CHAPTER I

Random Maps, Distribution, and Mathematical Expectation

In the spirit of a refresher, we begin with an overview of the measure-theoretic framework for probability. Readers for whom this is entirely new material may wish to consult the appendices for statements and proofs of basic theorems from analysis. A **measure space** is a triple $(S, \mathcal{S}, \mu)$, where S is a nonempty set; $\mathcal{S}$ is a collection of subsets of S, referred to as a **σ-field**, which includes $\emptyset$ and is closed under complements and countable unions; and $\mu : \mathcal{S} \to [0, \infty]$ satisfies (i) $\mu(\emptyset) = 0$, (ii) (**countable additivity**) $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ if $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{S}$. Subsets of S belonging to $\mathcal{S}$ are called **measurable sets**.. The pair $(S, \mathcal{S})$ is referred to as a **measurable space**, and the set function μ is called a **measure**. Familiar examples from real analysis are **Lebesgue measure** μ on $S = \mathbb{R}^k$, equipped with a σ-field $\mathcal{S}$ containing the class of all k-dimensional rectangles, say $R = (a_1, b_1] \times \cdots \times (a_k, b_k]$, with "volume" measure $\mu(R) = \prod_{j=1}^{k}(b_j - a_j)$; or **Dirac point mass measure** $\mu = \delta_x$ at $x \in S$ defined by $\delta_x(B) = 1$ if $x \in B$, $\delta_x(B) = 0$ if $x \in B^c$, for $B \in \mathcal{S}$. Such examples should suffice for the present, but see Appendix A for constructions of these and related measures based on the **Carathéodory extension theorem**. If $\mu(S) < \infty$ then μ is referred to as a **finite measure**. If one may write $S = \cup_{n=1}^{\infty} S_n$, where each $S_n \in \mathcal{S} (n \geq 1)$ and $\mu(S_n) < \infty, \forall n$, then μ is said to be a $\sigma-$**finite measure**.

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a nonempty set, $\mathcal{F}$ is a σ-field of subsets of Ω, and P is a finite measure on the measurable space $(\Omega, \mathcal{F})$ with $P(\Omega) = 1$. The measure P is referred to as a **probability**. Intuitively, Ω represents the set of all possible "outcomes" of a random experiment, real or conceptual, for some given coding of the results of the experiment. The set Ω is referred to as the **sample space** and the elements $\omega \in \Omega$ as **sample points** or possible outcomes.

The $\sigma-$field $\mathcal{F}$ comprises "events" $A \subseteq \Omega$ whose probability $P(A)$ of occurrence is well defined.

The finite total probability and countable additivity of a probability have many important consequences such as **finite additivity**, **finite** and **countable subadditivity**, **inclusion–exclusion**, **monotonicity**, and the formulas for both relative complements and universal complements. Proofs of these properties are left to the reader and included among the exercises.

Example 1 *(Finite Sampling of a Balanced Coin).* Consider m repeated tosses of a balanced coin. Coding the individual outcomes as 1 or 0 (or, say, H,T), the possible outcomes may be represented as sequences of binary digits of length m. Let $\Omega = \{0,1\}^m$ denote the set of all such sequences and $\mathcal{F} = 2^\Omega$, the power set of Ω. The condition that the coin be balanced may be defined by the requirement that $P(\{\omega\})$ is the same for each sequence $\omega \in \Omega$. Since Ω has cardinality $|\Omega| = 2^m$, it follows from the finite additivity and total probability requirements that

$$P(\{\omega\}) = \frac{1}{2^m} = \frac{1}{|\Omega|}, \quad \omega \in \Omega.$$

Using finite additivity this completely and explicitly specifies the model $(\Omega, \mathcal{F}, P)$ with

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \frac{|A|}{|\Omega|}, \quad A \subseteq \Omega.$$

The so-called **continuity properties** also follow from the definition as follows: A sequence of events $A_n, n \geq 1$, is said to be **increasing** (respectively, **decreasing**) with respect to set inclusion if $A_n \subseteq A_{n+1}, \forall n \geq 1$ (respectively $A_n \supseteq A_{n+1} \forall n \geq 1$). In the former case one defines $\lim_n A_n := \cup_n A_n$, while for decreasing measurable events $\lim_n A_n := \cap_n A_n$. In either case the continuity of a probability, from below or above, respectively, is the following consequence of countable additivity[1] (Exercise 3):

$$P(\lim_n A_n) = \lim_n P(A_n). \tag{1.1}$$

A bit more generally, if $\{A_n\}_{n=1}^\infty$ is a sequence of measurable events one defines

$$\limsup_n A_n := \cap_{n=1} \cup_{m \geq n} A_m \tag{1.2}$$

[1]With the exception of properties for "complements" and "continuity from above," these and the aforementioned consequences can be checked to hold for any measure.

and

$$\liminf_n A_n := \cup_{n=1}^{\infty} \cap_{m \geq n} A_m. \tag{1.3}$$

The event $\limsup_n A_n$ denotes the collection of outcomes $\omega \in \Omega$ that correspond to the occurrences of A_n for infinitely many n; i.e., the events A_n occur **infinitely often**. This event is also commonly denoted by $[A_n \ i.o.] := \limsup_n A_n$. On the other hand, $\liminf_n A_n$ is the set of outcomes ω that belong to A_n for all but finitely many n. Note that $[A_n \ i.o.]^c$ is the event that A_n^c occurs for all but finitely many n and equals $\liminf_n A_n^c$.

***Lemma 1** (Borel–Cantelli I).* Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A_n \in \mathcal{F}, n = 1, 2, \ldots$. If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \ i.o.) = 0$.

Proof. Apply (1.1) to the decreasing sequence of events $\cup_{m=1}^{\infty} A_m \supseteq \cup_{m=2}^{\infty} A_m \supseteq \cdots$ and then subadditivity of the probability to get

$$P(\limsup_n A_n) = \lim_{n \to \infty} P(\cup_{m=n}^{\infty} A_m) \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} P(A_m) = 0. \tag{1.4}$$

■

A partial converse (Borel–Cantelli II) will be given in the next chapter.

***Example 2** (Infinite Sampling of a Balanced Coin).* The possible outcomes of nonterminated repeated coin tosses can be coded as infinite binary sequences of 1's and 0's. Thus the sample space is the infinite product space $\Omega = \{0, 1\}^{\infty}$. Observe that a sequence $\omega \in \Omega$ may be viewed as the digits in a binary expansion of a number x in the unit interval. The binary expansion $x = \sum_{n=1}^{\infty} \omega_n(x) 2^{-n}$, where $\omega_n(x) \in \{0, 1\}$, is not unique for binary rationals, e.g., $\frac{1}{2} = .1000000\ldots = .011111\ldots$, however it may be made unique by requiring that infinitely many 0's occur in the expansion. Thus Ω and $[0, 1)$ may be put in one-to-one correspondence. Observe that for a given specification $\varepsilon_n \in \{0, 1\}, n = 1, \ldots, m$, of the first m tosses, the event $A = \{\omega = (\omega_1, \omega_2, \ldots) \in \Omega : \omega_n = \varepsilon_n, n \leq m\}$ corresponds to the subinterval $[\sum_{n=1}^{m} \varepsilon_n 2^{-n}, \sum_{n=1}^{m} \varepsilon_n 2^{-n} + 2^{-m})$ of $[0, 1)$ of length (Lebesgue measure) 2^{-m}. Again modeling the repeated tosses of a balanced coin by the requirement that for each fixed m, $P(A)$ not depend on the specified values $\varepsilon_n \in \{0, 1\}, 1 \leq n \leq m$, it follows from finite additivity and total probability one that $P(A) = 2^{-m} = |A|$, where $|A|$ denotes the one-dimensional Lebesgue measure of A. Based on these considerations, one may use Lebesgue measure on $[0, 1)$ to define a probability model for infinitely many tosses of a balanced coin. As we will see below, this is an essentially unique choice. For now, let us exploit the model with an illustration of the Borel–Cantelli Lemma I. Fix a nondecreasing sequence r_n of positive integers and let $A_n = \{x \in [0, 1) : \omega_k(x) = 1, k = n, n + 1, \ldots, n + r_n - 1\}$ denote the event that a run of 1's occurs of length at least r_n starting at the nth toss. Note that this is an

interval of length 2^{-r_n}. Thus if r_n increases so quickly that $\sum_{n=1}^{\infty} 2^{-r_n} < \infty$ then the Borel–Cantelli lemma I yields that $P(A_n \; i.o.) = 0$. For a concrete illustration, let $r_n = [\theta \log_2 n]$, for fixed $\theta > 0$, with $[\cdot]$ denoting the integer part. Then $P(A_n \; i.o.) = 0$ for $\theta > 1$.

