1

Measure Theory

In this chapter, we recall some definitions and results from measure theory. Our purpose here is to provide an introduction for readers who have not seen these concepts before and to review that material for those who have. Harder proofs, especially those that do not contribute much to one's intuition, are hidden away in the Appendix. Readers with a solid background in measure theory can skip Sections 1.4, 1.5, and 1.7, which were previously part of the Appendix.

1.1 Probability Spaces

Here and throughout the book, terms being defined are set in **boldface**. We begin with the most basic quantity. A **probability space** is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a set of "outcomes," $\mathcal{F}$ is a set of "events," and $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events. We assume that $\mathcal{F}$ is a σ**-field** (or σ**-algebra**), that is, a (nonempty) collection of subsets of Ω that satisfy

(i) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and
(ii) if $A_i \in \mathcal{F}$ is a countable sequence of sets then $\cup_i A_i \in \mathcal{F}$.

Here and in what follows, **countable** means finite or countably infinite. Since $\cap_i A_i = (\cup_i A_i^c)^c$, it follows that a σ-field is closed under countable intersections. We omit the last property from the definition to make it easier to check.

Without P, $(\Omega, \mathcal{F})$ is called a **measurable space**, that is, it is a space on which we can put a measure. A **measure** is a nonnegative countably additive set function; that is, a function $\mu : \mathcal{F} \to \mathbf{R}$ with

(i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$, and
(ii) if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu(\cup_i A_i) = \sum_i \mu(A_i)$$

If $\mu(\Omega) = 1$, we call μ a **probability measure**. In this book, probability measures are usually denoted by P.

The next result gives some consequences of the definition of a measure that we will need later. In all cases, we assume that the sets we mention are in $\mathcal{F}$.

Theorem 1.1.1. *Let μ be a measure on $(\Omega, \mathcal{F})$*

(i) **Monotonicity.** *If $A \subset B$ then $\mu(A) \le \mu(B)$.*

(ii) **Subadditivity.** *If $A \subset \cup_{m=1}^\infty A_m$ then $\mu(A) \le \sum_{m=1}^\infty \mu(A_m)$.*

(iii) **Continuity from below.** *If $A_i \uparrow A$ (i.e., $A_1 \subset A_2 \subset \ldots$ and $\cup_i A_i = A$) then $\mu(A_i) \uparrow \mu(A)$.*

(iv) **Continuity from above.** *If $A_i \downarrow A$ (i.e., $A_1 \supset A_2 \supset \ldots$ and $\cap_i A_i = A$), with $\mu(A_1) < \infty$ then $\mu(A_i) \downarrow \mu(A)$.*

Proof.

(i) Let $B - A = B \cap A^c$ be the **difference** of the two sets. Using $+$ to denote disjoint union, $B = A + (B - A)$ so

$$\mu(B) = \mu(A) + \mu(B - A) \ge \mu(A).$$

(ii) Let $A'_n = A_n \cap A$, $B_1 = A'_1$ and for $n > 1$, $B_n = A'_n - \cup_{m=1}^{n-1} (A'_m)^c$. Since the B_n are disjoint and have union A, we have, using (i) of the definition of measure, $B_m \subset A_m$, and (i) of this theorem,

$$\mu(A) = \sum_{m=1}^\infty \mu(B_m) \le \sum_{m=1}^\infty \mu(A_m)$$

(iii) Let $B_n = A_n - A_{n-1}$. Then the B_n are disjoint and have $\cup_{m=1}^\infty B_m = A$, $\cup_{m=1}^n B_m = A_n$ so

$$\mu(A) = \sum_{m=1}^\infty \mu(B_m) = \lim_{n\to\infty} \sum_{m=1}^n \mu(B_m) = \lim_{n\to\infty} \mu(A_n)$$

(iv) $A_1 - A_n \uparrow A_1 - A$ so (iii) implies $\mu(A_1 - A_n) \uparrow \mu(A_1 - A)$. Since $A_1 \supset B$, we have $\mu(A_1 - B) = \mu(A_1) - \mu(B)$ and it follows that $\mu(A_n) \downarrow \mu(A)$. ■

