1 A selective review of basic probability

This chapter reviews many of the main concepts in a first level course on probability theory, with more emphasis on axioms and the definition of expectation than is typical of a first course.

1.1 The axioms of probability theory

Random processes are widely used to model systems in engineering and scientific applications. This book adopts the most widely used framework of probability and random processes, namely the one based on Kolmogorov's axioms of probability. The idea is to assume a mathematically solid definition of the model. This structure encourages a modeler to have a consistent, if not accurate, model.

A *probability space* is a triplet $(\Omega, \mathcal{F}, \mathcal{P})$. The first component, Ω, is a nonempty set. Each element ω of Ω is called an *outcome* and Ω is called the *sample space*. The second component, $\mathcal{F}$, is a set of subsets of Ω called *events*. The set of events $\mathcal{F}$ is assumed to be a σ-algebra, meaning it satisfies the following axioms (see Appendix 11.1 for set notation):

A.1 $\Omega \in \mathcal{F}$,
A.2 If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,
A.3 If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$. Also, if $A_1, A_2, \ldots$ is a sequence of elements in $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

If $A, B \in \mathcal{F}$, then $AB \in \mathcal{F}$ by A.2, A.3 and the fact $AB = (A^c \cup B^c)^c$. By the same reasoning, if $A_1, A_2, \ldots$ is a sequence of elements in a σ-algebra $\mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Events A_i, $i \in I$, indexed by a set I are called *mutually exclusive* if the intersection $A_iA_j = \emptyset$ for all $i,j \in I$ with $i \neq j$. The final component, P, of the triplet $(\Omega, \mathcal{F}, P)$ is a probability measure on $\mathcal{F}$ satisfying the following axioms:

P.1 $P(A) \geq 0$ for all $A \in \mathcal{F}$,
P.2 If $A, B \in \mathcal{F}$ and A and B are mutually exclusive, $P(A \cup B) = P(A) + P(B)$. Also, if $A_1, A_2, \ldots$ is a sequence of mutually exclusive events in $\mathcal{F}$ then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.
P.3 $P(\Omega) = 1$.

The axioms imply a host of properties including the following. For any subsets A, B, C of $\mathcal{F}$:

- If $A \subset B$ then $P(A) \leq P(B)$,
- $P(A \cup B) = P(A) + P(B) - P(AB)$,
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$,
- $P(A) + P(A^c) = 1$,
- $P(\emptyset) = 0$.

Example 1.1 (Toss of a fair coin) Using "H" for "heads" and "T" for "tails," the toss of a fair coin is modeled by

$$\Omega = \{H, T\} \quad \mathcal{F} = \{\{H\}, \{T\}, \{H, T\}, \emptyset\}$$
$$P\{H\} = P\{T\} = \frac{1}{2} \quad P\{H, T\} = 1 \quad P(\emptyset) = 0.$$

Note that, for brevity, we omitted the parentheses and wrote $P\{H\}$ instead of $P(\{H\})$.

Example 1.2 (Standard unit-interval probability space) Take $\Omega = \{\omega : 0 \leq \omega \leq 1\}$. Imagine an experiment in which the outcome ω is drawn from Ω with no preference towards any subset. In particular, we want the set of events $\mathcal{F}$ to include intervals, and the probability of an interval $[a, b]$ with $0 \leq a \leq b \leq 1$ to be given by:

$$P(\ [a, b]\ ) = b - a. \tag{1.1}$$

Taking $a = b$, we see that $\mathcal{F}$ contains singleton sets $\{a\}$, and these sets have probability zero. Since $\mathcal{F}$ is to be a σ-algebra, it must also contain all the open intervals (a, b) in Ω, and for such an open interval, $P(\ (a, b)\ ) = b - a$. Any open subset of Ω is the union of a finite or countably infinite set of open intervals, so that $\mathcal{F}$ should contain all open and all closed subsets of Ω. Thus, $\mathcal{F}$ must contain any set that is the intersection of countably many open sets, the union of countably many such sets, and so on. The specification of the probability function P must be extended from intervals to all of $\mathcal{F}$. It is not a priori clear how large $\mathcal{F}$ can be. It is tempting to take $\mathcal{F}$ to be the set of all subsets of Ω. However, that idea doesn't work – see Problem 1.37 showing that the length of all subsets of $\mathbb{R}$ can't be defined in a consistent way. The problem is resolved by taking $\mathcal{F}$ to be the smallest σ-algebra containing all the subintervals of Ω, or equivalently, containing all the open subsets of Ω. This σ-algebra is called the Borel σ-algebra for [0, 1], and the sets in it are called Borel sets. While not every subset of Ω is a Borel subset, any set we are likely to encounter in applications is a Borel set. The existence of the Borel σ-algebra is discussed in Problem 1.38. Furthermore, extension theorems of measure theory[1] imply that P can be extended from its definition (1.1) for interval sets to all Borel sets.

[1] See, for example, (Royden 1968) or (Varadhan 2001). The σ-algebra $\mathcal{F}$ can be extended somewhat further by requiring the following completeness property: if $B \subset A \in \mathcal{F}$ with $P(A) = 0$, then $B \in \mathcal{F}$ (and also $P(B) = 0$).

The smallest σ-algebra, $\mathcal{B}$, containing the open subsets of $\mathbb{R}$ is called the Borel σ-algebra for $\mathbb{R}$, and the sets in it are called *Borel subsets of* $\mathbb{R}$. Similarly, the Borel σ-algebra $\mathcal{B}^n$ of subsets of $\mathbb{R}^n$ is the smallest σ-algebra containing all sets of the form $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$. Sets in $\mathcal{B}^n$ are called *Borel subsets of* $\mathbb{R}^n$. The class of Borel sets includes not only rectangle sets and countable unions of rectangle sets, but all open sets and all closed sets. Virtually any subset of $\mathbb{R}^n$ arising in applications is a Borel set.

Example 1.3 (Repeated binary trials) Suppose we would like to represent an infinite sequence of binary observations, where each observation is a zero or one with equal probability. For example, the experiment could consist of repeatedly flipping a fair coin, and recording a one each time it shows heads and a zero each time it shows tails. Then an outcome ω would be an infinite sequence, $\omega = (\omega_1, \omega_2, \cdots)$, such that for each $i \geq 1$, $\omega_i \in \{0, 1\}$. Let Ω be the set of all such ωs. The set of events can be taken to be large enough so that any set that can be defined in terms of only finitely many of the observations is an event. In particular, for any binary sequence $(b_1, \cdots, b_n)$ of some finite length n, the set $\{\omega \in \Omega : \omega_i = b_i \text{ for } 1 \leq i \leq n\}$ should be in $\mathcal{F}$, and the probability of such a set is taken to be 2^{-n}.

There are also events that don't depend on a fixed, finite number of observations. For example, let F be the event that an even number of observations is needed until a one is observed. Show that F is an event and then find its probability.

Solution

For $k \geq 1$, let E_k be the event that the first one occurs on the kth observation. So $E_k = \{\omega : \omega_1 = \omega_2 = \cdots = \omega_{k-1} = 0 \text{ and } \omega_k = 1\}$. Then E_k depends on only a finite number of observations, so it is an event, and $P\{E_k\} = 2^{-k}$. Observe that $F = E_2 \cup E_4 \cup E_6 \cup \ldots$, so F is an event by Axiom A.3. Also, the events $E_2, E_4, \ldots$ are mutually exclusive, so by the full version of Axiom P.2:

$$P(F) = P(E_2) + P(E_4) + \cdots = \frac{1}{4}\left(1 + \left(\frac{1}{4}\right) + \left(\frac{1}{4}\right)^2 + \cdots\right) = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}.$$

The following lemma gives a continuity property of probability measures which is analogous to continuity of functions on $\mathbb{R}^n$, reviewed in Appendix 11.3. If $B_1, B_2, \ldots$ is a sequence of events such that $B_1 \subset B_2 \subset B_3 \subset \ldots$, then we can think that B_j converges to the set $\cup_{i=1}^{\infty} B_i$ as $j \to \infty$. The lemma states that in this case, $P(B_j)$ converges to the probability of the limit set as $j \to \infty$.

