| \leq Z$, then $\lim_{n \rightarrow \infty} E[X_n] = E[\lim_{n \rightarrow \infty} X_n] = E[X]$. This convergence results by Lebesgue's theorem with $A = \Omega$. ◇

We conclude this section by noting that the family of random variables $\{X\}$ defined on a probability space (Ω, \mathcal{F}, P) with the property $E[|X|^p] < \infty$ constitute a vector space that, with the norm $\|X\| = (E[|X|^p])^{1/p}$ becomes a Hilbert space denoted by $L^p(\Omega, \mathcal{F}, P)$, where $p \geq 1$ is an integer. Properties of $L^p(\Omega, \mathcal{F}, P)$ relevant to our discussion are in Sect. B.5.

2.8 Radon–Nikodym Derivative

A direct application of Radon–Nikodym derivatives is the construction of improved Monte Carlo simulation algorithms for estimating properties of random elements.

Definition 2.29 Let (Ω, \mathcal{F}) be a measurable space and let $\mu, \nu: \Omega \rightarrow [0, \infty]$ be measures on this space. If $\mu(A) = 0, A \in \mathcal{F}$, implies $\nu(A) = 0$, we say that ν is absolutely continuous with respect to μ and indicate this property by the notation $\nu \ll \mu$. If $\nu \ll \mu$ and $\mu \ll \nu$, then ν and μ are said to be equivalent measures.

Example 2.35 Consider a measure space $(\Omega, \mathcal{F}, \mu)$ and a measurable function $h: (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$. Then

$$\nu(A) = \int_A h \, d\mu, \quad A \in \mathcal{F} \quad (2.29)$$

is a measure that is absolutely continuous with respect to μ . \diamond

Proof The set function ν is positive by definition. It is countably additive since for $A_n \in \mathcal{F}, n = 1, 2, \dots$, disjoint sets, we have

$$\begin{aligned} \nu(\cup_{n=1}^{\infty} A_n) &= \int_{\cup_{n=1}^{\infty} A_n} h \, d\mu = \int \sum_{n=1}^{\infty} h 1_{A_n} \, d\mu \\ &= \sum_{n=1}^{\infty} \int h 1_{A_n} \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} h \, d\mu = \sum_{n=1}^{\infty} \nu(A_n). \end{aligned}$$

The term by term integration is valid whether $\sum_{n=1}^{\infty} \int h 1_{A_n} \, d\mu$ is or is not finite since h is positive. The measure ν is absolutely continuous with respect to μ since

$$\nu(A) = \int_A h \, d\mu = \int_{\Omega} h 1_A \, d\mu \leq \sup_{\omega \in A} (h(\omega)) \mu(A)$$

so that $\mu(A) = 0$ implies $\nu(A) = 0$ with the convention $0 \cdot \infty = 0$. \blacktriangle

A converse of this example is provided by following theorem, referred to as the Radon–Nikodym theorem, that guarantees the existence of a measurable function h satisfying (2.29) for given measures μ and ν under some conditions.

Theorem 2.12 If μ and ν are σ -finite measures on a measurable space (Ω, \mathcal{F}) such that $\nu \ll \mu$, then there exists a measurable function

$$h = \frac{d\nu}{d\mu}: (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty))), \quad (2.30)$$

called the Radon–Nikodym derivative of ν with respect to μ , such that (2.29) holds ([5], Theorem 18, p. 116).

Example 2.36 Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) . The measure Q induced on $(\mathbb{R}, \mathcal{B})$ by P , that is, the distribution of X , is $Q(B) = P(X^{-1}(B))$, $B \in \mathcal{B}$. Set $F(x) = Q((-\infty, x])$ for $B = (-\infty, x]$, $x \in \mathbb{R}$.

Suppose Q is absolutely continuous with respect to the Lebesgue measure λ on the real line, that is, $\lambda(B) = 0$ implies $Q(B) = 0$ for all $B \in \mathcal{B}$. Since Q is a probability measure, it is finite and, hence, σ -finite. Theorem 2.12 shows that there exists a Radon-Nikodym derivative $h = dQ/d\lambda : (\mathbb{R}, \mathcal{B}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$ such that $Q(B) = \int_B h d\lambda$ for all $B \in \mathcal{B}$. For $B = (-\infty, x]$ we have $F(x) = Q(B) = \int_{-\infty}^x f(u) d\lambda(u)$ with the notation $h = f$. The function $f(x) = dF(x)/dx$ is called the probability density function or the density of X . \diamond

2.9 Distribution and Density Functions

Let X be an \mathbb{R}^d -valued random variable defined on a probability space (Ω, \mathcal{F}, P) . For $u, v \in \mathbb{R}^d$, we use the notation $u \leq v$ ($u < v$) to mean $u_i \leq v_i$ ($u_i < v_i$) for all $i = 1, \dots, d$. Similarly, $(-\infty, u]$ denotes the rectangle $\times_{i=1}^d (-\infty, u_i]$. We have seen that the distribution Q of X is the probability measure induced on $(\mathbb{R}^d, \mathcal{B}^d)$, that is, $Q(B) = P(X^{-1}(B))$, $B \in \mathcal{B}^d$ (Definition 2.12). It is common to view distributions of random variables as probability measures Q defined for intervals $B = \times_{i=1}^d (-\infty, x_i]$. This definition of Q is not restrictive ([11], Corollary 3.2.1 and Proposition 3.2.4).

Definition 2.30 The distribution function of X is

$$F(x) = P(X^{-1}((-\infty, x])) = P \circ X^{-1}((-\infty, x]) = Q((-\infty, x]), \quad x \in \mathbb{R}^d, \quad (2.31)$$

where $Q = P \circ X^{-1}$ is the probability measure induced by P on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Definition 2.31 Let F be the distribution of an \mathbb{R}^d -valued random variable that is absolutely continuous with respect to the Lebesgue measure λ on \mathbb{R}^d . The Radon-Nikodym derivative of F with respect to λ , referred to the density of X , exists and is given by

$$f(x) = \frac{\partial^d F(x)}{\partial x_1 \cdots \partial x_d}, \quad x \in \mathbb{R}^d. \quad (2.32)$$

Since $P(\cap_{i=1}^d \{X_i \in (x_i, x_i + dx_i]\}) \simeq f(x) dx$, the volume of f over an infinitesimal rectangle $\times_{i=1}^d (x_i, x_i + dx_i]$ gives the probability that X takes values in this rectangle.

Definition 2.32 The distribution of one or more coordinates of X can be obtained from the distribution or the density of X . For example, the distribution and the density of X_1 are $F_1(x_1) = F(x_1, \infty, \dots, \infty)$ and $f_1(x_1) = \int_{\mathbb{R}^{d-1}} f(x) dx_2 \cdots dx_d$ or $f_1(x_1) = dF_1(x_1)/dx_1$, respectively.

Let X be a real-valued random variable. Following is a list of properties of the distribution function F of X that are useful for calculations. F is a right continuous, increasing function with range $[0, 1]$; F can have only jump discontinuities and the set of these jumps is countable; F is continuous at $x \in \mathbb{R}$ if and only if $P(X = x) = 0$; $\lim_{x \rightarrow \infty} F(x) = 1$; $\lim_{x \rightarrow -\infty} F(x) = 0$; $P(a < X \leq b) = F(b) - F(a) \geq 0$ for $a \leq b$; and $P(a \leq X < b) = F(b) - F(a) + P(X = a) - P(X = b)$ for $a \leq b$. These facts follow the properties of probability measures and real-valued functions (Exercise 2.17, [7], Sect. 2.10.1, [11], Sect. 2.1).

If the distribution F of a real-valued random variable X is absolutely continuous with respect to the Lebesgue measure on the real line, it has a density $f(x) = dF(x)/dx$ with the following properties: $F(b) - F(a) = \int_a^b f(x) dx$ for $a \leq b$, $f(x) = F'(x)$ so that $\int_{-\infty}^x f(\xi) d\xi = F(x)$, $f \geq 0$ since F is an increasing function, and $\int_{-\infty}^{\infty} f(x) dx = 1$. Note that f is not a probability measure.

Example 2.37 Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) and let $g : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ be a measurable function. If $Y = g \circ X$, a random variable on (Ω, \mathcal{F}, P) , has finite expectation, then

$$\begin{aligned} E[Y] &= \int_{\Omega} Y(\omega) P(d\omega) = \int_{\Omega} g(X(\omega)) P(d\omega) \\ &= \int_{\mathbb{R}} g(x) Q(dx) = \int_{\mathbb{R}} g(x) dF(x) = \int_{\mathbb{R}} g(x) f(x) dx, \end{aligned} \quad (2.33)$$

where $Q(B) = P(X^{-1}(B))$, $B \in \mathcal{B}$, is the distribution of X . \diamond

Proof If $g = 1_B$, $B \in \mathcal{B}$, then $E[Y] = P(X \in B) = Q(B)$. If $g = \sum_{i \in I} b_i 1_{B_i}$, the subsets $B_i \in \mathcal{B}$ partition \mathbb{R} , I is a finite index set, and b_i are real constants, then $E[Y] = \sum_{i \in I} b_i P(X \in B_i) = \sum_{i \in I} b_i Q(B_i)$ since integration is a linear operator. If g is an arbitrary positive Borel function, there exists a sequence of simple, increasing, and measurable functions g_n , $n = 1, 2, \dots$, converging to g as $n \rightarrow \infty$. The expectations of $g_n(X)$ calculated by all formulas in (2.33) coincide, so that (2.33) holds also for $g(X)$ by the monotone convergence theorem. If g is an arbitrary Borel function, (2.33) holds since $g = g^+ - g^-$ and $\int g(X) dP = \int g^+(X) dP - \int g^-(X) dP$. \blacktriangle

Example 2.38 A real-valued random variable X with density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right), \quad x, \mu \in \mathbb{R}, \quad \sigma > 0, \quad (2.34)$$

is said to be Gaussian with mean μ , variance σ^2 , and standard deviation σ , a property denoted by $X \sim N(\mu, \sigma^2)$. The function $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$, $u \in \mathbb{R}$, is the density of the standard Gaussian variable $N(0, 1)$. \diamond

Example 2.39 Consider a Cauchy random variable X with density $f(x) = a/[\pi(a^2 + x^2)]$, $x \in \mathbb{R}$, where $a > 0$ is a constant. The expectation of X does not exist since

$E[X^+] = E[1_{X \geq 0} X] = \int_0^\infty x f(x) dx = a \log(a^2 + x^2)/(2\pi) \big|_0^\infty = +\infty$ and $E[X^-] = +\infty$. \diamond

Example 2.40 The distribution of $X = F^{-1} \circ \Phi(G)$ is F , where F is a continuous distribution, Φ denotes the distribution of $N(0,1)$, that is, $\Phi(x) = \int_{-\infty}^x \phi(u) du$, and $G \sim N(0, 1)$. \diamond

Proof Since X is a continuous function of G , it is a random variable. We have $P(X \leq x) = P(F^{-1} \circ \Phi(G) \leq x) = P(G \leq \Phi^{-1}(F(x))) = F(x)$, $x \in \mathbb{R}$. Similar arguments show that $\Phi(G) \sim U(0, 1)$ is a random variable uniformly distributed in $(0,1)$, that is, $P(\Phi(G) \leq x) = x$, $x \in [0, 1]$. The representation $X = F^{-1} \circ \Phi(G)$ is used to generate samples of random variables following arbitrary distributions. \blacktriangle

Suppose now that X is an \mathbb{R}^d -valued random variable and $d > 1$. As for the case $d = 1$, we list the properties of the distribution function F of X : $\lim_{x_k \rightarrow \infty} F(x)$, $1 \leq k \leq d$ is the joint distribution of $(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_d)$, $\lim_{x_k \rightarrow -\infty} F(x) = 0$ for $k \in \{1, \dots, d\}$, function $x_k \mapsto F(x)$ is increasing for each $k \in \{1, \dots, d\}$, and function $x_k \mapsto F(x)$ is right continuous for each $k \in \{1, \dots, d\}$.