The simplest setting, which should be familiar from undergraduate probability, is:

Example 1.1.1. Discrete probability spaces. Let $\Omega =$ a countable set, that is, finite or countably infinite. Let $\mathcal{F} =$ the set of all subsets of Ω. Let

$$P(A) = \sum_{\omega \in A} p(\omega) \text{ where } p(\omega) \ge 0 \text{ and } \sum_{\omega \in \Omega} p(\omega) = 1$$

A little thought reveals that this is the most general probability measure on this space. In many cases when Ω is a finite set, we have $p(\omega) = 1/|\Omega|$ where $|\Omega| =$ the number of points in Ω.

For a simple concrete example that requires this level of generality, consider the astragali, dice used in ancient Egypt made from the ankle bones of sheep. This die could come to rest on the top side of the bone for four points or on the bottom for three points. The side of the bone was slightly rounded. The die could come to rest on a flat and narrow piece for six points or somewhere on the rest of the side for one point. There is no reason to think that all four outcomes are equally likely, so we need probabilities p_1, p_3, p_4, and p_6 to describe P.

To prepare for our next definition, we need:

Exercise 1.1.1. (i) If $\mathcal{F}_i$, $i \in I$ are σ-fields, then $\cap_{i \in I} \mathcal{F}_i$ is. Here $I \neq \emptyset$ is an arbitrary index set (i.e., possibly uncountable). (ii) Use the result in (i) to show that if we are given a set Ω and a collection $\mathcal{A}$ of subsets of Ω, then there is a smallest σ-field containing $\mathcal{A}$. We will call this the **σ-field generated by** $\mathcal{A}$ and denote it by $\sigma(\mathcal{A})$.

Let $\mathbf{R}^d$ be the set of vectors $(x_1, \ldots x_d)$ of real numbers and $\mathcal{R}^d$ be the **Borel sets,** the smallest σ-field containing the open sets. When $d = 1$, we drop the superscript.

Example 1.1.2. Measures on the real line. Measures on $(\mathbf{R}, \mathcal{R})$ are defined by giving probability a **Stieltjes measure function** with the following properties:
(i) F is nondecreasing.
(ii) F is right continuous, that is, $\lim_{y \downarrow x} F(y) = F(x)$.

Theorem 1.1.2. *Associated with each Stieltjes measure function F there is a unique measure μ on $(\mathbf{R}, \mathcal{R})$ with $\mu(a, b]) = F(b) - F(a)$*

$$\mu((a, b]) = F(b) - F(a) \tag{1.1.1}$$

When $F(x) = x$ the resulting measure is called **Lebesgue measure**.

The proof of Theorem 1.1.2 is a long and winding road, so we will content ourselves with describing the main ideas involved in this section and hide the remaining details in the Appendix in Section A.1. The choice of "closed on the right" in $(a, b]$ is dictated by the fact that if $b_n \downarrow b$ then we have

$$\cap_n (a, b_n] = (a, b]$$

The next definition will explain the choice of "open on the left."

A collection $\mathcal{S}$ of sets is said to be a **semialgebra** if (i) it is closed under intersection, that is, $S, T \in \mathcal{S}$ implies $S \cap T \in \mathcal{S}$, and (ii) if $S \in \mathcal{S}$ then S^c is a finite disjoint union of sets in $\mathcal{S}$. An important example of a semialgebra is:

Example 1.1.3. $\mathcal{S}_d =$ the empty set plus all sets of the form

$$(a_1, b_1] \times \cdots \times (a_d, b_d] \subset \mathbf{R}^d \quad \text{where } -\infty \leq a_i < b_i \leq \infty$$

The definition in (1.1.1) gives the values of μ on the semialgebra $\mathcal{S}_1$. To go from semialgebra to σ-algebra we use an intermediate step. A collection $\mathcal{A}$ of subsets of Ω is called an **algebra** (or **field**) if $A, B \in \mathcal{A}$ implies A^c and $A \cup B$ are in $\mathcal{A}$. Since $A \cap B = (A^c \cup B^c)^c$, it follows that $A \cap B \in \mathcal{A}$. Obviously a σ-algebra is an algebra. An example in which the converse is false is:

Example 1.1.4. Let $\Omega = \mathbf{Z}$ = the integers. $\mathcal{A}$ = the collection of $A \subset \mathbf{Z}$ so that A or A^c is finite is an algebra.