In the previous example, probability considerations led us to conclude that under the identification of sequence space with the unit interval, the probabilities of events in a certain collection $\mathcal{C}$ coincide with their Lebesgue measures. Let us pursue this situation somewhat more generally. For a given collection $\mathcal{C}$ of subsets of Ω, the smallest σ-field that contains all of the events in $\mathcal{C}$ is called the **σ-field generated by $\mathcal{C}$** and is denoted by $\sigma(\mathcal{C})$; if $\mathcal{G}$ is any $\sigma-$field containing $\mathcal{C}$ then $\sigma(\mathcal{C}) \subseteq \mathcal{G}$. Note that, in general, if $\mathcal{F}_\lambda, \lambda \in \Lambda$, is an arbitrary collection of σ-fields of subsets of Ω, then $\cap_{\lambda\in\Lambda}\mathcal{F}_\lambda := \{F \subseteq \Omega : F \in \mathcal{F}_\lambda \, \forall \lambda \in \Lambda\}$ is a σ-field. On the other hand $\cup_{\lambda\in\Lambda}\mathcal{F}_\lambda := \{F \subseteq \Omega : F \in \mathcal{F}_\lambda \text{ for some } \lambda \in \Lambda\}$ is not generally a σ-field. Define the **join σ-field**, denoted by $\bigvee_{\lambda\in\Lambda} \mathcal{F}_\lambda$, to be the σ-field generated by $\cup_{\lambda\in\Lambda}\mathcal{F}_\lambda$.

It is not uncommon that $\mathcal{F} = \sigma(\mathcal{C})$ for a collection $\mathcal{C}$ *closed under finite intersections*; such a collection $\mathcal{C}$ is called a **π-system**, e.g., $\Omega = (-\infty, \infty)$, $\mathcal{C} = \{(a, b] : -\infty \leq a \leq b < \infty\}$, or infinite sequence space $\Omega = \mathbb{R}^\infty$, and $\mathcal{C} = \{(a_1, b_1]\times\cdots\times(a_k, b_k]\times\mathbb{R}^\infty : -\infty \leq a_i \leq b_i < \infty, i = 1, \ldots, k, k \geq 1\}$.

A **λ-system** is a collection $\mathcal{L}$ of subsets of Ω such that (i) $\Omega \in \mathcal{L}$, (ii) If $A \in \mathcal{L}$ then $A^c \in \mathcal{L}$, (iii) If $A_n \in \mathcal{L}, A_n \cap A_m = \emptyset, n \neq m, n, m = 1, 2, \ldots$, then $\cup_n A_n \in \mathcal{L}$. A $\sigma-$field is clearly also a λ-system. The following π-λ theorem provides a very useful tool for checking measurability.

***Theorem 1.1** (Dynkin's π-λ Theorem).* If $\mathcal{L}$ is a λ-system containing a π-system $\mathcal{C}$, then $\sigma(\mathcal{C}) \subseteq \mathcal{L}$.

Proof. Let $\mathcal{L}(\mathcal{C}) = \cap\mathcal{F}$, where the intersection is over all λ-systems $\mathcal{F}$ containing $\mathcal{C}$. We will prove the theorem by showing (i) $\mathcal{L}(\mathcal{C})$ is a π-system, and (ii) $\mathcal{L}(\mathcal{C})$ is a λ-system. For then $\mathcal{L}(\mathcal{C})$ is a σ-field (see Exercise 2), and by its definition $\sigma(\mathcal{C}) \subseteq \mathcal{L}(\mathcal{C}) \subseteq \mathcal{L}$. Now (ii) is simple to check. For clearly $\Omega \in \mathcal{F}$ for all $\mathcal{F}$, and hence $\Omega \in \mathcal{L}(\mathcal{C})$. If $A \in \mathcal{L}(\mathcal{C})$, then $A \in \mathcal{F}$ for all $\mathcal{F}$, and since every $\mathcal{F}$ is a λ-system, $A^c \in \mathcal{F}$ for every $\mathcal{F}$. Thus $A^c \in \mathcal{L}(\mathcal{C})$. If $A_n \in \mathcal{L}(\mathcal{C}), n \geq 1$, is a disjoint sequence, then for each $\mathcal{F}$, $A_n \in \mathcal{F}$, for all n and $A \equiv \cup_n A_n \in \mathcal{F}$ for all $\mathcal{F}$. Since this is true for every λ-system $\mathcal{F}$, one has $A \in \mathcal{L}(\mathcal{C})$. It remains to prove (i). For each set A, define the class $\mathcal{L}_A := \{B : A \cap B \in \mathcal{L}(\mathcal{C})\}$. It suffices to check that $\mathcal{L}_A \supseteq \mathcal{L}(\mathcal{C})$ for all $A \in \mathcal{L}(\mathcal{C})$. First note that if $A \in \mathcal{L}(\mathcal{C})$, then $\mathcal{L}_A$ is a λ-system, by arguments along the line of (ii) above (Exercise 2). In particular, if $A \in \mathcal{C}$, then $A \cap B \in \mathcal{C}$ for all $B \in \mathcal{C}$, since $\mathcal{C}$ is closed under finite intersections. Thus $\mathcal{L}_A \supseteq \mathcal{C}$. This implies, in turn, that $\mathcal{L}(\mathcal{C}) \subseteq \mathcal{L}_A$. This says that $A \cap B \in \mathcal{L}(\mathcal{C})$ for all $A \in \mathcal{C}$ and for all $B \in \mathcal{L}(\mathcal{C})$. Thus, if we fix $B \in \mathcal{L}(\mathcal{C})$, then $\mathcal{L}_B \equiv \{A : B \cap A \in \mathcal{L}(\mathcal{C})\} \supseteq \mathcal{C}$. Therefore $\mathcal{L}_B \supseteq \mathcal{L}(\mathcal{C})$. In other words, for every $B \in \mathcal{L}(\mathcal{C})$ and $A \in \mathcal{L}(\mathcal{C})$, one has $A \cap B \in \mathcal{L}(\mathcal{C})$. ■

In view of the additivity properties of a probability, the following is an immediate and important corollary to the π-λ theorem.

***Corollary 1.2** (Uniqueness).* If P_1, P_2 are two probability measures such that $P_1(C) = P_2(C)$ for all events C belonging to a $\pi-$system $\mathcal{C}$, then $P_1 = P_2$ on all of $\mathcal{F} = \sigma(\mathcal{C})$.

Proof. Check that $\{A \in \mathcal{F} : P_1(A) = P_2(A)\} \supseteq \mathcal{C}$ is a λ-system. ■

For a related application suppose that (S, ρ) is a metric space. The **Borel σ-field** of S, denoted by $\mathcal{B}(S)$, is defined as the σ-field generated by the collection $\mathcal{C} = \mathcal{T}$ of open subsets of S, the collection $\mathcal{T}$ being referred to as the topology on S specified by the metric ρ. More generally, one may specify a **topology** for a set S by a collection $\mathcal{T}$ of subsets of S that includes both $\emptyset$ and S, and is closed under arbitrary unions and finite intersections. Then $(S, \mathcal{T})$ is called a **topological space** and members of $\mathcal{T}$ define the open subsets of S. The topology is said to be **metrizable** when it may be specified by a metric ρ as above. In any case, one defines the Borel σ-field by $\mathcal{B}(S) := \sigma(\mathcal{T})$.