Lemma 1.1 *(Continuity of probability) Suppose $B_1, B_2, \ldots$ is a sequence of events.*

(a) *If $B_1 \subset B_2 \subset \cdots$ then $\lim_{j\to\infty} P(B_j) = P\left(\bigcup_{i=1}^{\infty} B_i\right)$.*
(b) *If $B_1 \supset B_2 \supset \cdots$ then $\lim_{j\to\infty} P(B_j) = P\left(\bigcap_{i=1}^{\infty} B_i\right)$.*

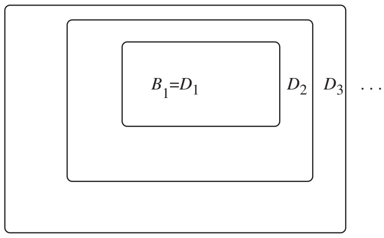

Figure 1.1 A sequence of nested sets.

Proof Suppose $B_1 \subset B_2 \subset \ldots$. Let $D_1 = B_1$, $D_2 = B_2 - B_1$, and, in general, let $D_i = B_i - B_{i-1}$ for $i \geq 2$, as shown in Figure 1.1. Then $P(B_j) = \sum_{i=1}^{j} P(D_i)$ for each $j \geq 1$, so

$$\lim_{j\to\infty} P(B_j) = \lim_{j\to\infty} \sum_{i=1}^{j} P(D_i) \stackrel{(a)}{=} \sum_{i=1}^{\infty} P(D_i) \stackrel{(b)}{=} P\left(\bigcup_{i=1}^{\infty} D_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right)$$

where (a) is true by the definition of the sum of an infinite series, and (b) is true by axiom $P.2$. This proves Lemma 1.1(a). Lemma 1.1(b) can be proved similarly, or can be derived by applying Lemma 1.1(a) to the sets B_j^c. □

Example 1.4 (Selection of a point in a square) Take Ω to be the square region in the plane,

$$\Omega = \{(x, y) : x, y \in [0, 1]\}.$$

Let $\mathcal{F}$ be the Borel σ-algebra for Ω, which is the smallest σ-algebra containing all the rectangular subsets of Ω that are aligned with the axes. Take P so that for any rectangle R,

$$P(R) = \text{area of } R.$$

(It can be shown that $\mathcal{F}$ and P exist.) Let T be the triangular region $T = \{(x, y) : 0 \leq y \leq x \leq 1\}$. Since T is not rectangular, it is not immediately clear that $T \in \mathcal{F}$, nor is it clear what $P(T)$ is. That is where the axioms come in. For $n \geq 1$, let T_n denote the region shown in Figure 1.2. Since T_n can be written as a union of finitely many mutually exclusive rectangles, it follows that $T_n \in \mathcal{F}$ and it is easily seen that $P(T_n) = \frac{1+2+\cdots+n}{n^2} = \frac{n+1}{2n}$. Since $T_1 \supset T_2 \supset T_4 \supset T_8 \cdots$ and $\cap_j T_{2^j} = T$, it follows that $T \in \mathcal{F}$ and $P(T) = \lim_{n\to\infty} P(T_n) = \frac{1}{2}$.

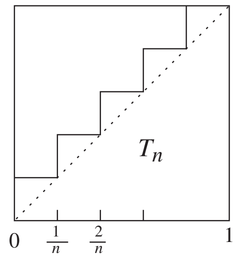

Figure 1.2 Approximation of a triangular region.

The reader is encouraged to show that if C is the diameter one disk inscribed within Ω then $P(C) = $ (area of C) $= \frac{\pi}{4}$.