Lemma 1.1.3. *If $\mathcal{S}$ is a semialgebra, then $\bar{\mathcal{S}}$ = {finite disjoint unions of sets in $\mathcal{S}$} is an algebra, called the* **algebra generated by** $\mathcal{S}$.

Proof. Suppose $A = +_i S_i$ and $B = +_j T_j$, where $+$ denotes disjoint union and we assume the index sets are finite. Then $A \cap B = +_{i,j} S_i \cap T_j \in \bar{\mathcal{S}}$. As for complements, if $A = +_i S_i$ then $A^c = \cap_i S_i^c$. The definition of $\mathcal{S}$ implies $S_i^c \in \bar{\mathcal{S}}$. We have shown that $\bar{\mathcal{S}}$ is closed under intersection, so it follows by induction that $A^c \in \bar{\mathcal{S}}$. ■

Example 1.1.5. Let $\Omega = \mathbf{R}$ and $\mathcal{S} = \mathcal{S}_1$. Then $\bar{\mathcal{S}}_1$ = the empty set plus all sets of the form

$$\cup_{i=1}^k (a_i, b_i] \quad \text{where } -\infty \le a_i < b_i \le \infty$$

Given a set function μ on $\mathcal{S}$, we can extend it to $\bar{\mathcal{S}}$ by

$$\mu\left(+_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$$

By a **measure on an algebra** $\mathcal{A}$, we mean a set function μ with

(i) $\mu(A) \ge \mu(\emptyset) = 0$ for all $A \in \mathcal{A}$, and
(ii) if $A_i \in \mathcal{A}$ are disjoint *and their union is in* $\mathcal{A}$, then

$$\mu\left(\cup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty \mu(A_i)$$

μ is said to be **σ-finite** if there is a sequence of sets $A_n \in \mathcal{A}$ so that $\mu(A_n) < \infty$ and $\cup_n A_n = \Omega$. Letting $A_1' = A_1$ and for $n \ge 2$,

$$A_n' = \cup_{m=1}^n A_m \quad \text{or} \quad A_n' = A_n \cap \left(\cap_{m=1}^{n-1} A_m^c\right) \in \mathcal{A}$$

we can without loss of generality assume that $A_n \uparrow \Omega$ or the A_n are disjoint.

The next result helps us to extend a measure defined on a semialgebra $\mathcal{S}$ to the σ-algebra it generates, $\sigma(\mathcal{S})$

Theorem 1.1.4. *Let $\mathcal{S}$ be a semialgebra and let μ defined on $\mathcal{S}$ have $\mu(\emptyset) = 0$. Suppose (i) if $S \in \mathcal{S}$ is a finite disjoint union of sets $S_i \in \mathcal{S}$ then $\mu(S) = \sum_i \mu(S_i)$, and (ii) if $S_i, S \in \mathcal{S}$ with $S = +_{i\ge1} S_i$ then $\mu(S) \le \sum_{i\ge1} \mu(S_i)$. Then μ has a unique* *extension $\bar{\mu}$ that is a measure on $\bar{\mathcal{S}}$ the algebra generated by $\mathcal{S}$. If $\bar{\mu}$ is sigma-finite then there is a unique extension ν that is a measure on $\sigma(\mathcal{S})$.*

In (ii) above, and in what follows, $i \geq 1$ indicates a countable union, while a plain subscript i or j indicates a finite union. The proof of Theorems 1.1.4 is rather involved, so it is given in Section A.1. To check condition (ii) in the theorem, the following is useful.

Lemma 1.1.5. *Suppose only that (i) holds.*
(a) If $A, B_i \in \bar{\mathcal{S}}$ with $A = +_{i=1}^n B_i$ then $\bar{\mu}(A) = \sum_i \bar{\mu}(B_i)$.
(b) If $A, B_i \in \bar{\mathcal{S}}$ with $A \subset \cup_{i=1}^n B_i$ then $\bar{\mu}(A) \leq \sum_i \bar{\mu}(B_i)$.