Definition 1.1. A class $\mathcal{C} \subseteq \mathcal{B}(S)$ is said to be **measure-determining** if for any two finite measures μ, ν such that $\mu(C) = \nu(C)\ \forall C \in \mathcal{C}$, it follows that $\mu = \nu$ on $\mathcal{B}(S)$.

One may directly apply the π-λ theorem, noting that S is both open and closed, to see that the class $\mathcal{T}$ of all open sets is measure-determining, as is the class $\mathcal{K}$ of all closed sets.

If $(S_i, \mathcal{S}_i)$, $i = 1, 2$, is a pair of measurable spaces then a function $f : S_1 \to S_2$ is said to be a **measurable map** if $f^{-1}(B) := \{x \in S_1 : f(x) \in B\} \in \mathcal{S}_1$ for all $B \in \mathcal{S}_2$. In usual mathematical discourse the σ-fields required for this definition may not be explicitly mentioned and will need to be inferred from the context. For example if $(S, \mathcal{S})$ is a measurable space, by a **Borel-measurable function** $f : S \to \mathbb{R}$ is meant measurability when $\mathbb{R}$ is given its Borel σ-field. A **random variable,** or a **random map,** X is a measurable map on a probability space $(\Omega, \mathcal{F}, P)$ into a measurable space $(S, \mathcal{S})$. Measurability of X means that each event[2] $[X \in B] := X^{-1}(B)$ belongs to $\mathcal{F}\ \forall\ B \in \mathcal{S}$. The term "random variable" is most often used to denote a real-valued random variable, i.e., where $S = \mathbb{R}$, $\mathcal{S} = \mathcal{B}(\mathbb{R})$. When $S = \mathbb{R}^k$, $\mathcal{S} = \mathcal{B}(\mathbb{R}^k)$, $k > 1$, one uses the term **random vector**.

A common alternative to the use of a metric to define a topology, is to indirectly characterize the topology by specifying what it means for a sequence to converge in the topology. That is, if $\mathcal{T}$ is a topology on S, then a sequence $\{x_n\}_{n=1}^{\infty}$ in S **converges to** $x \in S$ **with respect to the topology** $\mathcal{T}$ if for arbitrary $U \in \mathcal{T}$ such that $x \in U$, there is an N such that $x_n \in U$ for all $n \geq N$. A topological space $(S, \mathcal{T})$, or a topology $\mathcal{T}$, is said to be **metrizable** if $\mathcal{T}$ coincides with the class of open sets defined by a metric ρ on S. Using this notion, other commonly occurring measurable

[2]Throughout, this square-bracket notation will be used to denote events defined by inverse images.

image spaces may be described as follows: (i) $S = \mathbb{R}^\infty$—the space of all sequences of reals with the (metrizable) **topology of pointwise convergence**, and $\mathcal{S} = \mathcal{B}(\mathbb{R}^\infty)$, (ii) $S = C[0,1]$—the space of all real-valued continuous functions on the interval $[0,1]$ with the (metrizable) **topology of uniform convergence**, and $\mathcal{S} = \mathcal{B}(C[0,1])$, and (iii) $S = C([0,\infty) \to \mathbb{R}^k)$—the space of all continuous functions on $[0,\infty)$ into $\mathbb{R}^k$, with the (metrizable) **topology of uniform convergence on compact subsets of** $[0,\infty)$, $\mathcal{S} = \mathcal{B}(S)$ (see Exercise 7).

The relevant quantities for a random map X on a probability space $(\Omega, \mathcal{F}, P)$ are the probabilities with which X takes sets of values. In this regard, P determines the most important aspect of X, namely, its **distribution** $Q \equiv P \circ X^{-1}$ defined on the image space $(S, \mathcal{S})$ by $Q(B) := P(X^{-1}(B)) \equiv P(X \in B)$, $B \in \mathcal{S}$. The distribution is sometimes referred to as the **induced measure** of X under P. Note that given any probability measure Q on a measurable space $(S, \mathcal{S})$ one can construct a probability space $(\Omega, \mathcal{F}, P)$ and a random map X on $(\Omega, \mathcal{F})$ with distribution Q. The simplest such construction is given by letting $\Omega = S$, $\mathcal{F} = \mathcal{S}$, $P = Q$, and X the **identity map**: $X(\omega) = \omega$, $\omega \in S$. This is often called a **canonical construction**, and $(S, \mathcal{S}, Q)$ with the identity map X is called a **canonical model**.

If $X = \sum_{j=1}^m a_j \mathbf{1}_{A_j}$, $A_j \in \mathcal{F}$, $A_i \cap A_j = \emptyset (i \neq j)$, is a **discrete random variable** or, equivalently, a **simple random variable**, then $\mathbb{E}X \equiv \int_\Omega X dP := \sum_{j=1}^m a_j P(A_j)$. If $X : \Omega \to [0,\infty)$ is a random variable, then $\mathbb{E}X$ is defined by the "simple function approximation" $\mathbb{E}X \equiv \int_\Omega X dP := \sup\{\mathbb{E}Y : 0 \leq Y \leq X, Y \text{ simple}\}$. In particular, one may apply the standard simple function approximations $X = \lim_{n\to\infty} X_n$ given by the nondecreasing sequence

$$X_n := \sum_{j=0}^{n2^n-1} \frac{j}{2^n} \mathbf{1}_{[j2^{-n} \leq X < (j+1)2^{-n}]} + n\mathbf{1}_{[X \geq n]}, \quad n = 1, 2, \ldots, \tag{1.5}$$

to write

$$\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}X_n = \lim_{n\to\infty} \left\{ \sum_{j=0}^{n2^n-1} \frac{j}{2^n} P(j2^{-n} \leq X < (j+1)2^{-n}) + nP(X \geq n) \right\}. \tag{1.6}$$

Note that if $\mathbb{E}X < \infty$, then $nP(X > n) \to 0$ as $n \to \infty$ (Exercise 16). Now, more generally, if X is a real-valued random variable, then the **expected value** (or, **mean**) of X is defined as

$$\mathbb{E}(X) \equiv \int_\Omega X dP := \mathbb{E}X^+ - \mathbb{E}X^-, \tag{1.7}$$

provided at least one of $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ is finite, where $X^+ = X\mathbf{1}_{[X \geq 0]}$ and $X^- = -X\mathbf{1}_{[X \leq 0]}$. If both $\mathbb{E}X^+ < \infty$ and $\mathbb{E}X^- < \infty$, or equivalently, $\mathbb{E}|X| = \mathbb{E}X^+ + \mathbb{E}X^- < \infty$, then X is said to be **integrable** with respect to the probability P. Note that if X is bounded a.s., then applying (1.5) to X^+ and X^-, one obtains a sequence

$X_n (n \geq 1)$ of simple functions that *converge uniformly* to X, outside a P-null set. (Exercise 1(i)).

If X is a random variable with values in $(S, \mathcal{S})$ and if h is a real-valued Borel-measurable function on S, then using simple function approximations to h, one may obtain the following basic **change of variables formula** (Exercise 11)

$$\mathbb{E}(h(X)) \equiv \int_\Omega h(X(\omega))P(d\omega) = \int_S h(x)Q(dx), \tag{1.8}$$

where Q is the distribution of X, provided one of the two indicated integrals may be shown to exist. If $X = (X_1, X_2, \ldots, X_k)$ is a random vector, one defines $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_k))$.