1.2 Independence and conditional probability

Events A_1 and A_2 are defined to be *independent* if $P(A_1A_2) = P(A_1)P(A_2)$. More generally, events $A_1, A_2, \ldots, A_k$ are defined to be independent if

$$P(A_{i_1}A_{i_2}\ldots A_{i_j}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_j})$$

whenever j and $i_1, i_2, \ldots, i_j$ are integers with $j \geq 1$ and $1 \leq i_1 < i_2 < \cdots < i_j \leq k$. For example, events A_1, A_2, A_3 are independent if the following four conditions hold:

$$\begin{aligned}
P(A_1A_2) &= P(A_1)P(A_2), \\
P(A_1A_3) &= P(A_1)P(A_3), \\
P(A_2A_3) &= P(A_2)P(A_3), \\
P(A_1A_2A_3) &= P(A_1)P(A_2)P(A_3).
\end{aligned}$$

A weaker condition is sometimes useful: Events $A_1, \ldots, A_k$ are defined to be *pairwise independent* if A_i is independent of A_j whenever $1 \leq i < j \leq k$. Independence of k events requires that $2^k - k - 1$ equations hold: one for each subset of $\{1, 2, \ldots, k\}$ of size at least two. Pairwise independence only requires that $\binom{k}{2} = \frac{k(k-1)}{2}$ equations hold.

If A and B are events and $P(B) \neq 0$, then the *conditional probability* of A given B is defined by

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

It is not defined if $P(B) = 0$, which has the following meaning. If you were to write a computer routine to compute $P(A \mid B)$ and the inputs are $P(AB) = 0$ and $P(B) = 0$, your routine shouldn't simply return the value 0. Rather, your routine should generate an error message such as "input error – conditioning on event of probability zero." Such an error message would help you or others find errors in larger computer programs which use the routine.

As a function of A for B fixed with $P(B) \neq 0$, the conditional probability of A given B is itself a probability measure for Ω and $\mathcal{F}$. More explicitly, fix B with $P(B) \neq 0$. For each event A define $P'(A) = P(A \mid B)$. Then $(\Omega, \mathcal{F}, P')$ is a probability space, because P' satisfies the axioms $P1 - P3$. (Try showing that.)

If A and B are independent then A^c and B are independent. Indeed, if A and B are independent then

$$P(A^cB) = P(B) - P(AB) = (1 - P(A))P(B) = P(A^c)P(B).$$

Similarly, if A, B, and C are independent events then AB is independent of C. More generally, suppose $E_1, E_2, \ldots, E_n$ are independent events, suppose $n = n_1 + \cdots + n_k$ with $n_i \geq 1$ for each i, and suppose F_1 is defined by Boolean operations (intersections, complements, and unions) of the first n_1 events $E_1, \ldots, E_{n_1}$, F_2 is defined by Boolean operations on the next n_2 events, $E_{n_1+1}, \ldots, E_{n_1+n_2}$, and so on. Then $F_1, \ldots, F_k$ are independent.

Events $E_1, \ldots, E_k$ are said to form a *partition* of Ω if the events are mutually exclusive and $\Omega = E_1 \cup \ldots \cup E_k$. Of course for a partition, $P(E_1) + \ldots + P(E_k) = 1$. More generally, for any event A, the *law of total probability* holds because A is the union of the mutually exclusive sets $AE_1, AE_2, \ldots, AE_k$:

$$P(A) = P(AE_1) + \ldots + P(AE_k).$$

If $P(E_i) \neq 0$ for each i, this can be written as

$$P(A) = P(A \mid E_1)P(E_1) + \ldots + P(A \mid E_k)P(E_k).$$

Figure 1.3 illustrates the condition of the law of total probability.

Judicious use of the definition of conditional probability and the law of total probability leads to *Bayes' formula* for $P(E_i \mid A)$ (if $P(A) \neq 0$) in simple form

$$P(E_i \mid A) = \frac{P(AE_i)}{P(A)} = \frac{P(A \mid E_i)P(E_i)}{P(A)},$$

or in expanded form:

$$P(E_i \mid A) = \frac{P(A \mid E_i)P(E_i)}{P(A \mid E_1)P(E_1) + \ldots + P(A \mid E_k)P(E_k)}.$$

The remainder of this section gives the Borel–Cantelli lemma. It is a simple result based on continuity of probability and independence of events, but it is not typically

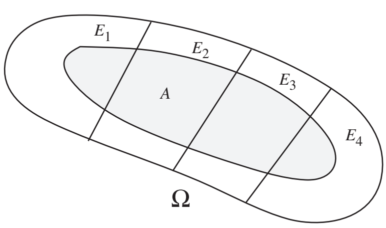

Figure 1.3 Partitioning a set A using a partition of Ω.

encountered in a first course on probability. Let $(A_n : n \geq 0)$ be a sequence of events for a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1.2 The event $\{A_n$ infinitely often$\}$ is the set of $\omega \in \Omega$ such that $\omega \in A_n$ for infinitely many values of n.