Proof. Observe that it follows from the definition that if $A = +_i B_i$ is a finite disjoint union of sets in $\bar{\mathcal{S}}$ and $B_i = +_j S_{i,j}$, then

$$\bar{\mu}(A) = \sum_{i,j} \mu(S_{i,j}) = \sum_i \bar{\mu}(B_i)$$

To prove (b), we begin with the case $n = 1$, $B_1 = B$. $B = A + (B \cap A^c)$ and $B \cap A^c \in \bar{\mathcal{S}}$, so

$$\bar{\mu}(A) \leq \bar{\mu}(A) + \bar{\mu}(B \cap A^c) = \bar{\mu}(B)$$

To handle $n > 1$ now, let $F_k = B_1^c \cap \cdots \cap B_{k-1}^c \cap B_k$ and note

$$\cup_i B_i = F_1 + \cdots + F_n$$
$$A = A \cap (\cup_i B_i) = (A \cap F_1) + \cdots + (A \cap F_n)$$

so using (a), (b) with $n = 1$, and (a) again

$$\bar{\mu}(A) = \sum_{k=1}^n \bar{\mu}(A \cap F_k) \leq \sum_{k=1}^n \bar{\mu}(F_k) = \bar{\mu}(\cup_i B_i) \qquad \blacksquare$$

Proof of Theorem 1.1.2. Let $\mathcal{S}$ be the semialgebra of half-open intervals $(a, b]$ with $-\infty \leq a < b \leq \infty$. To define μ on $\mathcal{S}$, we begin by observing that

$$F(\infty) = \lim_{x \uparrow \infty} F(x) \quad \text{and} \quad F(-\infty) = \lim_{x \downarrow -\infty} F(x) \quad \text{exist}$$

and $\mu((a, b]) = F(b) - F(a)$ makes sense for all $-\infty \leq a < b \leq \infty$ since $F(\infty) > -\infty$ and $F(-\infty) < \infty$.

If $(a, b] = +_{i=1}^n (a_i, b_i]$ then after relabeling the intervals we must have $a_1 = a$, $b_n = b$, and $a_i = b_{i-1}$ for $2 \leq i \leq n$, so condition (i) in Theorem 1.1.4 holds. To check (ii), suppose first that $-\infty < a < b < \infty$, and $(a, b] \subset \cup_{i \geq 1} (a_i, b_i]$ where (without loss of generality) $-\infty < a_i < b_i < \infty$. Pick $\delta > 0$ so that $F(a + \delta) < F(a) + \epsilon$ and pick η_i so that

$$F(b_i + \eta_i) < F(b_i) + \epsilon 2^{-i}$$

The open intervals $(a_i, b_i + \eta_i)$ cover $[a + \delta, b]$, so there is a finite subcover (α_j, β_j), $1 \le j \le J$. Since $(a + \delta, b] \subset \cup_{j=1}^{J} (\alpha_j, \beta_j]$, (b) in Lemma 1.1.5 implies

$$F(b) - F(a + \delta) \le \sum_{j=1}^{J} F(\beta_j) - F(\alpha_j) \le \sum_{i=1}^{\infty} (F(b_i + \eta_i) - F(a_i))$$

So, by the choice of δ and η_i,

$$F(b) - F(a) \le 2\epsilon + \sum_{i=1}^{\infty} (F(b_i) - F(a_i))$$

and since ϵ is arbitrary, we have proved the result in the case $-\infty < a < b < \infty$. To remove the last restriction, observe that if $(a, b] \subset \cup_i (a_i, b_i]$ and $(A, B] \subset (a, b]$ has $-\infty < A < B < \infty$, then we have

$$F(B) - F(A) \le \sum_{i=1}^{\infty} (F(b_i) - F(a_i))$$

Since the last result holds for any finite $(A, B] \subset (a, b]$, the desired result follows. ∎

Measures on $\mathbf{R}^d$

Our next goal is to prove a version of Theorem 1.1.2 for $\mathbf{R}^d$. The first step is to introduce the assumptions on the defining function F. By analogy with the case $d = 1$ it is natural to assume:

(i) It is nondecreasing, that is, if $x \le y$ (meaning $x_i \le y_i$ for all i), then $F(x) \le F(y)$.
(ii) F is right continuous, that is, $\lim_{y \downarrow x} F(y) = F(x)$ (here $y \downarrow x$ means each $y_i \downarrow x_i$).