This definition of expectation as an *integral in the sense of Lebesgue* is precisely the same as that used in real analysis to define $\int_S f(x)\mu(dx)$ for a real-valued Borel measurable function f on an arbitrary measure space $(S, \mathcal{S}, \mu)$; see Appendix A. One may exploit standard tools of real analysis (see Appendices A and C), such as *Lebesgue's dominated convergence theorem, Lebesgue's monotone convergence theorem, Fatou's lemma, Fubini–Tonelli theorem, Radon–Nykodym theorem*, for estimates and computations involving expected values.

Definition 1.2. A sequence $\{X_n\}_{n=1}^\infty$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$ is said to **converge in probability** to a random variable X if for each $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$. The convergence is said to be **almost sure** if the event $[X_n \not\to X] \equiv \{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}$ has P-measure zero.

Note that almost-sure convergence always implies convergence in probability, since for arbitrary $\varepsilon > 0$ one has $0 = P(\cap_{n=1}^\infty \cup_{m=n}^\infty [|X_m - X| > \varepsilon]) = \lim_{n\to\infty} P(\cup_{m=n}^\infty [|X_m - X| > \varepsilon]) \geq \limsup_{n\to\infty} P(|X_n - X| > \varepsilon)$. An equivalent formulation of convergence in probability can be cast in terms of almost-sure convergence as follows.

Proposition 1.3. A sequence of random variables $\{X_n\}_{n=1}^\infty$ on $(\Omega, \mathcal{F}, P)$ converges in probability to a random variable X on $(\Omega, \mathcal{F}, P)$ if and only if every subsequence has an a.s. convergent subsequence to X.

Proof. Suppose that $X_n \to X$ in probability as $n \to \infty$. Let $\{X_{n_k}\}_{k=1}^\infty$ be a subsequence, and for each $m \geq 1$ recursively choose $n_{k(0)} = 1, n_{k(m)} = \min\{n_k > n_{k(m-1)} : P(|X_{n_k} - X| > 1/m) \leq 2^{-m}\}$. Then it follows from the Borel–Cantelli lemma (Part I) that $X_{n_{k(m)}} \to X$ a.s. as $m \to \infty$. For the converse suppose that X_n does not converge to X in probability. Then there exists $\varepsilon > 0$ and a sequence $n_1, n_2, \ldots$ such that $\lim_k P(|X_{n_k} - X| > \varepsilon) = \alpha > 0$. Since a.s. convergence implies convergence in probability (see Appendix A, Proposition 2.4), there cannot be an a.s. convergent subsequence of $\{X_{n_k}\}_{k=1}^\infty$. ∎

The notion of measure-determining classes of sets extends to classes of functions as follows. Let μ, ν be arbitrary finite measures on the Borel σ-field of a metric space S. A class Γ of real-valued bounded Borel measurable functions on S is **measure-determining** if $\int_S g\, d\mu = \int_S g\, d\nu \ \forall g \in \Gamma$ implies $\mu = \nu$.

Proposition 1.4. The class $C_b(\mathcal{S})$ of real-valued bounded continuous functions on S is measure-determining.

Proof. To prove this, it is enough to show that for each (closed) $F \in \mathcal{K}$ there exists a sequence of nonnegative functions $\{f_n\} \subseteq C_b(S)$ such that $f_n \downarrow \mathbf{1}_F$ as $n \uparrow \infty$. Since F is closed, one may view $x \in F$ in terms of the equivalent condition that $\rho(x, F) = 0$, where $\rho(x, F) := \inf\{\rho(x, y) : y \in F\}$. Let $h_n(r) = 1 - nr$ for $0 \le r \le 1/n, h_n(r) = 0$ for $r \ge 1/n$. Then take $f_n(x) = h_n(\rho(x, F))$. In particular, $\mathbf{1}_F(x) = \lim_n f_n(x), x \in S$, and Lebesgue's dominated convergence theorem applies. ■

Note that the functions f_n in the proof of Proposition 1.4 are uniformly continuous, since $|f_n(x) - f_n(y)| \le (n\rho(x, y)) \wedge (2 \sup_x |f(x)|)$. It follows that the **set** $UC_b(S)$ **of bounded uniformly continuous functions on** S is measure determining.

Consider the L^p-space $L^p(\Omega, \mathcal{F}, P)$ of (real-valued) random variables X such that $\mathbb{E}|X|^p < \infty$. When random variables that differ only on a P-null set are identified, then for $p \ge 1$, it follows from Theorem 1.5(e) below that $L^p(\Omega, \mathcal{F}, P)$ is a normed linear space with norm $\|X\|_p := (\int_\Omega |X|^p dP)^{\frac{1}{p}} \equiv (\mathbb{E}|X|^p)^{\frac{1}{p}}$. It may be shown that with this norm (and distance $\|X - Y\|_p$), it is a complete metric space, and therefore a **Banach space** (Exercise 18). In particular, $L^2(\Omega, \mathcal{F}, P)$ is a **Hilbert space** with inner product (see Appendix C)

$$\langle X, Y \rangle = \mathbb{E}XY \equiv \int_\Omega XY dP, \quad ||X||_2 = \langle X, X \rangle^{\frac{1}{2}} . \tag{1.9}$$

The $L^2(S, \mathcal{S}, \mu)$ spaces are the only Hilbert spaces that are required in this text, where $(S, \mathcal{S}, \mu)$ is a σ-finite measure space; see Appendix C for an exposition of the essential structure of such spaces. Note that by taking S to be a countable set with counting measure μ, this includes the l^2 *sequence space*. Unlike the case of a measure space $(\Omega, \mathcal{F}, \mu)$ with an infinite measure μ, for finite measures it is always true that

$$L^r(\Omega, \mathcal{F}, P) \subseteq L^s(\Omega, \mathcal{F}, P) \quad \text{if } r > s \ge 1, \tag{1.10}$$

as can be checked using $|x|^s < |x|^r$ for $|x| > 1$. The basic inequalities in the following Theorem 1.5 are consequences of *convexity* at some level. So let us be precise about this notion.

Definition 1.3. A function φ defined on an open interval J is said to be a **convex function** if $\varphi(ta + (1 - t)b) \le t\varphi(a) + (1 - t)\varphi(b)$, for all $a, b \in J$, $0 \le t \le 1$.

If the function φ is sufficiently smooth, one may use calculus to check convexity, see Exercise 14. The following lemma is required to establish a geometrically obvious "line of support property" of convex functions.

***Lemma 2** (Line of Support).* Suppose φ is convex on an interval J. (a) If J is open, then (i) the left-hand and right-hand derivatives φ^- and φ^+ exist and are finite and nondecreasing on J, and $\varphi^- \le \varphi^+$. Also (ii) for each $x_0 \in J$ there is a constant $m = m(x_0)$ such that $\varphi(x) \ge \varphi(x_0) + m(x - x_0), \forall x \in J$. (b) If J has a left (or right) endpoint and the right-hand (left-hand) derivative is finite, then the line of support property holds at this endpoint x_0.