Definition 2.33 Let X be an \mathbb{R}^d -valued random variable with density f . Denote by $X^{(1)}$ and $X^{(2)}$, the first $d_1 < d$ and the last $d_2 = d - d_1$ coordinates of X for $d \geq 2$. The conditional density $f^{(1|2)}$ of $X^{(1)}$ given $X^{(2)} = z$ is

$$f^{(1|2)}(x^{(1)} | z) = \frac{f(x^{(1)}, z)}{f^{(2)}(z)}, \quad (2.35)$$

where $f^{(k)}$ denotes the density of $X^{(k)}$, $k = 1, 2$, and $x^{(1)} = (x_1, \dots, x_{d_1})$. If $f^{(1|2)}(x^{(1)} | z) = f^{(1)}(x^{(1)})$, then $X^{(1)}$ and $X^{(2)}$ are independent.

The conditional probability $P(A | B)$ in (2.5) provides a heuristic interpretation for the conditional density in (2.35). Since $P(A | B) \simeq f^{(1|2)}(x^{(1)} | z) dx^{(1)}$ for $A = \{X_1 \in (x_1, x_1 + dx_1], \dots, X_{d_1} \in (x_{d_1}, x_{d_1} + dx_{d_1}]\}$ and $B = \{X_{d_1+1} \in (z_1, z_1 + dz_1], \dots, X_d \in (z_{d_2}, z_{d_2} + dz_{d_2}]\}$, $f^{(1|2)}(x^{(1)} | z) dx^{(1)}$ represents the probability that $X^{(1)}$ is in the infinitesimal rectangle $(x_1, x_1 + dx_1] \times \dots \times (x_{d_1}, x_{d_1} + dx_{d_1}]$ under the condition $X^{(2)} = z$. A rigorous discussion on this topic can be found in [5] (Sect. 21.3, pp. 416–417).

Definition 2.34 An \mathbb{R}^d -valued random variable X is said to be Gaussian with mean vector μ and covariance matrix γ if it has the density

$$f(x) = [(2\pi)^d \det(\gamma)]^{-1/2} \exp\left[-\frac{1}{2}(x - \mu)' \gamma^{-1}(x - \mu)\right], \quad x \in \mathbb{R}^d, \quad (2.36)$$

where $(\cdot)'$ denotes matrix transposition. We use the notation $X \sim N(\mu, \gamma)$ to indicate that X has this property. If $d = 2$, $\mu_1 = \mu_2 = 0$, $\gamma_{1,1} = \gamma_{2,2} = 1$, and $\gamma_{1,2} = \gamma_{2,1} = \rho$, $\rho \in (-1, 1)$, the density f is denoted by $\phi(\cdot, \cdot; \rho)$ and has the expression

$$\phi(x_1, x_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right], \quad (x_1, x_2) \in \mathbb{R}^2, \quad (2.37)$$

and is referred to as the density of the standard bivariate Gaussian vector. The parameter ρ is the correlation coefficient between the coordinates of X .

Example 2.41 Let X be a bivariate random vector with the density in (2.37). The conditional density of $X_1 \mid (X_2 = z)$ is

$$f^{(1|2)}(x_1 \mid z) = \frac{\phi(x_1, z; \rho)}{\phi(z)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{x_1^2 - 2\rho x_1 z + z^2}{2(1-\rho^2)} + \frac{z^2}{2}\right]. \quad (2.38)$$

This density becomes $f^{(1|2)}(x_1 \mid z) = \phi(x_1)$ for $\rho = 0$ showing that X_1 and X_2 are independent for this value of ρ . \diamond

Example 2.42 Let X be an \mathbb{R}^d -valued random variable with density f_x . Let $Y = g(X)$ where $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a measurable function defining a one-to-one mapping between X and Y . The density f_y of Y is given by

$$f_y(y) = f_x(h(y))|J|, \quad x, y \in \mathbb{R}^d, \quad (2.39)$$

where $J = \{\partial x_i / \partial y_j, i, j = 1, \dots, d\}$ denotes the Jacobian matrix.

Proof Since $x \mapsto y = g(x)$ is a one-to-one mapping, J is nonzero everywhere and so is the Jacobian of the inverse mapping $y \mapsto x = g^{-1}(y)$. Let D_x be a neighborhood of $x \in \mathbb{R}^d$ and $D_y = \{\eta \in \mathbb{R}^d : \eta = g(\xi), \xi \in D_x\}$ denote the image of D_x by the transformation $x \mapsto y = g(x)$. The equality $P(X \in D_x) = P(Y \in D_y)$ can be written as $\int_{D_x} f_x(\xi) d\xi = \int_{D_y} f_y(\eta) d\eta$, or

$$\int_{D_y} f_x(g^{-1}(\eta)) \left| \frac{\partial(\xi_1, \dots, \xi_d)}{\partial(\eta_1, \dots, \eta_d)} \right| d\eta = \int_{D_y} f_y(\eta) d\eta$$

by a change of variables. The last equality gives (2.39).

If the mapping $y \mapsto x = g^{-1}(y)$ has multiple solutions, we can construct a partition $\{A_v\}$ of the x -space such that the mapping $y \mapsto x$ is one-to-one in each A_v . The probability mass of D_y is equal to the sum $\sum_v f_v |J_v|$ of the corresponding contributions of the subsets A_v , where each term $f_v |J_v|$ is equal to the right side of (2.39) for the restriction of $y \mapsto x = g^{-1}(y)$ to A_v . \blacktriangle

Example 2.43 Let $X \sim N(0, 1)$ and $Y = \cos(X)$. The distribution of Y is $P(Y \leq y) = \sum_{k \in \mathbb{Z}} (\Phi(2k\pi + \cos^{-1}(y)) - \Phi(2k\pi - \cos^{-1}(y)))$ for $|y| \leq 1$ since $\{Y \leq y\}$ if X belongs to $\cup_{k \in \mathbb{Z}} [2k\pi - \cos^{-1}(y), 2k\pi + \cos^{-1}(y)]$. \diamond

Example 2.44 Let $X \sim N(0, \rho)$ be an \mathbb{R}^d -valued standard Gaussian variable with $\rho_{ii} = 1$, and set $Y_i = F_i^{-1} \circ \Phi(X_i)$, where F_i are continuous distributions with densities f_i , $i = 1, \dots, d$. The density of $Y = (Y_1, \dots, Y_d) \in \mathbb{R}^d$ is

$$f_y(y_1, \dots, y_d) = [(2\pi)^d \det(\rho)]^{-1/2} \exp\left(-\frac{1}{2} x' \rho^{-1} x\right) \prod_{i=1}^d \frac{f_i(y_i)}{\phi(x_i)}, \quad (2.40)$$

where $x_i = \Phi^{-1} \circ F_i(y_i)$, $i = 1, \dots, d$. The non-Gaussian vector Y is said to be a translation random vector ([6], Sect. 3.1.1). \diamond

Proof The equality $P(\cap_{i=1}^d Y_i \leq y_i) = P(\cap_{i=1}^d X_i \leq x_i)$ holds since the mappings $y_i \mapsto x_i = \Phi^{-1} \circ F_i(y_i)$ are invertible. This gives (2.40) by differentiation. Note that (2.40) follows directly from (2.39) since $|J| = \prod_{i=1}^d f_i(y_i)/\phi(x_i)$. \blacktriangle

2.10 Characteristic Function

The characteristic function defines completely the probability law of random variables. We describe random variables by their distributions or characteristic functions depending on the objective of the analysis.

Definition 2.35 Let X be an \mathbb{R}^d -valued random variable with distribution F . The characteristic function of X is

$$\varphi(u) = E[e^{iu'X}] = \int_{\mathbb{R}^d} e^{iu'x} dF(x) = \int_{\mathbb{R}^d} e^{iu'x} f(x) dx, \quad u \in \mathbb{R}^d, \quad (2.41)$$

where f is the density of X , provided it exists and u' denotes the transpose of $u \in \mathbb{R}^d$. The expectation of the complex-valued random variable $\exp(iu'X)$ is obtained from the expectations of its real and imaginary parts, that is, $E[\exp(iu'X)] = E[\cos(u'X)] + iE[\sin(u'X)]$. The characteristic function is always defined.

Suppose first that X is a real-valued random variable. Following is a list of properties of the characteristic function of X that are useful for calculations: $|\varphi(u)| \leq \varphi(0) = 1$ for all $u \in \mathbb{R}$; $\varphi(-u) = \varphi(u)^*$, where z^* denotes the complex conjugate of $z \in \mathbb{C}$, φ is positive definite (Exercise 2.23); the characteristic and the density functions are Fourier pairs, that is,

$$\varphi(u) = \int_{\mathbb{R}} e^{iux} f(x) dx \quad \text{and} \quad f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \varphi(u) du;$$

φ is uniformly continuous in \mathbb{R} ; and, if $X \in L^q$, then $\varphi \in C^q(\mathbb{R})$ and $\varphi^{(k)}(0) = i^k E[X^k]$ for $k = 1, \dots, q$. Most of these properties result from the definition of φ but the proof of some properties requires technical arguments ([7], Sect. 2.10.3, [11], Sects. 9.2–9.5).

Example 2.45 Consider a Cauchy random variable X with density $f(x) = a/[\pi(a^2 + x^2)]$, $a > 0$, $x \in \mathbb{R}$ (Example 2.39). The characteristic function of X exists and is $\varphi(u) = \exp(-a|u|)$, $u \in \mathbb{R}$. However, X has no mean since $\varphi(u)$ is not differentiable at $u = 0$. \diamond

Example 2.46 Let $F(x) = \sum_{k \in I} p_k 1(x_k \leq x)$ be the distribution of a real-valued random variable X , where I is a finite index set, $p_k \geq 0$ such that $\sum_{k \in I} p_k = 1$, and $\{x_k\}$ is an increasing sequence of real numbers. The characteristic function of X is $\varphi(u) = \sum_{k \in I} p_k [\cos(ux_k) + i \sin(ux_k)]$ so that it does not vanish as $|u| \rightarrow \infty$. \diamond

Example 2.47 Let $X = \sum_{k=1}^N Y_k$ be a compound Poisson random variable, where N is a Poisson variable with intensity $\lambda > 0$ and Y_1, Y_2, \dots are iid random variables that are independent of N . The characteristic function of X is

$$\varphi(u) = \exp\left(-\lambda \int_{\mathbb{R}} (1 - e^{iu y}) dF_Y(y)\right) = \exp[-\lambda(1 - \varphi_Y(u))], \quad u \in \mathbb{R}, \quad (2.42)$$

where F_Y and φ_Y denote the distribution and the characteristic functions of Y_1 . \diamond

Proof We have $\varphi(u) = E[e^{iuX} 1(N \geq 1) + e^{iuX} 1(N = 0)] = E[e^{iu \sum_{k=1}^N Y_k} 1(N \geq 1)] + P(N = 0)$ and $E[e^{iu \sum_{k=1}^N Y_k} 1(N \geq 1)] = \sum_{k=1}^{\infty} (\varphi_Y(u))^k P(N = k)$. This gives $\varphi(u) = e^{-\lambda} \sum_{k=0}^{\infty} (\lambda \varphi_Y(u))^k / k! = \exp[-\lambda(1 - \varphi_Y(u))]$ since $P(N = n) = e^{-\lambda} \lambda^n / n!$, $n = 0, 1, \dots$, is the probability of the Poisson variable N . \blacktriangle

Suppose now that X is an \mathbb{R}^d -valued random variable and $d > 1$. Following is a list of relevant properties of the characteristic function φ of X : $|\varphi(u)| \leq \varphi(0) = 1$ for all $u \in \mathbb{R}^d$; the characteristic function and the density of X are Fourier pairs, that is,

$$\varphi(u) = \int_{\mathbb{R}^d} e^{iu'x} f(x) dx \quad \text{and} \quad f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iu'x} \varphi(u) du;$$

φ is uniformly continuous; and $\varphi(u) = \prod_{k=1}^d \varphi_k(u_k)$ if X has independent coordinates, where $\varphi_k(u_k) = E[\exp(iu_k X_k)]$.