Another way to describe $\{A_n$ infinitely often$\}$ is that it is the set of ω such that for any k, there is an $n \geq k$ such that $\omega \in A_n$. Therefore,

$$\{A_n \text{ infinitely often}\} = \cap_{k\geq 1} \left(\cup_{n\geq k} A_n \right).$$

For each k, the set $\cup_{n\geq k} A_n$ is a countable union of events, so it is an event, and $\{A_n$ infinitely often$\}$ is an intersection of countably many such events, so that $\{A_n$ infinitely often$\}$ is also an event.

Lemma 1.3 *(Borel–Cantelli lemma) Let $(A_n : n \geq 1)$ be a sequence of events and let $p_n = P(A_n)$.*

(a) If $\sum_{n=1}^{\infty} p_n < \infty$, then $P\{A_n$ infinitely often$\} = 0$.
(b) If $\sum_{n=1}^{\infty} p_n = \infty$ and $A_1, A_2, \ldots$ are mutually independent, then $P\{A_n$ infinitely often$\} = 1$.

Proof (a) Since $\{A_n$ infinitely often$\}$ is the intersection of the monotonically non-increasing sequence of events $\cup_{n\geq k} A_n$, it follows from the continuity of probability for monotone sequences of events (Lemma 1.1) that $P\{A_n$ infinitely often$\} = \lim_{k\to\infty} P(\cup_{n\geq k} A_n)$. Lemma 1.1, the fact that the probability of a union of events is less than or equal to the sum of the probabilities of the events, and the definition of the sum of a sequence of numbers, yield that for any $k \geq 1$,

$$P(\cup_{n\geq k} A_n) = \lim_{m\to\infty} P(\cup_{n=k}^{m} A_n) \leq \lim_{m\to\infty} \sum_{n=k}^{m} p_n = \sum_{n=k}^{\infty} p_n.$$

Therefore, $P\{A_n$ infinitely often$\} \leq \lim_{k\to\infty} \sum_{n=k}^{\infty} p_n$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $\lim_{k\to\infty} \sum_{n=k}^{\infty} p_n = 0$, which implies part (a) of the lemma.
(b) Suppose that $\sum_{n=1}^{\infty} p_n = +\infty$ and that the events $A_1, A_2, \ldots$ are mutually independent. For any $k \geq 1$, using the fact $1 - u \leq \exp(-u)$ for all u,

$$\begin{aligned} &P(\cup_{n\geq k} A_n) \\ &= \lim_{m\to\infty} P(\cup_{n=k}^{m} A_n) = \lim_{m\to\infty} 1 - \prod_{n=k}^{m} (1 - p_n) \\ &\geq \lim_{m\to\infty} 1 - \exp\left(- \sum_{n=k}^{m} p_n \right) = 1 - \exp\left(- \sum_{n=k}^{\infty} p_n \right) = 1 - \exp(-\infty) = 1. \end{aligned}$$

Therefore, $P\{A_n$ infinitely often$\} = \lim_{k\to\infty} P(\cup_{n\geq k} A_n) = 1$. □

Example 1.5 Consider independent coin tosses using biased coins, such that $P(A_n) = p_n = \frac{1}{n}$, where A_n is the event of getting heads on the nth toss. Since $\sum_{n=1}^{\infty} \frac{1}{n} = +\infty$, the part of the Borel–Cantelli lemma for independent events implies that $P\{A_n$ infinitely often$\} = 1$.