Proof. (a) In the definition of convexity, one may take $a < b$, $0 < t < 1$. Thus convexity is equivalent to the following inequality with the identification $a = x$, $b = z$, $t = (z - y)/(z - x)$: For any $x, y, z \in J$ with $x < y < z$,

$$\frac{\varphi(y) - \varphi(x)}{y - x} \le \frac{\varphi(z) - \varphi(y)}{z - y}. \tag{1.11}$$

More generally, use the definition of convexity to analyze monotonicity and bounds on the Newton quotients (slopes of secant lines) from the right and left to see that (1.11) implies $\frac{\varphi(y)-\varphi(x)}{y-x} \le \frac{\varphi(z)-\varphi(x)}{z-x} \le \frac{\varphi(z)-\varphi(y)}{z-y}$ (use the fact that $c/d \le e/f$ for $d, f > 0$ implies $c/d \le (c + e)/(d + f) \le e/f$). The first of these inequalities shows that $\frac{\varphi(y)-\varphi(x)}{y-x}$ decreases as y decreases, so that the right-hand derivative $\varphi^+(x)$ exists and $\frac{\varphi(y)-\varphi(x)}{y-x} \ge \varphi^+(x)$. Letting $z \downarrow y$ in (1.11), one gets $\frac{\varphi(y)-\varphi(x)}{y-x} \le \varphi^+(y)$ for all $y > x$. Hence φ^+ is finite and nondecreasing on J. Now fix $x_0 \in J$. By taking $x = x_0$ and $y = x_0$ in turn in these two inequalities for φ^+, it follows that $\varphi(y) - \varphi(x_0) \ge \varphi^+(x_0)(y - x_0)$ for all $y \ge x_0$, and $\varphi(x_0) - \varphi(x) \le \varphi^+(x_0)(x_0 - x)$ for all $x \le x_0$. Thus the "line of support" property holds with $m = \varphi^+(x_0)$. (b) If J has a left (right) endpoint x_0, and $\varphi^+(x_0)$ $(\varphi^-(x_0))$ is finite, then the above argument remains valid with $m = \varphi^+(x_0)$ $(\varphi^-(x_0))$.

A similar proof applies to the left-hand derivative $\varphi^-(x)$ (Exercise 14). On letting $x \uparrow y$ and $z \downarrow y$ in (1.11), one obtains $\varphi^-(y) \le \varphi^+(y)$ for all y. In particular, the line of support property now follows for $\varphi^-(x_0) \le m \le \varphi^+(x_0)$. ■

***Theorem 1.5** (Basic Inequalities).* Let X, Y be random variables on $(\Omega, \mathcal{F}, P)$.

(a) *(Jensen's Inequality)* If φ is a convex function on the interval J and $P(X \in J) = 1$, then $\varphi(\mathbb{E}X) \le \mathbb{E}(\varphi(X))$ provided that the indicated expectations exist. Moreover, if φ is strictly convex, then equality holds if and only if X is a.s. constant.
(b) *(Lyapounov Inequality)* If $0 < r < s$ then $(\mathbb{E}|X|^r)^{\frac{1}{r}} \le (\mathbb{E}|X|^s)^{\frac{1}{s}}$.
(c) *(Hölder Inequality)* Let $p \ge 1$. If $X \in L^p, Y \in L^q$, $\frac{1}{p} + \frac{1}{q} = 1$, then $XY \in L^1$ and $\mathbb{E}|XY| \le (\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}$.
(d) *(Cauchy–Schwarz Inequality)* If $X, Y \in L^2$ then $XY \in L^1$ and one has $|\mathbb{E}(XY)| \le \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}$.

(e) *(Minkowski Triangle Inequality)* Let $p \geq 1$. If $X, Y \in L^p$ then $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$.

(f) *(Markov and Chebyshev-type Inequalities)* Let $p \geq 1$. If $X \in L^p$ then $P(|X| \geq \lambda) \leq \frac{\mathbb{E}(|X|^p \mathbf{1}_{[|X|\geq\lambda]})}{\lambda^p} \leq \frac{\mathbb{E}|X|^p}{\lambda^p}, \quad \lambda > 0,$

Proof. The proof of Jensen's inequality hinges on the line of support property of convex functions in Lemma 2 by taking $x = X(\omega), \omega \in \Omega, x_0 = \mathbb{E}X$. The Lyapounov inequality follows from Jensen's inequality by writing $|X|^s = (|X|^r)^{\frac{s}{r}}$. for $0 < r < s$. For the Hölder inequality, let $p, q > 1$ be **conjugate exponents** in the sense that $\frac{1}{p} + \frac{1}{q} = 1$. Using convexity of the function $\exp(x)$ one sees that $|ab| = \exp(\ln(|a|^p)/p + \ln(|b|^q)/q)) \leq \frac{1}{p}|a|^p + \frac{1}{q}|b|^q$. Applying this to $a = \frac{|X|}{\|X\|_p}, b = \frac{|Y|}{\|Y\|_q}$ and integrating, it follows that $\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|Y|^q)^{\frac{1}{q}}$. The Cauchy–Schwarz inequality is the Hölder inequality with $p = q = 2$. For the proof of Minkowski's inequality, first use the inequality (1.21) to see that $|X + Y|^p$ is integrable from the integrability of $|X|^p$ and $|Y|^p$. Applying Hölder's inequality to each term of the expansion $\mathbb{E}(|X|+|Y|)^p = \mathbb{E}|X|(|X| + |Y|)^{p-1} + \mathbb{E}|Y|(|X| + |Y|)^{p-1}$, and solving the resulting inequality for $\mathbb{E}(|X|+|Y|)^p$ (using conjugacy of exponents), it follows that $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$. Finally, for the Markov and Chebyshev-type inequalities simply observe that since $\mathbf{1}_{\{|X|\geq\lambda\}} \leq \frac{|X|^p \mathbf{1}_{\{|X|\geq\lambda\}})}{\lambda^p} \leq \frac{|X|^p}{\lambda^p}$ on Ω, taking expectations yields $P(|X| \geq \lambda) \leq \frac{\mathbb{E}(|X|^p \mathbf{1}_{\{|X|\geq\lambda\}})}{\lambda^p} \leq \frac{\mathbb{E}|X|^p}{\lambda^p}, \quad \lambda > 0.$ ■

The Markov inequality refers to the case $p = 1$ in (f). Observe from the proofs that (c–e) hold with the random variables X, Y replaced by measurable functions, in fact complex-valued, on an arbitrary (not necessarily finite) measure space $(S, \mathcal{S}, \mu)$; see Exercise 19.

The main text includes use of another limit result for Lebesgue integrals, **Scheffé's theorem**, which is more particularly suited to probability applications in which one may want to include the consequence of convergence almost everywhere in terms of convergences in other metrics. It is included here for ease of reference. To state it, suppose that $(S, \mathcal{S}, \mu)$ is an arbitrary measure space and $g : S \to [0, \infty)$ a Borel-measurable function, though not necessarily integrable. One may use g as a **density** to define another measure ν on $(S, \mathcal{S})$, i.e., with g as its Radon–Nykodym derivative $d\nu/d\mu = g$, also commonly denoted by $d\nu = g\, d\mu$, and meaning that $\nu(A) = \int_A g\, d\mu$, $A \in \mathcal{S}$; see Appendix C for a full treatment of the Radon–Nikodym theorem[3].

Recall that a sequence of measurable functions $\{g_n\}_{n=1}^{\infty}$ on S is said to **converge μ-a.e.** to a measurable function g on S if and only if $\mu(\{x \in S : \lim_n g_n(x) \neq g(x)\}) = 0$.

[3]A probabilistic proof can be given for the Radon–Nikodym theorem based on martingales. Such a proof is given in the text on stochastic processes.