Example 2.48 Let $X \sim N(\mu, \gamma)$ be an \mathbb{R}^d -valued Gaussian variable with density given by (2.36). The characteristic function of X is

$$\varphi(u) = \exp\left(iu'\mu - \frac{1}{2}u'\gamma u\right), \quad u \in \mathbb{R}^d. \quad (2.43)$$

If γ is a diagonal matrix, that is, the coordinates of X are uncorrelated, they are also independent since $\varphi(u) = \prod_{k=1}^d \varphi_k(u_k)$, where $\varphi_k(u_k) = E[\exp(iu_k X_k)] = \exp(iu_k \mu_k - \gamma_{kk} u_k^2 / 2)$ and $X_k \sim N(\mu_k, \gamma_{kk})$. Generally, lack of correlation does not imply independence. However, independence and lack of correlation are equivalent for Gaussian variables. \diamond

Example 2.49 Let $X \sim N(\mu, \gamma)$ be an \mathbb{R}^d -valued Gaussian variable and set $Y = aX + b$, where a and b are (d', d) and $(d', 1)$ matrices with constant entries. Then Y is an $\mathbb{R}^{d'}$ -valued Gaussian variable with mean $a\mu + b$ and covariance matrix $a\gamma a'$, that is, linear transformations of Gaussian vectors are Gaussian vectors. \diamond

Proof The mean vector and the covariance matrix of Y can be obtained by direct calculations using the definition of Y and the linearity of the expectation operator. The characteristic function of Y is

$$E[e^{iv'Y}] = E[e^{iv'aX}] e^{iv'b} = \exp\left(iu'\mu - \frac{1}{2}u'\gamma u\right) e^{iv'b}, \quad v \in \mathbb{R}^{d'}, \quad u = a'v \in \mathbb{R}^d,$$

so that Y is an $\mathbb{R}^{d'}$ -valued Gaussian variable with the stated properties. \blacktriangle

Example 2.50 Let $X \sim N(\mu, \gamma)$ be an \mathbb{R}^d -valued random variable and denote the first $d_1 < d$ and the last $d_2 = d - d_1$ coordinates of X by $X^{(1)}$ and $X^{(2)}$, respectively. The conditional vector $\hat{X} = X^{(1)} \mid (X^{(2)} = z)$ is Gaussian with mean vector $\hat{\mu}$ and covariance matrix $\hat{\gamma}$ given by

$$\begin{aligned}\hat{\mu} &= \mu^{(1)} + \gamma^{(1,2)}(\gamma^{(2,2)})^{-1}(z - \mu^{(2)}) \quad \text{and} \\ \hat{\gamma} &= \gamma^{(1,1)} - \gamma^{(1,2)}(\gamma^{(2,2)})^{-1}\gamma^{(2,1)},\end{aligned}\tag{2.44}$$

where $\mu^{(r)} = E[X^{(r)}]$ and $\gamma^{(r,s)} = E[(X^{(r)} - \mu^{(r)})(X^{(s)} - \mu^{(s)})']$, $r, s = 1, 2$. \diamond

Example 2.51 Let $X \sim N(\mu, \gamma)$ be a bivariate Gaussian vector. The random variable,

$$\hat{X} = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(X_2 - \mu_2),\tag{2.45}$$

is the optimal, mean square, linear estimator for X_1 given X_2 , where $\sigma_1^2 = \gamma_{1,1}$, $\sigma_2^2 = \gamma_{2,2}$, and $\rho\sigma_1\sigma_2 = \gamma_{1,2}$. \diamond

Proof Let $Z = aX_2 + b$, $a, b \in \mathbb{R}$, be a linear estimator for X_1 , and impose the conditions that Z is unbiased and minimizes the mean square error $E[(Z - X_1)^2]$. The first condition implies $E[Z] = \mu_1$ so that $a\mu_2 + b = \mu_1$ and $Z = a(X_2 - \mu_2) + \mu_1$. The mean square error

$$E[(Z - X_1)^2] = E[(a(X_2 - \mu_2) - (X_1 - \mu_1))^2] = a^2\sigma_2^2 + \sigma_1^2 - 2a\rho\sigma_1\sigma_2$$

of the estimator Z takes its minimum value at $a = \rho\sigma_1/\sigma_2$. \blacktriangle

2.11 Conditional Expectation

Consider a real-valued random variable X defined on a probability space (Ω, \mathcal{F}, P) such that $E[|X|] < \infty$, that is, the mean of X exists and is finite. Our objective is to define the conditional expectation $E[X \mid \mathcal{G}]$ of X with respect to a sub- σ -field \mathcal{G} of \mathcal{F} . The expectation $E[X \mid \mathcal{G}]$ can be viewed as a local average of X since \mathcal{G} is coarser than \mathcal{F} .

Example 2.52 Let X be an \mathbb{R}^2 -valued Gaussian variable with mean zero and covariances $E[X_i^2] = 1$, $i = 1, 2$, and $E[X_1X_2] = \rho$, $|\rho| < 1$. The joint density of X and the density of $X_1 \mid (X_2 = z)$ are given by (2.37) and (2.38). These densities show that $X_1 \mid (X_2 = z) \sim N(\rho z, 1 - \rho^2)$ so that $E[X_1 \mid X_2 = z] = \rho z$, a result that can also be obtained from (2.44). \diamond

Example 2.53 The sample space, the σ -field \mathcal{F} , and the probability measure for the experiment of rolling two dice are $\Omega = \{\omega = (i, j) : i, j = 1, \dots, 6\}$, the

collection of all parts of Ω , and $P(\{\omega\}) = 1/36$, respectively. Let X be a random variable defined on this space by $X(\omega) = i + j$. The expectation of X is $E[X] = 7$. Let $\Lambda_n = \{\omega = (i, j) : i \wedge j = n\}$, $n = 1, \dots, 6$, be measurable sets partitioning Ω , for example, $\Lambda_4 = \{(4, 4), (4, 5), (5, 4), (4, 6), (6, 4)\}$. The probability that an outcome (i, j) is in Λ_n is equal to the cardinality of Λ_n divided by 36. Let $\mathcal{G} = \sigma(\Lambda_1, \dots, \Lambda_6)$ denote the σ -field generated by $\{\Lambda_n, n = 1, \dots, 6\}$, so that the members of \mathcal{G} are unions of members of $\{\Lambda_n, n = 1, \dots, 6\}$.

The local averages $E[X | \Lambda_n]$ of X over Λ_n can be calculated simply. For example, $E[X | \Lambda_4] = ((4 + 4) + (4 + 5) + (5 + 4) + (4 + 6) + (6 + 4))(1/5) = 46/5$ by the definition of X and the fact that the members of Λ_n are equally likely. Similar calculations give the other local averages, for example, $E[X | \Lambda_6] = 12$. In general, we have

$$E[X | \Lambda_n] = \sum_{\omega \in \Lambda_n} X(\omega) \frac{1}{\text{card}(\Lambda_n)} = \frac{1}{\text{card}(\Lambda_n)/36} \sum_{\omega \in \Lambda_n} X(\omega) \frac{1}{36} = \frac{1}{P(\Lambda_n)} \int_{\Lambda_n} X dP.$$

Hence, the conditional expectation of X with respect to \mathcal{G} is a simple random variable denoted by $E[X | \mathcal{G}]$, that takes the values $\{E[X | \Lambda_n]\}$ with probabilities $\{P(\Lambda_n)\}$, that is, the random variable

$$E[X | \mathcal{G}] = \sum_{n=1}^6 E[X | \Lambda_n] 1_{\Lambda_n}.$$

This random variable is \mathcal{G} -measurable and has the properties $\int_A E[X | \mathcal{G}] dP = \int_A X dP$ for all $A \in \mathcal{G}$ and $E\{E[X | \mathcal{G}]\} = E[X]$. \diamond

Proof It is obvious that $E[X | \mathcal{G}]$ is a random variable on (Ω, \mathcal{G}, P) . The members of \mathcal{G} are union of Λ_n , so that if, for example, $A = \Lambda_k \cup \Lambda_l$, $k \neq l$, we have $\int_A E[X | \mathcal{G}] dP = \sum_{n=1}^6 E[X | \Lambda_n] \int_{\Lambda_k \cup \Lambda_l} 1_{\Lambda_n} dP = E[X | \Lambda_k]P(\Lambda_k) + E[X | \Lambda_l]P(\Lambda_l)$. Direct calculations give $\int_A X dP = \int_A E[X | \mathcal{G}] dP$ for $A \in \mathcal{G}$ arbitrary. We also have $E\{E[X | \mathcal{G}]\} = \sum_{n=1}^6 E[X | \Lambda_n]P(\Lambda_n) = \sum_{n=1}^6 \int_{\Lambda_n} X dP = \int_{\Omega} X dP = E[X]$. \diamond

Example 2.54 Let X and Y be real-valued random variables defined on a probability space (Ω, \mathcal{F}, P) . Suppose Y is discrete taking distinct values y_i , $i = 1, 2, \dots$, so that the sets $B_i = \{Y = y_i\}$ partition Ω . If $P(B_i) > 0$, the expectation of X conditional on Y is a discrete random variable denoted by $E[X | Y]$ taking the values $E[X | Y](\omega) = E[X | B_i]$ for $\omega \in B_i$, where $E[X | B_i] = E[X | Y = y_i] = \int_{B_i} x dF(x)/P(B_i)$ and F denotes the distribution of X . \diamond

Proof The distribution of X conditional on B_i is $F(x | B_i) = P(X \leq x, B_i)/P(B_i)$ so that the conditional expectation of X given B_i can be calculated from $E[X | B_i] = \int x dF(x | B_i) = \int_{B_i} x dF(x)/P(B_i)$.

The conditional expectation $E[X | Y]$ is equal to the local average of X over B_i . It constitutes a coarser version of X that can be viewed as an approximation of this random variable. Since the measurable partition $\{B_i\}$ of Ω generates the

σ -field $\sigma(Y)$, we may write $E[X | \sigma(Y)]$ for $E[X | Y]$. The conditional expectation $E[X | \sigma(Y)]$ has the same properties as $E[X | \mathcal{G}]$ in the previous example. \blacktriangle

It is not possible to extend the definition of $E[X | Y]$ in Example 2.54 to continuous random variables Y since, for example, $\{\omega \in \Omega : a < Y(\omega) \leq b\}$, $a < b$, does not belong to the σ -field generated by the sets $\{\omega \in \Omega : Y(\omega) = y\}$, $y \in \mathbb{R}$. In agreement with an observation in this example, we set $E[X | Y]$ to be the conditional expectation $E[X | \sigma(Y)]$.

Definition 2.36 Let X be a real-valued integrable random variable defined on a probability space (Ω, \mathcal{F}, P) , and let \mathcal{G} be a sub- σ -field of \mathcal{F} . The conditional expectation $E[X | \mathcal{G}]$ of X with respect to \mathcal{G} is the class of \mathcal{G} -measurable functions satisfying the defining relation

$$\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}] dP, \quad \forall \Lambda \in \mathcal{G}. \quad (2.46)$$

Note that $E[X | \mathcal{G}]$ exists ([4], Theorem 9.1.1) and is such that $E[X | \mathcal{G}] = E[X]$ for $\mathcal{G} = \{\emptyset, \Omega\}$, $E[X | \mathcal{G}] = X$ for $\mathcal{G} = \mathcal{F}$ (Exercise 2.32), and $E\{E[X | \mathcal{G}]\} = E[X]$ by (2.46) with $\Lambda = \Omega$.

Theorem 2.13 Let X be a real-valued integrable random variable defined on a probability space (Ω, \mathcal{F}, P) , \mathcal{G} a sub- σ -field of \mathcal{F} , and Z a real-valued \mathcal{G} -measurable function. Then

$$\begin{aligned} E[(X - E[X | \mathcal{G}])Z] &= 0, \quad \forall Z \in \mathcal{G} \\ E[XZ | \mathcal{G}] &= ZE[X | \mathcal{G}] \quad \text{a.s., } \forall Z \in \mathcal{G}. \end{aligned} \quad (2.47)$$

Proof The first equality in (2.47) holds for $Z = 1_{\Delta}$, $\Delta \in \mathcal{G}$, by the defining relation. It also holds for simple random variables $Z = \sum_n b_n 1_{\Delta_n}$, $\Delta_n \in \mathcal{G}$, since expectation is a linear operator.

The random variables $E[XZ | \mathcal{G}]$ and $ZE[X | \mathcal{G}]$ in the second equality of (2.47) are \mathcal{G} -measurable. If $Z = 1_{\Delta}$, $\Delta \in \mathcal{G}$, (2.47) holds a.s., because for $\Lambda \in \mathcal{G}$ the left and the right sides of this equation are $\int_{\Lambda} E[X 1_{\Delta} | \mathcal{G}] dP = \int_{\Lambda} X 1_{\Delta} dP = \int_{\Lambda \cap \Delta} X dP$ and $\int_{\Lambda} 1_{\Delta} E[X | \mathcal{G}] dP = \int_{\Lambda \cap \Delta} E[X | \mathcal{G}] dP = \int_{\Lambda \cap \Delta} X dP$, respectively, by the defining relation. This equality also holds for simple random variables Z by the linearity of conditional expectation.