Example 1.6 Let $(\Omega, \mathcal{F}, P)$ be the standard unit-interval probability space defined in Example 1.2, and let $A_n = [0, \frac{1}{n})$. Then $p_n = \frac{1}{n}$ and $A_{n+1} \subset A_n$ for $n \geq 1$. The events are not independent, because for $m < n$, $P(A_m A_n) = P(A_n) = \frac{1}{n} \neq P(A_m)P(A_n)$. Of course $0 \in A_n$ for all n. But for any $\omega \in (0, 1]$, $\omega \notin A_n$ for $n > \frac{1}{\omega}$. Therefore, $\{A_n$ infinitely often$\} = \{0\}$. The single point set $\{0\}$ has probability zero, so $P\{A_n$ infinitely often$\} = 0$. This conclusion holds even though $\sum_{n=1}^{\infty} p_n = +\infty$, illustrating the need for the independence assumption in Lemma 1.3(b).

1.3 Random variables and their distribution

Let a probability space $(\Omega, \mathcal{F}, P)$ be given. By definition, a random variable is a function X from Ω to the real line $\mathbb{R}$ that is $\mathcal{F}$ measurable, meaning that for any number c,

$$\{\omega : X(\omega) \leq c\} \in \mathcal{F}. \tag{1.2}$$

If Ω is finite or countably infinite, then $\mathcal{F}$ can be the set of all subsets of Ω, in which case any real-valued function on Ω is a random variable.

If $(\Omega, \mathcal{F}, P)$ is the standard unit-interval probability space described in Example 1.2, then the random variables on $(\Omega, \mathcal{F}, P)$ are called the Borel measurable functions on Ω. Since the Borel σ-algebra contains all subsets of $[0, 1]$ that come up in applications, for practical purposes we can think of any function on $[0, 1]$ as being a random variable. For example, any piecewise continuous or piecewise monotone function on $[0, 1]$ is a random variable for the standard unit-interval probability space.

The cumulative distribution function (CDF) of a random variable X is denoted by F_X. It is the function, with domain the real line $\mathbb{R}$, defined by

$$\begin{aligned} F_X(c) &= P\{\omega : X(\omega) \leq c\} \\ &= P\{X \leq c\} \text{ (for short).} \end{aligned}$$

If X denotes the outcome of the roll of a fair die ("die" is singular of "dice") and if Y is uniformly distributed on the interval $[0, 1]$, then F_X and F_Y are shown in Figure 1.4.

The CDF of a random variable X determines $P\{X \leq c\}$ for any real number c. But what about $P\{X < c\}$ and $P\{X = c\}$? Let $c_1, c_2, \ldots$ be a monotone nondecreasing sequence that converges to c from the left. This means $c_i \leq c_j < c$ for $i < j$ and $\lim_{j\to\infty} c_j = c$. Then the events $\{X \leq c_j\}$ are nested: $\{X \leq c_i\} \subset \{X \leq c_j\}$ for $i < j$, and the union of all such events is the event $\{X < c\}$. Thus, by Lemma 1.1

$$P\{X < c\} = \lim_{i\to\infty} P\{X \leq c_i\} = \lim_{i\to\infty} F_X(c_i) = F_X(c-).$$

Therefore, $P\{X = c\} = F_X(c) - F_X(c-) = \triangle F_X(c)$, where $\triangle F_X(c)$ is defined to be the size of the jump of F at c. For example, if X has the CDF shown in Figure 1.5 then $P\{X = 0\} = \frac{1}{2}$. The collection of all events A such that $P\{X \in A\}$ is determined by F_X

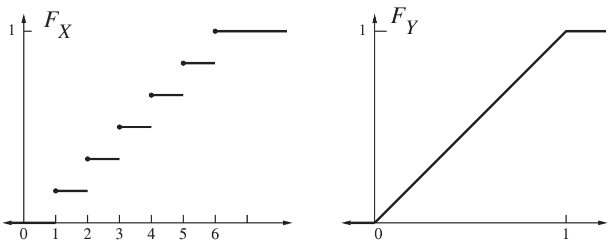

Figure 1.4 Examples of CDFs.

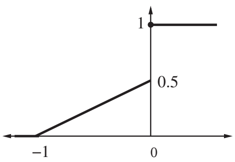

Figure 1.5 An example of a CDF.

is a σ-algebra containing the intervals, and thus this collection contains all Borel sets. That is, $P\{X \in A\}$ is determined by F_X for any Borel set A.