Theorem 1.6 (Scheffé). Let $(S, \mathcal{S}, \mu)$ be a measure space and suppose that $\nu, \{\nu_n\}_{n=1}^{\infty}$ are measures on $(S, \mathcal{S})$ with respective nonnegative densities $g, \{g_n\}_{n=1}^{\infty}$ with respect to μ, such that

$$\int_S g_n \, d\mu = \int_S g \, d\mu < \infty, \quad \forall n = 1, 2, \ldots.$$

If $g_n \to g$ as $n \to \infty$, μ-a.e., then

$$\sup_{A \in \mathcal{S}} \left| \int_A g \, d\mu - \int_A g_n \, d\mu \right| \leq \int_S |g - g_n| \, d\mu \to 0, \text{ as } n \to \infty.$$

Proof. The indicated bound on the supremum follows from the triangle inequality for integrals. Since $\int_S (g - g_n) \, d\mu = 0$ for each n, $\int_S (g - g_n)^+ \, d\mu = \int_\Omega (g - g_n)^- \, d\mu$. In particular, since $|g - g_n| = (g - g_n)^+ + (g - g_n)^-$,

$$\int_S |g - g_n| \, d\mu = 2 \int_S (g - g_n)^+ \, d\mu.$$

But $0 \leq (g - g_n)^+ \leq g$. Since g is μ-integrable, one obtains $\int_S (g - g_n)^+ \, d\mu \to 0$ as $n \to \infty$ from Lebesgue's dominated convergence theorem. ■

For a measurable space $(S, \mathcal{S})$, a useful metric (see Exercise 1) defined on the space $\mathcal{P}(S)$ of probabilities on $\mathcal{S} = \mathcal{B}(S)$ is furnished by the **total variation distance** defined by

$$d_v(\mu, \nu) := \sup\{|\mu(A) - \nu(A)| : A \in \mathcal{B}(S)\}, \quad \mu, \nu \in \mathcal{P}(S). \tag{1.12}$$

Proposition 1.7. Suppose that $(S, \mathcal{S})$ is a measurable space. Then

$$d_v(\mu, \nu) = \frac{1}{2} \sup \left\{ \left| \int_S f \, d\mu - \int_S f \, d\nu \right| : f \in B(S), |f| \leq 1 \right\},$$

where $B(S)$ denotes the space of bounded Borel-measurable functions on S. Moreover, $(\mathcal{P}(S), d_v)$ is a complete metric space.

Proof. Let us first establish the formula for the total variation distance. By standard simple function approximation it suffices to consider bounded simple functions in the supremum. Fix arbitrary $\mu, \nu \in \mathcal{P}(S)$. Let $f = \sum_{i=1}^k a_i \mathbf{1}_{A_i} \in B(S)$ with $|a_i| \leq 1, i = 1, \ldots, k$ and disjoint sets $A_i \in \mathcal{S}$, $1 \leq i \leq k$. Let $I^+ := \{i \leq k : \mu(A_i) \geq \nu(A_i)\}$. Let I^- denote the complementary set of indices. Then by definition of the integral of a simple function and splitting the sum over $I^\pm$ one has upon twice using the triangle inequality that

$$\left| \int_S f d\mu - \int_S f d\nu \right| \leq \sum_{i \in I^+} |a_i| (\mu(A_i) - \nu(A_i)) + \sum_{i \in I^-} |a_i| (\nu(A_i) - \mu(A_i))$$

$$\leq \sum_{i\in I^+}(\mu(A_i)-\nu(A_i)) + \sum_{i\in I^-}(\nu(A_i)-\mu(A_i))$$
$$= \mu(\cup_{i\in I^+}A_i) - \nu(\cup_{i\in I^+}A_i) + \nu(\cup_{i\in I^-}A_i) - \mu(\cup_{i\in I^-}A_i)$$
$$\leq 2\sup\{|\mu(A)-\nu(A)| : A\in\mathcal{S}\}. \tag{1.13}$$

On the other hand, taking $f = \mathbf{1}_A - \mathbf{1}_{A^c}, A\in\mathcal{S}$, one has

$$\left|\int_S f d\mu - \int_S f d\nu\right| = |\mu(A)-\mu(A^c)-\nu(A)+\nu(A^c)|$$
$$= |\mu(A)-\nu(A)-1+\mu(A)+1-\nu(A)|$$
$$= 2|\mu(A)-\nu(A)|. \tag{1.14}$$

Thus, taking the supremum over sets $A\in\mathcal{S}$ establishes the asserted formula for the total variation distance. Next, to prove that the space $\mathcal{P}(S)$ of probabilities is complete for this metric, let $\{\mu_n\}_{n=1}^\infty$ be a Cauchy sequence in $\mathcal{P}(S)$. Since the closed interval $[0,1]$ of real numbers is complete, one may define $\mu(A) := \lim_n \mu_n(A), A\in\mathcal{S}$. Because this convergence is uniform over $\mathcal{S}$, it is simple to check that $\mu\in\mathcal{P}(S)$ and $\mu_n\to\mu$ in the metric d_v; see Exercise 1. ■

So we note that Scheffé's theorem provides conditions under which a.s. convergence implies $L^1(S,\mathcal{S},\mu)$-convergence of the densities g_n to g, and convergence in the total variation metric of the probabilities ν_n to ν.

We will conclude this chapter with some further basic convergence theorems for probability spaces. For this purpose we require a definition.

Definition 1.4. A sequence $\{X_n\}_{n=1}^\infty$ of random variables on a probability space $(\Omega,\mathcal{F},P)$ is said to be **uniformly integrable** if $\lim_{\lambda\to\infty}\sup_n \mathbb{E}\{|X_n|\mathbf{1}_{[|X_n|\geq\lambda]}\} = 0$.

Theorem 1.8 *(L^1-Convergence Criterion).* Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables on a probability space $(\Omega,\mathcal{F},P), X_n\in L^1$ $(n\geq 1)$. Then $\{X_n\}_{n=1}^\infty$ converges in L^1 to a random variable X if and only if (i) $X_n\to X$ in probability as $n\to\infty$, and (ii) $\{X_n\}_{n=1}^\infty$ is uniformly integrable.

Proof. *(Necessity)* If $X_n\to X$ in L^1 then convergence in probability (i) follows from the Markov inequality. Also

$$\int_{[|X_n|\geq\lambda]}|X_n|dP \leq \int_{[|X_n|\geq\lambda]}|X_n-X|dP + \int_{[|X_n|\geq\lambda]}|X|dP$$
$$\leq \int_\Omega |X_n-X|dP + \int_{[|X|\geq\lambda/2]}|X|dP$$

$$+\int_{[|X|<\lambda/2,|X_n-X|\geq\lambda/2]}|X|dP. \tag{1.15}$$

The first term of the last sum goes to zero as $n\to\infty$ by hypothesis. For each $\lambda>0$ the third term goes to zero by the dominated convergence theorem as $n\to\infty$. The second term goes to zero as $\lambda\to\infty$ by the dominated convergence theorem too. Thus there are numbers $n(\varepsilon)$ and $\lambda(\varepsilon)$ such that for all $\lambda\geq\lambda(\varepsilon)$,

$$\sup_{n\geq n(\varepsilon)}\int_{[|X_n|\geq\lambda]}|X_n|dP\leq\varepsilon. \tag{1.16}$$

Since a *finite* sequence of integrable random variables $\{X_n : 1\leq n\leq n(\varepsilon)\}$ is always uniformly integrable, it follows that the full sequence $\{X_n\}$ is uniformly integrable.

(*Sufficiency)* Under the hypotheses (i), (ii), given $\varepsilon>0$ one has for all n that

$$\int_\Omega|X_n|dP\leq\int_{[|X_n|\geq\lambda]}|X_n|dP+\lambda\leq\varepsilon+\lambda(\varepsilon) \tag{1.17}$$

for sufficiently large $\lambda(\varepsilon)$. In particular, $\{\int_\Omega|X_n|dP\}_{n=1}^\infty$ is a bounded sequence. Thus $\int_\Omega|X|dP<\infty$ by Fatou's lemma. Now

$$\begin{aligned}\int_{[|X_n-X|\geq\lambda]}|X_n-X|dP&=\int_{[|X_n-X|\geq\lambda,|X_n|\geq\lambda/2]}|X_n-X|dP\\&\quad+\int_{[|X_n|<\lambda/2,|X_n-X|\geq\lambda]}|X_n-X|dP\\&\leq\int_{[|X_n|\geq\lambda/2]}|X_n|dP+\int_{[|X_n-X|\geq\lambda/2]}|X|dP\\&\quad+\int_{[|X_n|<\lambda/2,|X_n-X|\geq\lambda]}(\frac{\lambda}{2}+|X|)dP.\end{aligned} \tag{1.18}$$

Now, using (ii), given $\varepsilon>0$, choose $\lambda=\lambda(\varepsilon)>0$ so large that the first term of the last sum is smaller than ε. With this value of $\lambda=\lambda(\varepsilon)$ the second and third terms go to zero as $n\to\infty$ by Lebesgue's dominated convergence theorem, using (i). Thus,

$$\limsup_{n\to\infty}\int_{[|X_n-X|\geq\lambda(\varepsilon)]}|X_n-X|dP\leq\varepsilon. \tag{1.19}$$

But again applying the dominated convergence theorem one also has

$$\limsup_{n\to\infty}\int_{[|X_n-X|<\lambda(\varepsilon)]}|X_n-X|dP=0. \tag{1.20}$$

Thus the conditions are also sufficient for L^1 convergence to X. ■

The next result follows as a corollary.