The extension of (2.47) to an arbitrary random variable Z results from the representation of Z by a difference of two positive random variables, which can be defined as limits of simple random variables ([4], Sect. 9.1). \blacktriangle

Corollary 2.1 The conditional expectation $E[X | \mathcal{G}]$ is the projection of X on \mathcal{G} and $X - E[X | \mathcal{G}]$ is orthogonal to \mathcal{G} . Moreover, $E[X | \mathcal{G}]$ represents the best mean square (m.s.) estimator for X given the information content of \mathcal{G} .

Proof That $X - E[X | \mathcal{G}]$ is orthogonal to \mathcal{G} follows from the first equality in (2.47). Since $E[X | \mathcal{G}]$ is the orthogonal projection of X on $\{Z : Z \in \mathcal{G}\}$, it is the best m.s. estimator for X given \mathcal{G} ([8], Sects. 4.3 and 4.4). \blacktriangle

Example 2.55 The conditional expectation $E[X \mid \sigma(Z)] = E[X \mid Z]$ is the best m.s. estimator of X with respect to the information content of $\sigma(Z)$, where $X, Z \in L^2(\Omega, \mathcal{F}, P)$. The best m.s. linear estimator of X is

$$\hat{X} = E[X] + \frac{E[XZ] - E[X]E[Z]}{E[Z^2] - E[Z]^2} (Z - E[Z]). \quad (2.48)$$

The estimator represents the linear regression of X with respect to Z , and becomes $\hat{X} = E[X]$ if X and Z are uncorrelated. \diamond

Proof The function $\hat{X} = aZ + b$ is $\sigma(Z)$ -measurable for any constants a, b . It is the conditional expectation of X with respect to $\sigma(Z)$ if $E[X1_\Lambda] = E[(aZ + b)1_\Lambda]$ for all $\Lambda \in \sigma(Z)$ by (2.46), which gives $E[X] = E[\hat{X}] = aE[Z] + b$ for $\Lambda = \Omega$. The orthogonality condition in (2.47) implies $E[(X - \hat{X})Z] = 0$ or $E[XZ] = aE[Z^2] + bE[Z]$. The solutions a, b of these equations introduce in $\hat{X} = aZ + b$ give the expression of \hat{X} in (2.48). The resulting estimator has the property that its m.s. error $E[(\hat{X} - X)^2]$ is smaller than the error $E[(aZ + b - X)^2]$ for all the other values of a and b . \blacktriangle

The following three theorems give properties of the conditional expectations that are useful for calculations. The properties listed below are similar to those of expectation and hold a.s. As indicated at the beginning of this section, we consider real-valued random variables defined on the same probability space (Ω, \mathcal{F}, P) and a sub- σ -field \mathcal{G} of \mathcal{F} .

Theorem 2.14 *If X and X_n are integrable random variables, then ([11], Sect. 10.3)*

$$\begin{aligned} X \in \mathcal{G} &\text{ implies } E[X \mid \mathcal{G}] = X, \\ E[aX_1 + bX_2 \mid \mathcal{G}] &= aE[X_1 \mid \mathcal{G}] + bE[X_2 \mid \mathcal{G}] \quad (\text{linearity}), \\ X_1 \leq X_2 &\text{ implies } E[X_1 \mid \mathcal{G}] \leq E[X_2 \mid \mathcal{G}] \quad (\text{monotonicity}), \\ |E[X \mid \mathcal{G}]| &\leq E[|X| \mid \mathcal{G}] \quad (\text{modulus inequality}), \\ X_n \uparrow (\downarrow) X &\text{ implies } E[X_n \mid \mathcal{G}] \uparrow (\downarrow) E[X \mid \mathcal{G}] \quad (\text{monotone convergence}), \\ |X_n| \leq Y \text{ a.s., } E[Y] < \infty \text{ a.s., and } X_n \rightarrow X &\text{ imply } E[X_n \mid \mathcal{G}] \rightarrow E[X \mid \mathcal{G}] \\ &\quad (\text{dominated convergence}). \end{aligned} \quad (2.49)$$

Theorem 2.15 *The Cauchy-Schwarz and Jensen inequalities are, respectively, ([4], Sect. 9.1)*

$$(E[XY \mid \mathcal{G}])^2 \leq E[X^2 \mid \mathcal{G}]E[Y^2 \mid \mathcal{G}] \quad \text{and} \quad (2.50)$$

$$g(E[X \mid \mathcal{G}]) \leq E[g(X) \mid \mathcal{G}], \quad \text{where } g: \mathbb{R} \rightarrow \mathbb{R} \text{ is a convex function.} \quad (2.51)$$

Theorem 2.16 *If \mathcal{G}_1 and \mathcal{G}_2 are sub- σ -fields of \mathcal{F} such that $\mathcal{G}_1 \subset \mathcal{G}_2$, then ([4], Sect. 9.1)*

$$E[X | \mathcal{G}_1] = E[X | \mathcal{G}_2] \iff E[X | \mathcal{G}_2] \in \mathcal{G}_1 \quad \text{and} \quad (2.52)$$

$$E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} = E[X | \mathcal{G}_1] = E\{E[X | \mathcal{G}_1] | \mathcal{G}_2\}. \quad (2.53)$$

Definition 2.37 Let (Ω, \mathcal{F}, P) be a probability space and \mathcal{G} a sub- σ -field of \mathcal{F} . The conditional probability with respect to \mathcal{G} is

$$P(A | \mathcal{G}) = E[1_A | \mathcal{G}], \quad A \in \mathcal{F}. \quad (2.54)$$

The definition is meaningful since 1_A is \mathcal{F} -measurable. The random variable $P(A | \mathcal{G})$ is integrable, \mathcal{G} -measurable, and satisfies $\int_A P(A | \mathcal{G}) dP = P(A \cap A)$ for all $A \in \mathcal{G}$. The latter equality holds since $\int_A P(A | \mathcal{G}) dP = \int_A E[1_A | \mathcal{G}] dP = \int_A 1_A dP$ by the defining relation in (2.46).

Example 2.56 Let A and B be events on a probability space (Ω, \mathcal{F}, P) such that $P(B) > 0$ and $P(B^c) > 0$, and consider the sub- σ -field $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$ of \mathcal{F} . The conditional probability in (2.54) is

$$\begin{aligned} P(A | \mathcal{G}) &= E[1_A | \mathcal{G}] = E[1_A | B]1_B + E[1_A | B^c]1_{B^c} \\ &= \frac{P(A \cap B)}{P(B)}1_B + \frac{P(A \cap B^c)}{P(B^c)}1_{B^c}, \end{aligned} \quad (2.55)$$

where that latter equality holds by Example 2.53. This shows that $P(A | \mathcal{G})$ is a random variable taking the values $P(A \cap B)/P(B)$ and $P(A \cap B^c)/P(B^c)$ with probabilities $P(B)$ and $P(B^c)$, respectively. Note that $P(A | \mathcal{G})$ extends the definition of the conditional probability in (2.5). \diamond

Example 2.57 Let $X \geq 0$ a.s. be a random variable defined on a probability space (Ω, \mathcal{F}, P) and \mathcal{G} a sub- σ -field of \mathcal{F} . If $E[X] < \infty$, the set function $Q(A) = E[X1_A]$, $A \in \mathcal{G}$, is a finite measure on the measurable space (Ω, \mathcal{G}) . The conditional expectation of X with respect to \mathcal{G} is the Radon–Nikodym derivative $E[X | \mathcal{G}] = dQ/dP$.

This definition becomes $E[X | \mathcal{G}] = E[X^+ | \mathcal{G}] - E[X^- | \mathcal{G}]$ for random variables with $E[|X|] < \infty$ and $E[X | \mathcal{G}] = (E[X_1 | \mathcal{G}], \dots, E[X_d | \mathcal{G}])$ for \mathbb{R}^d -valued random variables with finite mean. \diamond

Proof The defining relation for conditional expectation gives $Q(A) = \int_A X dP = \int_A E[X | \mathcal{G}] dP$, $A \in \mathcal{G}$. Since Q and P are finite measures and Q is absolutely continuous with respect to P , the equality $Q(A) = \int_A E[X | \mathcal{G}] dP$, $A \in \mathcal{G}$, implies $E[X | \mathcal{G}] = dQ/dP$ by Theorem 2.12. \blacktriangle

2.12 Discrete Time Martingales

Martingales are essentials for constructing stochastic integrals. This section provides a primer on discrete time martingales. An example is used to introduce an elementary version of stochastic integrals.

Definition 2.38 Let (Ω, \mathcal{F}) be a measurable space. An increasing collection $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \mathcal{F}_n \subseteq \dots \subseteq \mathcal{F}$ of sub- σ -fields of \mathcal{F} is said to be a filtration in (Ω, \mathcal{F}) . A probability space (Ω, \mathcal{F}, P) endowed with a filtration $(\mathcal{F}_n)_{n \geq 0}$ is called a filtered probability space and is denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, P)$. It is assumed that \mathcal{F}_0 contains all the P -null sets of \mathcal{F} .

Example 2.58 Suppose the sequence $X = (X_1, X_2, \dots)$ gives outcomes of coin tosses. The information content of the σ -field $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ is sufficient to decide whether an event related to the first n tosses has or has not occurred. For example, the event $A = \{\text{at least 2 heads in the first five tosses}\}$ is \mathcal{F}_5 -measurable because we can decide after five tosses whether A has or has not occurred. If $\{\text{tail, tail, head, tail}\}$ is a sample of the first four tosses, the event A remains undecided so that $A \notin \mathcal{F}_4$. \diamond

Definition 2.39 Let (Ω, \mathcal{F}) and (Ψ, \mathcal{G}) be measurable spaces, $(\mathcal{F}_n)_{n \geq 0}$ a filtration on (Ω, \mathcal{F}) , and $X = (X_0, X_1, \dots)$ a sequence of measurable functions from (Ω, \mathcal{F}) to (Ψ, \mathcal{G}) . The sequence X is said to be adapted to the filtration $(\mathcal{F}_n)_{n \geq 0}$ or \mathcal{F}_n -adapted if X_n is \mathcal{F}_n -measurable for each $n \geq 0$. The minimal or natural filtration of $X = (X_0, X_1, \dots)$, that is, the smallest σ -field with respect to which X is adapted, is the filtration $\sigma(X_0, X_1, \dots, X_n)$, $n \geq 0$.

Example 2.59 Let $X = (X_0, X_1, \dots)$ be a real-valued sequence defined on a probability space (Ω, \mathcal{F}, P) and set $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. The sequences $Y_n = g(X_n)$ and $Y_n = \max_{0 \leq i \leq n} \{X_i\}$ are \mathcal{F}_n -adapted, where $g : \mathbb{R} \rightarrow \mathbb{R}$ is a Borel measurable function. \diamond

Definition 2.40 Let $X = (X_0, X_1, X_2, \dots)$ be a sequence of real-valued random variables defined on a probability space (Ω, \mathcal{F}, P) endowed with a filtration $(\mathcal{F}_n)_{n \geq 0}$. The sequence X_0, X_1, X_2, \dots is said to be an \mathcal{F}_n -martingale if (1) $E[|X_n|] < \infty$, $n \geq 0$, (2) X is \mathcal{F}_n -adapted, and (3) $E[X_n | \mathcal{F}_m] = X_m$ for $0 \leq m \leq n$.

If the equality in the third condition is replaced by \geq and \leq , then X is said to be an \mathcal{F}_n -submartingale and \mathcal{F}_n -supermartingale, respectively. If the random variables X_n are in $L^p(\Omega, \mathcal{F}, P)$, X is called a p -integrable martingale, submartingale, or supermartingale. If $p = 2$, then X is said to be a square integrable martingale, submartingale, or supermartingale.