Proposition 1.4 *A function F is the CDF of some random variable if and only if it has the following three properties:*

- ***F.1*** *F is nondecreasing,*
- ***F.2*** $\lim_{x\to+\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$,
- ***F.3*** *F is right continuous.*

Proof The "only if" part is proved first. Suppose that F is the CDF of some random variable X. If $x < y$, $F(y) = P\{X \le y\} = P\{X \le x\} + P\{x < X \le y\} \ge P\{X \le x\} = F(x)$ so that F.1 is true. Consider the events $B_n = \{X \le n\}$. Then $B_n \subset B_m$ for $n \le m$. Thus, by Lemma 1.1,

$$\lim_{n\to\infty} F(n) = \lim_{n\to\infty} P(B_n) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = P(\Omega) = 1.$$

This and the fact F is nondecreasing imply the following. Given any $\epsilon > 0$, there exists N_ϵ so large that $F(x) \ge 1 - \epsilon$ for all $x \ge N_\epsilon$. That is, $F(x) \to 1$ as $x \to +\infty$. Similarly,

$$\lim_{n\to-\infty} F(n) \quad = \quad \lim_{n\to\infty} P(B_{-n}) = P\left(\bigcap_{n=1}^{\infty} B_{-n}\right) = P(\emptyset) = 0.$$

so that $F(x) \to 0$ as $x \to -\infty$. Property F.2 is proved.

The proof of F.3 is similar. Fix an arbitrary real number x. Define the sequence of events A_n for $n \geq 1$ by $A_n = \{X \leq x + \frac{1}{n}\}$. Then $A_n \subset A_m$ for $n \geq m$ so

$$\lim_{n\to\infty} F\left(x + \frac{1}{n}\right) = \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{k=1}^{\infty} A_k\right) = P\{X \leq x\} = F_X(x).$$

Convergence along the sequence $x + \frac{1}{n}$, together with the fact that F is nondecreasing, implies that $F(x+) = F(x)$. Property F.3 is thus proved. The proof of the "only if" portion of Proposition 1.4 is complete.

To prove the "if" part of Proposition 1.4, let F be a function satisfying properties F.1–F.3. It must be shown that there exists a random variable with CDF F. Let $\Omega = \mathbb{R}$ and let $\mathcal{F}$ be the set $\mathcal{B}$ of Borel subsets of $\mathbb{R}$. Define $\tilde{P}$ on intervals of the form $(a, b]$ by $\tilde{P}((a, b]) = F(b) - F(a)$. It can be shown by an extension theorem of measure theory that $\tilde{P}$ can be extended to all of $\mathcal{F}$ so that the axioms of probability are satisfied. Finally, let $\tilde{X}(\omega) = \omega$ for all $\omega \in \Omega$. Then

$$\tilde{P}(\tilde{X} \in (a, b]) = \tilde{P}((a, b]) \;=\; F(b) - F(a).$$

Therefore, $\tilde{X}$ has CDF F. So F is a CDF, as was to be proved. □

The vast majority of random variables described in applications are one of two types, to be described next. A random variable X is a discrete random variable if there is a finite or countably infinite set of values $\{x_i : i \in I\}$ such that $P\{X \in \{x_i : i \in I\}\} = 1$. The probability mass function (pmf) of a discrete random variable X, denoted $p_X(x)$, is defined by $p_X(x) = P\{X = x\}$. Typically the pmf of a discrete random variable is much more useful than the CDF. However, the pmf and CDF of a discrete random variable are related by $p_X(x) = \triangle F_X(x)$ and conversely,

$$F_X(x) = \sum_{y:y\leq x} p_X(y), \tag{1.3}$$

where the sum in (1.3) is taken only over y such that $p_X(y) \neq 0$. If X is a discrete random variable with only finitely many mass points in any finite interval, then F_X is a piecewise constant function.

A random variable X is a *continuous* random variable if the CDF is the integral of a function:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy.$$

The function f_X is called the *probability density function* (pdf). If the pdf f_X is continuous at a point x, then the value $f_X(x)$ has the following nice interpretation:

$$\begin{aligned} f_X(x) &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{x}^{x+\varepsilon} f_X(y)dy \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} P\{x \leq X \leq x + \varepsilon\}. \end{aligned}$$