Theorem 1.9 *(L^p-Convergence Criterion).* Let $p \geq 1$. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P), X_n \in L^p (n \geq 1)$. Then $\{X_n\}_{n=1}^{\infty}$ converges in L^p to a random variable X if and only if (i) $X_n \to X$ in probability as $n \to \infty$, and (ii) $\{|X_n|^p\}_{n=1}^{\infty}$ is uniformly integrable.

Proof. Apply the preceding result to the sequence $\{|X_n - X|^p\}_{n=1}^{\infty}$. The proof of necessity is analogous to (1.15) and (1.16) using the following elementary inequalities:

$$|a+b|^p \leq (|a|+|b|)^p \leq (2\max\{|a|,|b|\})^p \leq 2^p(|a|^p + |b|^p). \tag{1.21}$$

For sufficiency, note as in (1.17) that (i), (ii) imply $X \in L^p$, and then argue as in (1.18) that the uniform integrability of $\{|X_n|^p : n \geq 1\}$ implies that of $\{|X_n - X|^p : n \geq 1\}$. ■

Chebyshev-type inequalities often provide useful ways to check uniform integrability of $\{|X_n|^p\}_{n=1}^{\infty}$ in the case that $\{\mathbb{E}|X_n|^m\}$ can be shown to be a bounded sequence for some $m > p$ (see Exercise 15).

EXERCISES

Exercise Set I

1. Let $(S, \mathcal{S})$ be a measurable space. (i) Show that if f is a real-valued bounded measurable function, $|f(x)| \leq c$ for all x, then the standard simple function approximations (1.5) to f^+ and f^- provide a sequence of simple functions f_n converging to f *uniformly* on S, and satisfying $|f_n(x)| \leq c$ for all x and for all n. (ii) Show that (1.12) defines a metric on $\mathcal{P}(S)$ i.e., is a well-defined nonnegative symmetric function on $\mathcal{P}(S) \times \mathcal{P}(S)$ satisfying the triangle inequality with $d_v(\mu, \nu) = 0$ if and only if $\mu = \nu$. Also show for a Cauchy sequence $\{\mu_n\}_{n=1}^{\infty}$ in $\mathcal{P}(S)$, that the set function defined by $\mu(A) := \lim_n \mu_n(A) \in [0,1], A \in \mathcal{S}$ is a probability measure. [*Hint*: The convergence of the real numbers $\mu_n(A) \to \mu(A)$ is uniform for $A \in \mathcal{S}$.]

2. Show that if $\mathcal{L}$ is a π-system and a λ-system, then it is a σ-field. In the proof of Dynkin's π-λ theorem, show that if $A \in \mathcal{L}(\mathcal{C})$, then $\mathcal{L}_A$ is a λ-system. [*Hint:* $A \cap B^c = (A^c \cup (A \cap B))^c$.]

3. Let $(\Omega, \mathcal{F}, P)$ be an arbitrary probability space and let $A_1, A_2, \ldots$ be measurable events. Prove each of the following.
   (i) (Finite Additivity). If $A_1, \ldots, A_m$ are disjoint then $P(\cup_{j=1}^m A_j) = \sum_{j=1}^m P(A_j)$.
   (ii) (Monotonicity). If $A_1 \subseteq A_2$ then $P(A_1) \leq P(A_2)$.
   (iii) (Inclusion–Exclusion). $P(\cup_{j=1}^m A_j) = \sum_{k=1}^m (-1)^{k+1} \sum_{1 \leq j_1 < \cdots < j_k \leq m} P(A_{j_1} \cap \cdots \cap A_{j_k})$.
   (iv) (Subadditivity). $P(\cup_j A_j) \leq \sum_j P(A_j)$.
   (v) Show that the property $\mu(A_n) \uparrow \mu(A)$ if $A_n \uparrow A$, holds for all measures μ. [*Hint*: $A = \cup_n B_n$, $B_1 = A_1, B_2 = A_1^c \cap A_2, \ldots, B_n = A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n$, so that $A_n = \cup_{j=1}^n B_j$.]

(vi) Show that the property: $\mu(A_n) \downarrow \mu(A)$ if $A_n \downarrow A$ holds for *finite* measures. Show by counterexample that it does not, in general, hold for measures μ that are not finite.

4. (*Bonferroni Inequalities*) Show that for odd $m \in \{1, 2, \ldots, n\}$, (a) $P(\cup_{j=1}^n A_j) \leq \sum_{k=1}^m \sum_{1 \leq j_1 \leq j_2 \leq \cdots \leq j_k \leq n} (-1)^{k+1} P(A_{j_1} \cap \cdots \cap A_{j_k})$, and for even $m \in \{2, \ldots, n\}$, (b) $P(\cup_{j=1}^n A_j) \geq \sum_{k=1}^m \sum_{1 \leq j_1 \leq j_2 \leq \cdots \leq j_k \leq n} (-1)^{k+1} P(A_{j_1} \cap \cdots \cap A_{j_k})$.

5. Let $(\Omega, \mathcal{F}, P)$ be an arbitrary probability space and suppose $A, B \in \mathcal{F}$ are *independent* events, i.e., $P(A \cap B) = P(A)P(B)$, and $P(A) \geq \frac{1}{2} \leq P(B)$. Show that $P(A \cup B) \geq \frac{3}{4}$.

6. Show that the Borel σ-field of $\mathbb{R}$ is generated by any one of the following classes of sets: (i) $\mathcal{C} = \{(a, b) : -\infty \leq a \leq b \leq \infty\}$; (ii) $\mathcal{C} = \{(a, b] : -\infty \leq a \leq b < \infty\}$; (iii) $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}$.

7. In each case below, show that ρ is a metric for the indicated topology.
   (i) For $S = \mathbb{R}^\infty$, $\rho(x, y) = \sum_{k=1}^\infty 2^{-k} |x_k - y_k| / (1 + |x_k - y_k|)$, for $x = (x_1, x_2, \ldots)$, $y = (y_1, y_2, \ldots) \in \mathbb{R}^\infty$ metrizes the topology of pointwise convergence: $x^{(n)} \to x$ if and only if $x_k^{(n)} \to x_k$ for each k, as $n \to \infty$.
   (ii) For $S = C[0, 1]$, $\rho(f, g) = \max\{|f(x) - g(x)| : x \in [0, 1]\}$ metrizes the topology of uniform convergence of continuous functions on $[0, 1]$.
   (iii) For $S = C([0, \infty) \to \mathbb{R}^k)$, $\rho(f, g) = \sum_{n=1}^\infty 2^{-n} \|f - g\|_n / (1 + \|f - g\|_n)$, where $\|f - g\|_n := \max\{\|f(x) - g(x)\| : x \in [0, n]\}$, $\|\cdot\|$ denoting the Euclidean norm on $\mathbb{R}^k$, metrizes the topology of uniform convergence on compacts.