Example 2.60 Let $R_n = \sum_{i=1}^n X_i$, $n \geq 1$, and $R_0 = 0$ be a random walk, where $\{X_i\}$ are iid random variables. If the random variables X_i have finite mean, $R = (R_0, R_1, R_2, \dots)$ is an \mathcal{F}_n -supermartingale, martingale, or submartingale depending on the sign of expectation $E[X_1]$, where $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, $n = 1, 2, \dots$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. \diamond

Proof Note that (R_0, R_1, \dots, R_n) and (R_1, \dots, R_n) are \mathcal{F}_n -measurable for $n \geq 1$ so that $R = (R_0, R_1, R_2, \dots)$ is \mathcal{F}_n -adapted. Also, for $n \geq m \geq 0$, we have $E[R_n | \mathcal{F}_m] = R_m + \sum_{i=m+1}^n E[X_i] = R_m + (n - m)E[X_1]$ so that R is a supermartingale, martingale, or submartingale if $E[X_1]$ is negative, zero, or positive. \blacktriangle

Example 2.61 Let $R = (R_0, R_1, \dots)$ be as in Example 2.60. If the random variables X_i have finite variance and mean zero, the sequence $S_n = R_n^2 = \sum_{i,j=1}^n X_i X_j$, $n \geq 1$, with $S_0 = 0$ is an \mathcal{F}_n -submartingale and $S_n - nE[X_1^2]$ is an \mathcal{F}_n -martingale. \diamond

Proof The sequences S_n and $S_n - nE[X_1^2]$ have the first two defining properties for martingales. For the third property, note that

$$\begin{aligned} E[S_{n+1} | \mathcal{F}_n] &= E[(R_n + X_{n+1})^2 | \mathcal{F}_n] = E[R_n^2 + 2R_n X_{n+1} + X_{n+1}^2 | \mathcal{F}_n] \\ &= R_n^2 + 2R_n E[X_{n+1}] + E[X_{n+1}^2] = S_n + E[X_1^2] \geq S_n \end{aligned}$$

since R_n is \mathcal{F}_n -measurable, X_{n+1} is independent of \mathcal{F}_n , and $E[X_{n+1}] = 0$. Hence, $S = (S_0, S_1, \dots)$ is a submartingale. Since $E[S_{n+1} - (n+1)E[X_1^2] | \mathcal{F}_n] = (S_n + E[X_1^2]) - (n+1)E[X_1^2] = S_n - nE[X_1^2]$, $S_n - nE[X_1^2]$ is a martingale. \blacktriangle

Following are martingale properties resulting from their definition, Doob's decomposition, an elementary construction of stochastic integrals, two martingale inequalities, and a brief discussion on stopped martingales.

Let $X = (X_0, X_1, \dots)$ be a sequence of random variables defined on a probability space (Ω, \mathcal{F}, P) with a filtration $(\mathcal{F}_n)_{n \geq 0}$. Then (1) if X is a submartingale, martingale, and supermartingale, its expectation is an increasing, constant, and decreasing function of time, (2) X is a martingale if it is both a submartingale and a supermartingale, (3) if X is a submartingale, then $-X$ is a supermartingale, and (4) the third defining condition for martingales can be replaced with $E[X_{n+1} | \mathcal{F}_n] = X_n$, $n \geq 0$.

Theorem 2.17 Let $X = (X_0, X_1, \dots)$ be a martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, P)$. The series $Y = (Y_0, Y_1, \dots)$ defined by $Y_0 = X_0 - E[X_0]$ and $Y_n = X_n - X_{n-1}$, $n \geq 1$, is orthogonal, that is, $E[Y_m Y_n] = 0$ for $m \neq n$.

Proof The properties of X imply that Y_n is integrable, \mathcal{F}_n -measurable, and satisfies $E[Y_n | \mathcal{F}_m] = 0$, $n > m$. For $n > m$, we have $E[Y_n Y_m] = E\{E[Y_n Y_m | \mathcal{F}_m]\} = E\{Y_m E[Y_n | \mathcal{F}_m]\} = 0$ since Y_m is \mathcal{F}_m -measurable and $E[Y_n | \mathcal{F}_m] = 0$. \blacktriangle

Theorem 2.18 (Doob decomposition) If $X = (X_0, X_1, \dots)$ is an \mathcal{F}_n -submartingale, then there is an \mathcal{F}_n -martingale M_n and an increasing process A_n with the properties $A_0 = 0$ and $A_n \in \mathcal{F}_{n-1}$, $n \geq 1$, such that the representation

$$X_n = A_n + M_n, \quad n \geq 0, \quad (2.56)$$

holds and is unique. The representation shows that submartingales have a predictable part A_n that can be told ahead of time and an unpredictable part M_n .

Proof We first show that (2.56) is unique provided it exists. Note that $E[X_n | \mathcal{F}_{n-1}] = E[A_n | \mathcal{F}_{n-1}] + E[M_n | \mathcal{F}_{n-1}] = A_n + M_{n-1}$ for $n \geq 1$. Substituting M_{n-1} in this equation with its expression from (2.56), we obtain the recurrence formula $A_n = A_{n-1} + E[X_n | \mathcal{F}_{n-1}] - X_{n-1}$, which defines A uniquely since $A_0 = 0$.

We now show that the decomposition in (2.56) exists, that is, that there are processes A and M with the stated properties. Let A_n , $n = 0, 1, \dots$, be defined by

the above recurrence formula with $A_0 = 0$. Note that $A_n \in \mathcal{F}_{n-1}$ and $A_n \geq A_{n-1}$ since X_n is a submartingale, so that A_n , $n = 0, 1, \dots$, has the stated properties. We also have

$$\begin{aligned} E[M_n | \mathcal{F}_{n-1}] &= E[X_n - A_n | \mathcal{F}_{n-1}] = E[X_n | \mathcal{F}_{n-1}] \\ &\quad - E[A_{n-1} + E[X_n | \mathcal{F}_{n-1}] - X_{n-1} | \mathcal{F}_{n-1}] = -A_{n-1} + X_{n-1} = M_{n-1} \end{aligned}$$

so that M is an \mathcal{F}_n -martingale. \blacktriangle

Example 2.62 Let X_n , $n \geq 0$, be a square integrable martingale. Then X_n^2 is a submartingale that admits the Doob decomposition in (2.56) with $M_n = X_n^2 - A_n$ and $A_n = \sum_{i=1}^n E[X_i^2 - X_{i-1}^2 | \mathcal{F}_{i-1}]$. \diamond

Proof The process X_n^2 satisfies the first two defining conditions for martingales, and

$$\begin{aligned} E[X_n^2 | \mathcal{F}_{n-1}] &= E[(X_n - X_{n-1})^2 + 2X_n X_{n-1} - X_{n-1}^2 | \mathcal{F}_{n-1}] \\ &= E[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}] + X_{n-1}^2 \geq X_{n-1}^2 \end{aligned}$$

since $E[(X_n - X_{n-1})^2 | \mathcal{F}_{n-1}] \geq 0$, X_{n-1} is \mathcal{F}_{n-1} -measurable, and X_n is a martingale. Hence, X_n^2 is a submartingale, so that (2.56) holds with $\{A_n\}$ given by $A_n = A_{n-1} + E[X_n^2 | \mathcal{F}_{n-1}] - X_{n-1}^2$, $n \geq 1$, and $A_0 = 0$. \blacktriangle

Example 2.63 Let X_n denote our fortune after n rounds of a game with unit stake. Suppose $m \geq 1$ rounds have been completed in this game, so that $X_n - X_m$, $n > m$, gives our net total winnings/losses in the future $m+1, \dots, n$ rounds. The best m.s. estimator of $X_n - X_m$, $n > m$, given our knowledge \mathcal{F}_m after m rounds is the conditional expectation $E[X_n - X_m | \mathcal{F}_m]$, where $\mathcal{F}_m = \sigma(X_1, \dots, X_m)$, $m \geq 1$, and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. If X_0, X_1, \dots is a martingale, then $E[X_n - X_m | \mathcal{F}_m] = 0$, that is, our average fortune $E[X_n | \mathcal{F}_m]$ at time $n > m$ is equal to our current fortune X_m .

Suppose now that stakes A_i , $i = 0, 1, \dots$, other than one are allowed, where $A_0 = 0$. Since stakes for round $m+1$ are decided based on knowledge \mathcal{F}_m accumulated after m rounds, A_{m+1} is \mathcal{F}_m -measurable. Processes with this property are said to be predictable processes. The sequence $M = (M_0, M_1, M_2, \dots)$ defined by

$$M_n = \sum_{i=1}^n A_i(X_i - X_{i-1}), \quad n = 1, 2, \dots, \quad (2.57)$$

with $M_0 = 0$ and $X_0 = 0$ gives our total fortune after $n \geq 1$ rounds and constitutes a discrete version of the stochastic integral studied later in the book (Chap. 4). The integrand $\{A_i\}$ is a predictable process and the integrator is defined by increments $\{X_i - X_{i-1}\}$ of a martingale. \diamond

Theorem 2.19 Let $X = (X_0, X_1, \dots)$ be a square integrable \mathcal{F}_n -martingale and $A = (A_0, A_1, \dots)$ be an \mathcal{F}_n -predictable process such that $A_0 = 0$ and $E[A_n^2] < \infty$. Then M_n in (2.57) is an \mathcal{F}_n -martingale.

Proof Note that $E[|M_n|] \leq \sum_{i=1}^n E[|A_i||X_i - X_{i-1}|] \leq \sum_{i=1}^n (E[A_i^2]E[(X_i - X_{i-1})^2])^{1/2}$ by the Cauchy–Schwarz inequality, so that $E[|M_n|] < \infty$ since A_i and X_i have finite second moments. That $M_n \in \mathcal{F}_n$ follows from its definition and the properties of A_n and X_n . For $n > m$, we have

$$\begin{aligned} E[M_n | \mathcal{F}_m] &= E\left[M_m + \sum_{i=m+1}^n A_i(X_i - X_{i-1}) \mid \mathcal{F}_m\right] \\ &= M_m + \sum_{i=m+1}^n E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_m] \\ &= M_m + \sum_{i=m+1}^n E\{E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_{i-1}] \mid \mathcal{F}_m\} = M_m \end{aligned}$$

since $A_i \in \mathcal{F}_{i-1}$ and X_i is a martingale so that $E[A_i(X_i - X_{i-1}) \mid \mathcal{F}_{i-1}] = A_i E[X_i - X_{i-1} \mid \mathcal{F}_{i-1}] = 0$. \blacktriangle

Definition 2.41 An $\{0, 1, \dots\}$ -valued random variable T defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 1}, P)$ is a stopping time with respect to \mathcal{F}_n , $n = 0, 1, \dots$, or an \mathcal{F}_n -stopping time if $\{T \leq n\} \in \mathcal{F}_n$ for each $n \geq 0$.

Stopping times are useful for both applications and theoretical considerations. For example, suppose $\{X_n\}$ is the state of a physical system that performs according to specifications as long as its state does not exceed a critical value x_{cr} . The failure time $T = \min\{n : X_n > x_{\text{cr}}\}$ is a stopping time. Stopping times are also useful tools for constructing stochastic integrals (Sect. 4.4.3). Useful information on stopping times can be found in [1] (Sect. 2.2), [7] (Sect. 2.16), [10] (Sect. 1.5), and [13] (Sect. 2.2).

Theorem 2.20 T is a stopping time if and only if $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Proof If T is a stopping time, then $\{T \leq n\} \in \mathcal{F}_n$ and $\{T \leq n-1\}^c \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$ so that $\{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c \in \mathcal{F}_n$. If $\{T = n\} \in \mathcal{F}_n$ for each $n \geq 0$, then $\{T = m\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$ for $m \leq n$ and $\{T \leq n\} = \bigcup_{m=0}^n \{T = m\} \in \mathcal{F}_n$. \blacktriangle

Definition 2.42 Let $X = (X_0, X_1, \dots)$ be an \mathcal{F}_n -submartingale, martingale, or supermartingale and let T denote an \mathcal{F}_n -stopping time. Then $X_n^T(\omega) = X_{n \wedge T(\omega)}(\omega)$, $n = 0, 1, \dots$, is called the sequence X stopped at T . Note that the samples of $\{X_n^T\}$ are constant at times exceeding T .

Theorem 2.21 If T is an \mathcal{F}_n -stopping time and X_n is an \mathcal{F}_n -submartingale, martingale, and supermartingale so is X_n^T .

Proof Since

$$E[|X_{n \wedge T}|] = \sum_{k=0}^n \int_{\{T=k\}} |X_k| dP + \int_{\{T>n\}} |X_n| dP,$$

$E[|X_n|] < \infty$, $\int_{\{T=k\}} |X_k| dP \leq E[|X_k|]$, and $\int_{\{T>n\}} |X_n| dP \leq E[|X_n|]$, X_n^T is integrable. $X_{n \wedge T}$ is \mathcal{F}_n -measurable for all $n \geq 0$ since $X_{n \wedge T} = \sum_{k=0}^{n-1} X_k 1(T = k) + X_n 1(T \geq n)$, $X_k \in \mathcal{F}_k \subseteq \mathcal{F}_n$ for $k \leq n$, $1(T = k) \in \mathcal{F}_k \subseteq \mathcal{F}_n$, and $1(T \geq n) \in \mathcal{F}_n$. The representation $X_{n \wedge T} = \sum_{k=0}^{n-1} X_k 1(T = k) + X_n 1(T \geq n)$ implies $X_{(n+1) \wedge T} - X_{n \wedge T} = (X_{n+1} - X_n) 1(T > n)$ so that $E[X_{(n+1) \wedge T} - X_{n \wedge T} | \mathcal{F}_n] = 1(T > n) E[X_{n+1} - X_n | \mathcal{F}_n]$ since $1(T > n)$ is \mathcal{F}_n -measurable. If X is a submartingale, martingale, or supermartingale, $E[X_{n+1} - X_n | \mathcal{F}_n]$ is positive, zero, or negative, that is, X_n^T is a submartingale, martingale, or supermartingale, respectively. \blacktriangle

Theorem 2.22 (Optional stopping theorem) *If (1) X is an \mathcal{F}_n -martingale, (2) T is a stopping time with respect to \mathcal{F}_n such that $T < \infty$ a.s., (3) X_T is integrable, and (4) $E[X_{n+1} 1(T > n)] \rightarrow 0$ as $n \rightarrow \infty$, then $E[X_T] = \mu$, where $\mu = E[X_0]$.*

Proof Since $X_T = X_{n \wedge T} + (X_T - X_n) 1(T > n)$ and X is a martingale, we have $E[X_T] = \mu + E[X_T 1(T > n)] - E[X_n 1(T > n)]$. The expectation $E[X_n 1(T > n)]$ converges to zero as $n \rightarrow \infty$ by hypothesis. The expectation $E[X_T 1(T > n)] = \sum_{k=n+1}^{\infty} E[X_k 1(T = k)]$ also converges to zero as $n \rightarrow \infty$ since $|E[X_T 1(T > n)]| \leq E[|X_T|]$ and X_T is integrable by assumption so that the series $\sum_{k=0}^{\infty} E[X_k 1(T = k)]$ is convergent. Hence, the expectation of X_T is μ . \blacktriangle

We conclude with two inequalities that are useful in applications. The proof of these and other inequalities can be found in, for example, [5] (Chap 24) and [11] (Chap. 10).

Theorem 2.23 (Doob maximal inequality) *If $X = (X_0, X_1, \dots)$ is a positive \mathcal{F}_n -submartingale and $\lambda > 0$ is an arbitrary constant, then*

$$P\left(\max_{0 \leq k \leq n} X_k \geq \lambda\right) \leq \frac{1}{\lambda} E\left[X_n 1\left(\max_{0 \leq k \leq n} X_k \geq \lambda\right)\right]. \quad (2.58)$$

Theorem 2.24 (Doob maximal L^2 inequality) *If $X = (X_0, X_1, \dots)$ is a square integrable positive \mathcal{F}_n -submartingale, then*

$$E\left[\left(\max_{0 \leq k \leq n} X_k\right)^2\right] \leq 4E[X_n^2]. \quad (2.59)$$

2.13 Monte Carlo Simulation

Let X be an \mathbb{R}^d -valued random variable with distribution F that is defined on a probability space (Ω, \mathcal{F}, P) . Our objectives are to generate independent samples of X and estimate properties of X from its samples.

2.13.1 Gaussian Variables

Let $X \sim N(\mu, \gamma)$ be an \mathbb{R}^d -valued Gaussian variable with mean μ and covariance matrix γ . A useful representation of X is provided by the Cholesky decomposition showing that

$$X \stackrel{d}{=} \mu + \beta G \sim N(\mu, \gamma), \quad (2.60)$$

where G is an \mathbb{R}^d -valued random variable with independent $N(0,1)$ coordinates and β is a lower triangular matrix whose non-zero entries are

$$\beta_{ij} = \frac{\gamma_{ij} - \sum_{r=1}^{j-1} \beta_{ir} \beta_{jr}}{\left[\gamma_{jj} - \sum_{r=1}^{j-1} \beta_{jr}^2 \right]^{1/2}}, \quad 1 \leq j \leq i \leq d, \quad (2.61)$$

with the convention $\sum_{r=1}^0 \beta_{ir} \beta_{jr} = 0$.

Samples of X can be calculated from (2.60) in which G is replaced with samples of this vector generated by, for example, the MATLAB function `randn`. Alternatively, algorithms using memoryless transformations of some random variables can be used to generate independent samples of $N(0, 1)$ variables. For example, $Z_1 = \sqrt{-2 \ln(U_1)} \cos(2\pi U_2)$ and $Z_2 = \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)$ are independent $N(0,1)$ variables, where $U_1, U_2 \sim U(0, 1)$ are independent random variables distributed uniformly in $(0,1)$ [3, 6, 12].

2.13.2 Non-Gaussian Variables

Let X be a real-valued, non-Gaussian random variable with distribution F that is continuous. Samples of X can be calculated from samples of $U(0,1)$ by the following transformation.

Theorem 2.25 *If X is a real-valued random variable with continuous distribution F , then*

$$X \stackrel{d}{=} F^{-1}(U(0, 1)). \quad (2.62)$$

Proof Since $P(F^{-1}(U(0, 1)) \leq z) = P(U(0, 1) \leq F(z)) = F(z)$, (2.62) holds, so that samples of X can be calculated from samples of $U(0,1)$ and the representation of X in (2.62). For example, n independent samples of an exponential random variable with mean $1/\lambda$, $\lambda > 0$, can be obtained from $-\ln(1 - \text{rand}(n, 1))/\lambda \stackrel{d}{=} -\ln(\text{rand}(n, 1))/\lambda$, where `rand` is a MATLAB function generating samples of $U(0,1)$. ▲

Suppose now that X is an \mathbb{R}^d -valued random variable with continuous distribution F . Let F_1 and $F_{k|k-1,\dots,1}$, $k = 2, \dots, d$ denote the distributions of the coordinate X_1 of X and of the conditional random variable $X_k | (X_{k-1}, \dots, X_1)$, respectively.

Theorem 2.26 Let $Z = (Z_1, \dots, Z_d)$ be an \mathbb{R}^d -valued random variable defined by

$$\begin{aligned} F_1(Z_1) &= U_1, \\ F_{k|k-1,\dots,1}(Z_k | Z_{k-1}, \dots, Z_1) &= U_k, \quad k = 2, \dots, d, \end{aligned} \quad (2.63)$$

where $\{U_k, k = 1, \dots, d\}$ are independent $U(0,1)$ random variables. Then, X and Z have the same distribution.

Proof That $P(Z_1 \leq z_1) = F_1(z_1)$ follows from Theorem 2.25. Since $Z_2 | Z_1 \stackrel{d}{=} F_{2|1}^{-1}(U_2 | Z_1)$, we have

$$\begin{aligned} P(Z_2 \leq z_2 | Z_1 = z_1) &= P(F_{2|1}^{-1}(U_2 | z_1) \leq z_2) \\ &= P(U_2 \leq F_{2|1}(z_2 | z_1)) = F_{2|1}(z_2 | z_1), \end{aligned}$$

and so on. \blacktriangle

Let (u_1, \dots, u_d) be a sample of (U_1, \dots, U_d) . The corresponding sample (z_1, \dots, z_d) of Z can be calculated from (2.63) sequentially beginning with $z_1 = F_1^{-1}(u_1)$ and continuing with $F_{k|k-1,\dots,1}(z_k | z_{k-1}, \dots, z_1) = u_k$ for increasing values of $k \geq 2$.

Example 2.64 Let $X = (X_1, X_2)$ be a non-Gaussian vector with $X_1 \sim N(\mu, \sigma^2)$ and $X_2 | (X_1 = x_1) \sim N(x_1, \beta^2)$. The density of X is

$$f(x_1, x_2) = \frac{1}{\sigma\beta} \phi\left(\frac{x_1 - \mu}{\sigma}\right) \phi\left(\frac{x_2 - x_1}{\beta}\right).$$

The mapping in (2.63) becomes $Z_1 = \mu + \sigma\Phi^{-1}(U_1)$ and $Z_2 | (Z_1 = z_1) = z_1 + \beta\Phi^{-1}(U_2)$, where U_1 and U_2 are independent copies of $U(0,1)$. \diamond

Example 2.65 Let $X \in \mathbb{R}^d$ be a translation vector, that is, $X = g(Y)$, where Y is an \mathbb{R}^d -valued Gaussian vector with mean zero, covariance matrix $\rho = \{\rho_{i,j} = E[Y_i Y_j]\}$ such that $\rho_{i,i} = 1$, $i = 1, \dots, d$, and $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Borel measurable. Samples of X can be generated from samples of Y and the definition of X or from samples of $U(0, 1)$, the distribution of X , and the algorithm in (2.63). The latter approach is less efficient for translation random vectors. If the mapping $Y \mapsto X$ is given by $X_i = F_i^{-1}(\Phi(Y_i)) = g_i(Y_i)$, $i = 1, \dots, d$, where F_i are continuous distributions, then the coordinates of X have the distributions F_i . \diamond

Proof Let $y_i = g_i^{-1}(x_i)$ and $y = (y_1, \dots, y_d)$. The distribution,

$$P(X_1 \leq x_1, \dots, X_d \leq x_d) = P(Y_1 \leq y_1, \dots, Y_d \leq y_d) = \Phi_d(y; \rho),$$

called multivariate translation distribution, can be used as input to the Monte Carlo simulation algorithm based on (2.63), where $\Phi_d(\cdot; \rho)$ denotes the joint distribution function of the Gaussian vector $Y \sim N(0, \rho)$. If $X_i = F_i^{-1}(\Phi(Y_i))$, then $P(X_i \leq x_i) = P(Y_i \leq \Phi^{-1}(F_i(x_i))) = F_i(x_i)$, $i = 1, \dots, d$. Details on translation random variables can be found in [6] (Sect. 3.1.1). \blacktriangle

2.13.3 Estimators

Let X be a real-valued random variable with distribution F and $h : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function. Our objective is to estimate the expectation $E[h(X)]$ from n independent samples of X . The expectation $E[h(X)]$ is of interest in applications since it provides useful information on X . For example, if $h(X) = X^r$ and $r \geq 1$ is an integer, then $E[h(X)]$ is the moment of order r of X . If $h(X) = 1(X > a)$, $a \in \mathbb{R}$, then $E[h(X)] = E[1(X > a)] = P(X > a)$.

Theorem 2.27 *Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function and let X_1, \dots, X_n be n independent copies of X such that $E[h(X)^2] < \infty$. The estimator,*

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n h(X_i), \quad (2.64)$$

is unbiased, that is, $E[\hat{Y}] = E[Y]$, and $\text{Var}[\hat{Y}] \rightarrow 0$ as $n \rightarrow \infty$, where $Y = h(X)$.

Proof We have $E[\hat{Y}] = (1/n) \sum_{i=1}^n E[h(X_i)] = E[Y]$ since X_i have the same distribution. Also,

$$E[\hat{Y}^2] = \frac{1}{n^2} \left[nE[h(X_1)^2] + (n^2 - n)(E[h(X_1)])^2 \right]$$

so that $\text{Var}[\hat{Y}] = \text{Var}[Y]/n = \text{Var}[h(X_1)]/n$. The coefficient of variation of estimator \hat{Y} is $\text{cov}[Y] = (\text{Var}[Y])^{1/2}/E[Y] = \text{cov}[h(X_1)]/\sqrt{n}$. \blacktriangle

The estimator \hat{Y} is guaranteed to approximate $E[Y] = E[h(X)]$ accurately for a sufficiently large n . Yet, the required sample size n may be so large that the use of \hat{Y} becomes impractical. For example, suppose our objective is to estimate the probability $P(X > a)$, $X \sim N(0, 1)$, that is, the expectation $E[1(X > a)]$. The mean and variance of \hat{Y} are $P(X > a)$ and $P(X > a)P(X \leq a)/n$, respectively, so that $\text{cov}[\hat{Y}] = \sqrt{P(X \leq a)/(nP(X > a))}$. For $a = 5$ we have $E[\hat{Y}] = E[1(X > a)] = \Phi(-5) = 2.8665 \times 10^{-7}$, $\text{Var}[\hat{Y}] \simeq \Phi(-5)/n$, and $\text{cov}[\hat{Y}] \simeq 1/\sqrt{n\Phi(-5)}$. To have a coefficient of variation of 0.1 we need at least 10^5 samples. A much larger sample size would be needed for a threshold $a > 5$.