8. Let X be a random map on $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{S})$. Show that $\mathcal{G} := \{[X \in A] : A \in \mathcal{S}\}$ is the smallest sub-σ-field of $\mathcal{F}$ such that $X : \Omega \to S$ is a random map on $(\Omega, \mathcal{G})$, i.e., such that $[X \in A] \in \mathcal{G}$ for all $A \in \mathcal{S}$.

9. Let $\Omega = \{(1, 1), (2, 2), (1, 2), (2, 1)\}$ equipped with the power set $\mathcal{F}$. Define a simple random variable by $X(\omega) = \omega_1 + \omega_2, \quad \omega = (\omega_1, \omega_2) \in \Omega$. Give an explicit description of $\sigma(X)$ as a subcollection of sets in $\mathcal{F}$ and give an example of a set in $\mathcal{F}$ that is not in $\sigma(X)$.

10. (i) Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{P} = \{A_1, A_2, \ldots, A_m\}$, $\emptyset \neq A_j \in \mathcal{F}$, $1 \leq j \leq m$, be a disjoint partition of Ω. Let $(S, \mathcal{S})$ be an arbitrary measurable space such that $\mathcal{S}$ contains all of the singleton sets $\{x\}$ for $x \in S$. Show that a random map $X : \Omega \to S$ is $\sigma(\mathcal{P})$-measurable if and only if X is a $\sigma(\mathcal{P})$-measurable simple function. Give a counterexample in the case that $\mathcal{S}$ does not contain singletons. (ii) Let $A_1, \ldots, A_k$ be nonempty subsets of Ω. Describe the smallest σ-field containing $\{A_1, \ldots, A_k\}$ and show that its cardinality is at most 2^k.

11. Give a proof of the change of variables formula. [*Hint*: (Method of simple function approximation) Begin with h an indicator function, then h a simple function, then $h \geq 0$, and finally write $h = h^+ - h^-$.]

12. Let X_1, X_2 be real-valued random variables on $(\Omega, \mathcal{F}, P)$. Suppose that $F_i(x) = P(X_i \leq x)$, $x \in \mathbb{R} (i = 1, 2)$ are two distribution functions on $(\mathbb{R}, \mathcal{B})$ and $F_1 = F_2$. Show that X_1 and X_2 have the same distribution.

13. Suppose that X_1 and X_2 are two bounded real-valued random variables on $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}X_1^m = \mathbb{E}X_2^m$, $m = 1, 2, \ldots$. Show that X_1 and X_2 must have the same distribution. [*Hint*: According to the Weierstrass approximation theorem, a continuous function on a closed and bounded interval may be approximated by polynomials uniformly over the interval (see Appendix B).]

14. (i) Show that for a convex function φ on an open interval J, φ^- is finite and nondecreasing, and the "line of support" property holds with $m = \varphi^-(x_0)$, as well as with any $m \in [\varphi^-(x_0), \varphi^+(x_0)]$. (ii) Show that while a convex φ is continuous on an open interval, it need not be so on an interval with left-hand and/or right-hand endpoints. (iii) Show that if φ has a continuous, nondecreasing derivative φ' on J, then φ is convex. In particular, if φ is twice differentiable and $\varphi'' \geq 0$ on J, then φ is convex. [*Hint*: Use the mean value theorem from calculus.]

15. Let $p \geq 1, X_n \in L^m(\Omega, \mathcal{F}, P)$ for some $m > p$. Suppose there is an M such that $\mathbb{E}|X_n|^m \leq M, \forall n \geq 1$. Show that $\{|X_n|^p\}_{n=1}^\infty$ is uniformly integrable. [*Hint*: Use a Chebyshev-type inequality.]

16. Let X be a nonnegative random variable. (i) Show that $nP(X > n) \to 0$ as $n \to \infty$ if $\mathbb{E}X < \infty$. [*Hint*: $nP(X > n) \leq \mathbb{E}X\mathbf{1}_{[X>n]}$.] (ii) Prove that $\sum_{n=1}^\infty P(X > n) \leq \mathbb{E}X \leq \sum_{n=0}^\infty P(X > n)$. [*Hint*: $\sum_{n=1}^\infty (n-1)P(n-1 < X \leq n) \leq \mathbb{E}X \leq \sum_{n=1}^\infty nP(n-1 < X \leq n)$.]

17. Let $\{f_n : n \geq 1\}$ be a Cauchy sequence in measure: $\mu(|f_n - f_m| > \varepsilon) \to 0$ as $n, m \to \infty$, $\forall \varepsilon > 0$. Prove that there exists a measurable function f such that $f_n \to f$ in measure. [*Hint*: Find a sequence $n_1 < n_2 < \cdots$ such that $\mu(|f_{n_k} - f_{n_{k+1}}| > 2^{-k}) < 2^{-k}, k = 1, 2, \ldots$. Let $B = [|f_{n_k} - f_{n_{k+1}}| > 2^{-k}\ i.o.]$, and show that $\mu(B) = 0$. On B^c, $\{f_{n_k}\}_{k=1}^\infty$ is a Cauchy sequence, converging to some function f. Also for every $\varepsilon > 0$, $\mu(|f_n - f| > \varepsilon) \leq \mu(|f_n - f_{n_k}| > \varepsilon/2) + \mu(|f_{n_k} - f| > \varepsilon/2)$. The first term on the right of this inequality is $o(1)$ as $k \to \infty, n \to \infty$. Also, outside $B_k := \cup_{m=k}^\infty [|f_{n_m} - f_{n_{m+1}}| > 2^{-m}]$, one has $|f_{n_k} - f| \leq \sum_{m=k}^\infty 2^{-m} = 2^{-(k-1)}$. By choosing k_0 such that $2^{-(k_0-1)} < \varepsilon/2$, one gets $\mu(|f_{n_k} - f| > \varepsilon/2) \leq \mu(B_{k_0}) \leq \varepsilon/2$ for all $k \geq k_0$.]

18. Show that for every $p \geq 1$, $L^p(S, \mathcal{S}, \mu)$ is a complete metric space.

19. (*Integration of Complex-Valued Functions*) A Borel measurable function $f = g + ih$ on a measure space $(S, \mathcal{S}, \mu)$ into $\mathbb{C}$, (i.e., g, h are real-valued Borel-measurable), is said to be *integrable* if its real and imaginary parts g and h are both integrable. Since $2^{-\frac{1}{2}}(|g| + |h|) \leq |f| \equiv \sqrt{g^2 + h^2} \leq |g| + |h|$, f is integrable if and only if $|f|$ is integrable. The following extend a number of standard results for measurable real-valued functions to measurable complex-valued functions.
    (a) Extend Lebesgue's dominated convergence theorem (Appendix A) to complex-valued functions.
    (b) Extend the inequalities of Lyapounov, Hölder, Minkowski, and Markov–Chebyshev (Theorem 1.5(b),(c),(e),(f)) to complex-valued functions.
    (c) For $p \geq 1$, let the L^p-space of complex-valued functions be defined by equivalence classes of complex-valued functions f induced by equality a.e. such that $|f|^p$ is integrable. Show that this L^p-space is a Banach space over the field of complex numbers with norm $\|f\|_p = (\int_S |f|^p d\mu)^{\frac{1}{p}}$.
    (d) Show that the L^2-space of complex-valued square-integrable functions is a Hilbert space with inner product $\langle f_1, f_2 \rangle = \int_S f_1 \overline{f}_2 \, d\mu$, where $\overline{f}_2$ is the complex conjugate of f_2.
    (e) Show that for the special case of real-valued functions, the L^p-norm defined above reduces to that introduced in the text.

20. Suppose that $X_1, X_2, \ldots$ is a sequence of identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Show that if $\mathbb{E}e^{|X_1|} < \infty$, then a.s. $\limsup_{n\to\infty} \frac{|X_n|}{\ln n} \leq 1$.