More efficient Monte Carlo algorithms, referred here to as improved Monte Carlo algorithms, can be constructed by measure change. Let P and Q be two probabilities on a measurable space (Ω, \mathcal{F}) such that $P \ll Q$, that is, P is absolutely continuous

with respect to Q (Definition 2.29). Then, there exists a positive measurable function $g = dP/dQ : (\Omega, \mathcal{F}) \rightarrow ([0, \infty), \mathcal{B}([0, \infty)))$, called the Radon–Nikodym derivative, such that $P(A) = \int_A g(\omega) Q(d\omega)$, $A \in \mathcal{F}$ (Sect. 2.8). Following are examples illustrating the construction of Monte Carlo simulation algorithms based on the Radon–Nikodym derivative.

Example 2.66 Let $X = \sum_{i=k}^m a_k 1_{A_k}$ be a simple random variable defined on a probability space (Ω, \mathcal{F}, P) , where $\{A_k \in \mathcal{F}, k = 1, \dots, m\}$ is a measurable partition of Ω such that $P(A_k) > 0$ and $|a_k| < \infty$, $k = 1, \dots, m$. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function. The expectation of $h(X)$ with respect to the probability measure P is $E_P[h(X)] = \sum_{k=1}^m h(a_k) P(A_k)$, for example, $E_P[h(X)] = P(X > a)$ for $h(x) = 1(x > a)$.

Consider another probability measure Q on the measurable space (Ω, \mathcal{F}) such that $Q(A_k) > 0$, $k = 1, \dots, m$. We have

$$E_P[h(X)] = \sum_{k=1}^m h(a_k) P(A_k) = \sum_{k=1}^m \left[h(a_k) \frac{P(A_k)}{Q(A_k)} \right] Q(A_k), \quad (2.65)$$

that is, $E_P[h(X)]$ can be calculated as the expectation of random variable $\tilde{X} = \sum_{k=1}^m h(a_k) (P(A_k)/Q(A_k)) 1_{A_k}$ with respect to the probability measure Q . \diamond

Example 2.67 Suppose our objective is to estimate the probability $P(X > a)$ by Monte Carlo simulation, where X is a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) . Let

$$\hat{p}_{MC}(a) = \frac{1}{n} \sum_{i=1}^n 1(x_i > a) \quad \text{and} \quad \hat{p}_{IMC}(a) = \frac{1}{n} \sum_{i=1}^n 1(z_i > a) \frac{f(z_i)}{q(z_i)},$$

be estimates of $P(X > a)$ by direct and improved Monte Carlo simulation, where x_i and z_i are independent samples generated from the densities f and q , respectively, where $f(\xi) = dP(X \leq \xi)/d\xi$, $q(\xi) = dQ(X \leq \xi)/d\xi$, and Q is a measure on (Ω, \mathcal{F}) such that $P \ll Q$.

If $X = \exp(Y)$, $Y \sim N(1, (0.2)^2)$, then $P(X > x)$ is 0.3712, 0.0211, 0.1603 $\times 10^{-3}$, 0.3293 $\times 10^{-4}$, and 0.7061 $\times 10^{-5}$ for $a = 3, 5, 8, 9$, and 10, respectively. The corresponding estimates $\hat{p}_{MC}(x)$ based on 10000 samples are 0.3766, 0.0235, 0.2 $\times 10^{-3}$, 0, and 0. The estimates $\hat{p}_{IMC}(x)$ based on the same number of samples are 0.3733, 0.0212, 0.1668 $\times 10^{-3}$, 0.3304 $\times 10^{-4}$, and 0.7084 $\times 10^{-5}$ for $a = 3, 5, 8, 9$, and 10, where $q(z) = \phi((z - a)/\sigma)/\sigma$ is the density of a Gaussian variable with mean a and variance σ^2 . While the estimators $\hat{p}_{IMC}(a)$ are satisfactory up to $a = 10$, the estimates $\hat{p}_{MC}(a)$ are inaccurate for $a \geq 8$. \diamond

Proof The required probability is $P(X > a) = \int_{\mathbb{R}} 1(\xi > a) f(\xi) d\xi = E_P[1(X > a)]$, where E_P denotes the expectation operator under P . We also have

$$P(X > a) = \int_{\mathbb{R}} \left[1(\xi > a) \frac{f(\xi)}{q(\xi)} \right] q(\xi) d\xi = E_Q \left[1(X > a) \frac{f(X)}{q(X)} \right].$$

The density q has been selected such that 50% of its samples exceed a and the ratio f/q is bounded in \mathbb{R} . ▲

Example 2.68 Suppose our objective is to estimate the probability $P(X > a)$, where $X \sim N(0, 1)$. The coefficient of variation of the Monte Carlo estimator \hat{Y}_{MC} in (2.64) with $h(X) = 1(X > a)$ is $2.30/\sqrt{n}$, $6.55/\sqrt{n}$, and $1870/\sqrt{n}$ for $a = 1, 2$, and 5 , respectively. Consider also an improved estimator for $P(X > a)$ defined by

$$\hat{Y}_{IMC} = \frac{1}{n} \sum_{i=1}^n 1(X_i > a) \frac{\phi(X_i)}{\phi^*(X_i)},$$

where $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$, $\phi^*(u) = \exp(-(u-a)^2/(2\sigma^2))/(\sqrt{2\pi}\sigma)$, $\sigma > 0$, and $\{X_i\}$ denote independent copies of X . The coefficient of variation of \hat{Y}_{IMC} is approximately $91.97/\sqrt{n}$, $97.21/\sqrt{n}$, and $67.06/\sqrt{n}$ for $a = 1, 2$, and 5 if $\sigma = 0.1$ and $1.21/\sqrt{n}$, $1.59/\sqrt{n}$, and $2.41/\sqrt{n}$ for $a = 1, 2$, and 5 if $\sigma = 1.0$. Note that \hat{Y}_{MC} deteriorates rapidly as a increases and that the efficiency of \hat{Y}_{IMC} depends strongly on the selection of sampling distribution ϕ^* . ◇

Example 2.68 shows that the efficiency of improved Monte Carlo simulation depends essentially on the measure proposed for calculations. Since the coefficient of variation of \hat{Y}_{IMC} is not available analytically, the selection of an optimal density $\phi^*(\cdot)$ requires extensive calculations. The coefficients of variation reported for \hat{Y}_{IMC} have been estimated from Monte Carlo experiments performed for various σ .

There is no efficient procedure for selecting measures yielding accurate and efficient estimators even for one-dimensional problems, as considered in Example 2.68. The construction of improved estimators encounters additional difficulties when dealing with multidimensional problems. For example, consider the estimation of the expectation $E[1(X \in D)]$, where X is an \mathbb{R}^d -valued random variable and D is a subset of \mathbb{R}^d . The selection of a new measure for X such that its samples under this measure fall in equal proportion in D and D^c is a rather complex task. The following example presents a multidimensional problem for which an improved Monte Carlo algorithm can be constructed simply.

Example 2.69 Let $p_s = P(X \in D)$ and $p_f = P(X \in D^c)$, where $D = \{x \in \mathbb{R}^d : \|x\| \leq r\}$ is a sphere of radius $r > 0$ centered at the origin of \mathbb{R}^d and X is an \mathbb{R}^d -valued random variable with independent $N(0, 1)$ coordinates defined on a probability space (Ω, \mathcal{F}, P) . The probability p_f can be calculated from

$$\begin{aligned} p_f &= \int_{\mathbb{R}^d} 1(x \in D^c) f(x) dx = E_P[1(X \in D^c)] \quad \text{or} \\ p_f &= \int_{\mathbb{R}^d} \left[1(x \in D^c) \frac{f(x)}{q(x)} \right] q(x) dx = E_Q \left[1(X \in D^c) \frac{f(X)}{q(X)} \right], \end{aligned} \quad (2.66)$$

where Q is a measure on (Ω, \mathcal{F}) such that $P \ll Q$. The densities f and q of the distributions induced by probability measures P and Q are

$$f(x) = (2\pi)^{-d/2} \exp \left[-\frac{1}{2} \sum_{i=1}^d x_i^2 \right] \quad \text{and}$$

$$q(z) = [(2\pi)\sigma^2]^{-d/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[(z_1 - r)^2 + \sum_{i=2}^d z_i^2 \right] \right\}.$$

Monte Carlo estimators based on the first and the second expressions of p_f in (2.66) are denoted by $\hat{p}_{f,MC}$ and $\hat{p}_{f,IMC}$, respectively. The exact probability p_f can be calculated from

$$1 - P_f = P \left(\sum_{i=1}^d X_i^2 \leq r^2 \right) = \frac{1}{\Gamma(d/2)} \int_0^{r^2/2} \xi^{d/2-1} e^{-\xi} d\xi,$$

and is equal to 0.0053, 0.8414×10^{-4} , and 0.4073×10^{-6} for $r = 5, 6$, and 7 , where $\Gamma(\cdot)$ denotes the gamma function. Monte Carlo estimates $\hat{p}_{f,MC}$ of p_f based on 10000 independent samples of X are 0.0053, 0.0, 0.0 for $r = 5, 6$, and 7 . Improved Monte Carlo estimates $\hat{p}_{f,IMC}$ of p_f based on the same number of samples for $r = 5$ are 0.0, 0.0009, 0.0053, and 0.0050 if $\sigma = 0.5, 1.0, 2.0$, and 3.0 . For $r = 6$, the estimates $\hat{p}_{f,IMC}$ are 0.0001×10^{-4} , 0.1028×10^{-4} , 0.5697×10^{-4} , 1.1580×10^{-4} , and 1.1350×10^{-4} if $\sigma = 0.5, 1.0, 2.0, 3.0$, and 4.0 . For $r = 7$, the estimates $\hat{p}_{f,IMC}$ are 0.0, 0.0016×10^{-6} , 0.1223×10^{-6} , 0.6035×10^{-6} , and 0.4042×10^{-6} if $\sigma = 0.5, 1.0, 2.0, 3.0$, and 4.0 . Note that the density q corresponds to an \mathbb{R}^d -valued variable with independent Gaussian coordinates with variance σ^2 and mean 0, except for a coordinate that has mean r . Monte Carlo estimates are unsatisfactory for relatively large values of r . The accuracy of the estimators $\hat{p}_{f,IMC}$ depends essentially on q ; for example, they are unsatisfactory for $\sigma = 0.5$ and accurate for σ in the range [3, 4]. \diamond

2.14 Exercises

Exercise 2.1 Show that $\sigma(\mathcal{A})$ defined by (2.1) is a σ -field, and that $\sigma(\mathcal{A})$ is the smallest σ -field including \mathcal{A} .

Exercise 2.2 Prove the properties of the probability measure P in (2.2).

Exercise 2.3 Show that the inclusion–exclusion formula in (2.3) is valid.

Hint: Use the fourth formula in (2.2) to calculate the probability of $\cup_{i=1}^m A_i$ by viewing this event as the union of $\cup_{i=1}^{m-1} A_i$ and A_m for $m \leq n$.

Exercise 2.4 Show that the conditional probability $P(A \mid B)$ in (2.5) is a probability measure on (Ω, \mathcal{F}, P) .

Exercise 2.5 Prove the law of total probability and the Bayes formula in (2.6).

Exercise 2.6 Consider the events $A_1 = \{(6, 2)\}$, $A_2 = \{(6, 2), (4, 4), (1, 6)\}$, and $B = \{\omega = (i, j) \in \Omega : i + j = 8\}$ in the experiment of rolling two dice. Calculate the conditional probabilities $P(A_1 \mid B)$ and $P(A_2 \mid B)$ by using the definition in (2.5) and by direct arguments.

Exercise 2.7 Let X be a random element. Show that the $\sigma(X)$ in Definition 2.13 is a σ -field and that this field is the smallest with respect to which X is measurable.

Exercise 2.8 Suppose a random variable X defined on a probability space (Ω, \mathcal{F}, P) takes a finite number of values $a_1, \dots, a_n \in \mathbb{R}$. Construct the σ -field $\sigma(X)$ generated by this variable.

Exercise 2.9 Prove Fatou's lemma for sequences of events, that is, show

$$P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n),$$

where $\{A_n\}$ are events on a probability space (Ω, \mathcal{F}, P) .

Hint Note that $P(\liminf_{n \rightarrow \infty} A_n) = P(\lim_{n \rightarrow \infty} \cap_{k \geq n} A_k) = \lim_{n \rightarrow \infty} P(\cap_{k \geq n} A_k)$, where the latter equality holds by Theorem 2.6 since $\cap_{k \geq n} A_k$ is an increasing sequence of events. Since $P(\cap_{k \geq n} A_k) \leq P(A_n)$, we have $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n)$. Similar arguments can be used to show $\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n)$. The inequality $\liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n)$ is valid since $\{P(A_n)\}$ is a numerical sequence.

Exercise 2.10 Consider two \mathbb{R}^d -valued random variables X and Y defined on a probability space (Ω, \mathcal{F}, P) . Show that $P(\{X \leq x\} \cap \{Y < y\}) = P(X \leq x)P(Y < y)$ implies the independence of X and Y , where the notation $X \leq x$ means $(X_1 \leq x_1, \dots, X_d \leq x_d)$.

Exercise 2.11 Show that X defined by (2.15) is an \mathbb{R}^d -value random variable and that the collection of simple random variables constitutes a vector space.

Exercise 2.12 Prove Jensen's inequality in (2.18) for finite-valued simple random variables.

Hint Use the following fact. If $g : \mathbb{R} \rightarrow \mathbb{R}$ is convex, then g is continuous and $g(x) = \sup\{l(x) : l(u) \leq g(u), \forall u \in \mathbb{R}\}$, where l denotes a linear function.

Exercise 2.13 Prove Fatou's lemma given by (2.28).

Exercise 2.14 Show that $X \leq Y$ a.s. and $E[|X|], E[|Y|] < \infty$ imply $E[X1_A] \leq E[Y1_A]$, $A \in \mathcal{F}$.

Exercise 2.15 Show that the expectation of a positive random variable X is given by $E[X] = \int_{[0, \infty)} P(X > x) dx$.

Hint The mapping $(x, \omega) \mapsto 1(X(\omega) > x)$ is measurable from $([0, \infty) \times \Omega, \mathcal{B}([0, \infty)) \times \mathcal{F})$ to $(\{0, 1\}, \mathcal{K})$, where $\mathcal{K} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. Fubini's theorem and the equality $\int_{[0, \infty)} 1(X(\omega) > x) dx = \int_{[0, X(\omega))} dx = X(\omega)$ give

$$\begin{aligned}
\int_{[0,\infty)} P(X > x) dx &= \int_{[0,\infty)} \left[\int_{\Omega} 1(X(\omega) > x) P(d\omega) \right] dx \\
&= \int_{\Omega} \left[\int_{[0,\infty)} 1(X(\omega) > x) dx \right] P(d\omega) = \int_{\Omega} X(\omega) P(d\omega) = E[X].
\end{aligned}$$

Exercise 2.16 Show that the correlation and covariance matrices of a random vector with finite variance are positive definite.

Exercise 2.17 Prove the properties of the distribution function stated following Definition 2.32.

Hint Set $B_n = \{\omega : X(\omega) \leq x_n\}$, and $B = \{\omega : X(\omega) \leq x\}$, where $\{x_n\}$ is a decreasing numerical series converging to x . The sequence of events B_n is decreasing so that $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n = B$ implying $\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(B_n) = P(\lim_{n \rightarrow \infty} B_n) = P(B) = F(x)$.

Since F is a bounded, increasing, and right continuous function, it can only have jump discontinuities. Recall that F has a jump discontinuity at c if the left and right limits of F at c are finite but not equal. To show that F has at most a countable number of jump discontinuities, consider two distinct jump points $\xi < \xi'$ of F and the open intervals $I_{\xi} = (F(\xi-), F(\xi))$ and $I_{\xi'} = (F(\xi'-), F(\xi'))$ associated with these jumps. Since $\xi \neq \xi'$, there exists $\tilde{\xi} \in (\xi, \xi')$ such that $F(\xi) \leq F(\tilde{\xi}) \leq F(\xi'-)$, showing that I_{ξ} and $I_{\xi'}$ are disjoint intervals. The collection of intervals I_{ξ} is countable since each I_{ξ} contains a rational number and the set of rational number is countable. The sum of all jumps of F is $\sum_{\xi \in J} [F(\xi+) - F(\xi-)] = \sum_{\xi \in J} [F(\xi) - F(\xi-)] \leq 1$, where J denotes the collection of jump points of F . Hence, $\varepsilon n_{\varepsilon} \leq 1$ so that $n_{\varepsilon} \leq 1/\varepsilon$, where n_{ε} denotes the number of jumps of F larger than $\varepsilon > 0$.

Exercise 2.18 Show that the central moments $E[(X - \mu)^q]$ of $X \sim N(\mu, \sigma^2)$ are zero if q is odd and equal to $q! \sigma^q / (2^{q/2} (q/2)!)$ if q is even.

Exercise 2.19 Let $X \sim N(0, 1)$ and set $Y = X^2$. Find the covariance matrix of (X, Y) . Are X and Y correlated? Are X and Y independent?

Exercise 2.20 Find the characteristic function of $X = a + bN$, where $a, b \in \mathbb{R}$ are constants and N is a Poisson random variable with intensity $\lambda > 0$, that is, an $\{0, 1, \dots\}$ -valued variable with probability $P(N = n) = \lambda^n e^{-\lambda} / n!$, $n \geq 0$.

Exercise 2.21 Let X and Y be random variables defined on the same probability space. Show that (1) if X and Y are independent, they are uncorrelated, (2) uncorrelated random variables can be dependent, and (3) uncorrelated Gaussian variables are independent.

Exercise 2.22 Calculate the expectation of random variable $X_2 \mid (X_1 = z)$ with density in (2.38).

Exercise 2.23 Show that the characteristic function φ of a real-valued random variable X is positive definite.

Hint The function $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ is positive definite if the matrix $\{\varphi(u_k - u_l), k, l = 1, \dots, n\}$ is positive semi-definite for all $n \geq 1$ and $u_k \in \mathbb{R}$. Note that $0 \leq E[ZZ^*] = \sum_{k,l=1}^n z_k z_l^* \varphi(u_k - u_l)$ for $Z = \sum_{k=1}^n z_k \exp(iu_k X)$ and $z_k \in \mathbb{C}$ arbitrary.

Exercise 2.24 Find the expression of the characteristic function for $X \sim N(\mu, \gamma)$ given by (2.43).

Exercise 2.25 Calculate the moment generating function $m(u) = E[\exp(uX)]$, $u \in \mathbb{R}$, for $X \sim N(\mu, \sigma^2)$.

Exercise 2.26 Find the properties of the conditional Gaussian vector in (2.44).

Exercise 2.27 Let $X(\omega) = 2 + \sin(2\pi\omega)$ be a random variable defined on a probability space $(\Omega = [0, 1], \mathcal{F} = \mathcal{B}[0, 1], P(d\omega) = d\omega)$ and let $\Lambda_1 = [0, 1/4]$, $\Lambda_2 = [1/4, 3/4]$, $\Lambda_3 = [3/4, 1]$, and $\Lambda_4 = \{1\}$ be a measurable partition of Ω . Calculate the conditional expectation $E[X | \mathcal{G}]$, where $\mathcal{G} = \sigma(\Lambda_i, i = 1, \dots, 4)$. Plot $E[X | \mathcal{G}]$ and $E[X]$ against $\omega \in \Omega = [0, 1]$.

Exercise 2.28 Prove Theorem 2.15.

Exercise 2.29 Prove the relationships (2.52) and (2.53) in Theorem 2.16.

Hint The defining relation gives $\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_1] dP$ for all $\Lambda \in \mathcal{G}_1$, so that $\int_{\Lambda} E[X | \mathcal{G}_1] dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP$ under the assumption $E[X | \mathcal{G}_1] = E[X | \mathcal{G}_2]$. We have $\int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP$ for all $\Lambda \in \mathcal{G}_1$ so that $E[X | \mathcal{G}_2]$ is \mathcal{G}_1 -measurable. Conversely, if $E[X | \mathcal{G}_2]$ is \mathcal{G}_1 -measurable, then $E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} = E[X | \mathcal{G}_2]$. The proof is completed by using (2.53).

Note that $\int_{\Lambda} E\{E[X | \mathcal{G}_2] | \mathcal{G}_1\} dP = \int_{\Lambda} E[X | \mathcal{G}_2] dP = \int_{\Lambda} X dP = \int_{\Lambda} E[X | \mathcal{G}_1] dP$ holds for all $\Lambda \in \mathcal{G}_1 \subset \mathcal{G}_2$ by the defining relation, which gives the first equality in (2.53). The second equality in this formula results since $E[X | \mathcal{G}_1]$ is \mathcal{G}_1 -measurable so that \mathcal{G}_2 -measurable, which gives $E\{E[X | \mathcal{G}_1] | \mathcal{G}_2\} = E[X | \mathcal{G}_1]$.

Exercise 2.30 Prove (2.55) by noting that \mathcal{G} is generated by the partition $\{B, B^c\}$ of Ω .

Exercise 2.31 Show that if X and Z are random variables on the same probability space, then $E[X | Z] = E[X | \sigma(Z)]$, where $E[X | Z] = \int u f_{X|Z}(u | z) du$ and $f_{X|Z}$ is the density of X conditional on Z .

Exercise 2.32 Let X be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) and \mathcal{G} be a sub- σ -field of \mathcal{F} . Show that $E[X | \mathcal{G}] = E[X]$ for $\mathcal{G} = \{\emptyset, \Omega\}$ and $E[X | \mathcal{G}] = X$ for $\mathcal{G} = \mathcal{F}$.

Hint If $\mathcal{G} = \mathcal{F}$, then $E[X | \mathcal{G}]$ is \mathcal{F} -measurable and $\int_{\Lambda} (X - E[X | \mathcal{F}]) dP = 0$ for all $\Lambda \in \mathcal{F}$ by the defining relation, so that $X = E[X | \mathcal{F}]$ a.s.

Exercise 2.33 Let X be a random variable with distribution F and $a \in \mathbb{R}$ such that $F(a) \in (0, 1)$. Show that

$$E[X | \mathcal{G}] = \frac{\int_{-\infty}^a x dF(x)}{F(a)} 1_A + \frac{\int_a^{\infty} x dF(x)}{1 - F(a)} 1_{A^c},$$

where $\mathcal{G} = \{\emptyset, \Omega, A, A^c\}$ and $A = X^{-1}((-\infty, a])$.

Exercise 2.34 Suppose A_i and X_i in (2.57) have finite variance. Calculate the mean and variance of M_n in this equation.

Exercise 2.35 Calculate the mean and the coefficient of variation of the estimator \hat{Y}_{IMC} in Example 2.68.

References

1. Applebaum D (2004) Lévy processes and stochastic calculus. Cambridge University Press, Cambridge
2. Brabenec RL (1990) Introduction to real analysis. PWS-KENT Publishing Company, Boston
3. Bratley P, Fox BL, Schrage LE (1987) A guide to simulation. 2nd edn. Springer, New York
4. Chung KL (1974) A course in probability theory. Academic Press Inc., New York
5. Fristedt B, Gray L (1997) A modern approach to probability theory. Birkhäuser, Boston
6. Grigoriu M (1995) Applied non-Gaussian processes: Examples, theory, simulation, linear random vibration, and MATLAB solutions. Prentice Hall, Englewoods Cliffs
7. Grigoriu M (2002) Stochastic calculus. Applications in science and engineering. Birkhäuser, Boston
8. Hernández DB (1995) Lectures on probability and second order random fields. World Scientific, London
9. Iranpour R, Chacon P (1988) The Mark Kac lectures. Macmillan Publishing Company, New York
10. Protter P (1990) Stochastic integration and differential equations. Springer, New York
11. Resnick SI (1998) A probability path. Birkhäuser, Boston
12. Rubinstein R (1981) Simulation and the Monte Carlo method. Wiley, New York
13. Steele JM (2001) Stochastic calculus and financial applications. Springer, New York



<http://www.springer.com/978-1-4471-2326-2>

Stochastic Systems

Uncertainty Quantification and Propagation

Grigoriu, M.D.

2012, XII, 532 p., Hardcover

ISBN: 978-1-4471-2326-2