However this time it is not enough. Consider the following F:

$$F(x_1, x_2) = \begin{cases} 1 & \text{if } x_1, x_2 \ge 1 \\ 2/3 & \text{if } x_1 \ge 1 \text{ and } 0 \le x_2 < 1 \\ 2/3 & \text{if } x_2 \ge 1 \text{ and } 0 \le x_2 < 1 \\ 0 & \text{otherwise} \end{cases}$$

See Figure 1.1 for a picture. A little thought shows that

$$\begin{aligned} \mu((a_1, b_1] \times (a_2, b_2]) &= \mu((-\infty, b_1] \times (-\infty, b_2]) - \mu((-\infty, a_1] \times (-\infty, b_2]) \\ &\quad - \mu((-\infty, b_1] \times (-\infty, a_2]) + \mu((-\infty, a_1] \times (-\infty, a_2]) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \end{aligned}$$

Using this with $a_1 = a_2 = 1 - \epsilon$ and $b_1 = b_2 = 1$ and letting $\epsilon \to 0$, we see that

$$\mu(\{1, 1\}) = 1 - 2/3 - 2/3 + 0 = -1/3$$

0	2/3	1
0	0	2/3
0	0	0

Figure 1.1. Picture of the counterexample.

Similar reasoning shows that $\mu(\{1,0\}) = \mu(\{0,1\} = 2/3$.

To formulate the third and final condition for F to define a measure, let

$$A = (a_1, b_1] \times \cdots \times (a_d, b_d]$$
$$V = \{a_1, b_1\} \times \cdots \times \{a_d, b_d\}$$

where $-\infty < a_i < b_i < \infty$. To emphasize that ∞'s are not allowed, we will call A a finite rectangle. Then $V =$ the vertices of the rectangle A. If $v \in V$, let

$$\text{sgn}\,(v) = (-1)^{\#\text{ of } a\text{'s in } v}$$
$$\Delta_A F = \sum_{v \in V} \text{sgn}\,(v) F(v)$$

We will let $\mu(A) = \Delta_A F$, so we must assume

(iii) $\Delta_A F \geq 0$ for all rectangles A.

Theorem 1.1.6. *Suppose $F : \mathbf{R}^d \to [0,1]$ satisfies (i)–(iii) given above. Then there is a unique probability measure μ on $(\mathbf{R}^d, \mathcal{R}^d)$ so that $\mu(A) = \Delta_A F$ for all finite rectangles.*

Example 1.1.6. Suppose $F(x) = \prod_{i=1}^d F_i(x)$, where the F_i satisfy (i) and (ii) of Theorem 1.1.2. In this case,

$$\Delta_A F = \prod_{i=1}^{d} (F_i(b_i) - F_i(a_i))$$

When $F_i(x) = x$ for all i, the resulting measure is Lebesgue measure on $\mathbf{R}^d$.

Proof. We let $\mu(A) = \Delta_A F$ for all finite rectangles and then use monotonicity to extend the definition to $\mathcal{S}_d$. To check (i) of Theorem 1.1.4, call $A = +_k B_k$ a **regular subdivision** of A if there are sequences $a_i = \alpha_{i,0} < \alpha_{i,1} \ldots < \alpha_{i,n_i} = b_i$ so that each rectangle B_k has the form

$$(\alpha_{1,j_1-1}, \alpha_{1,j_1}] \times \cdots \times (\alpha_{d,j_d-1}, \alpha_{d,j_d}] \quad \text{where } 1 \leq j_i \leq n_i$$

Cambridge University Press
978-0-521-76539-8 - Probability: Theory and Examples, Fourth Edition
Rick Durrett
Excerpt
More information

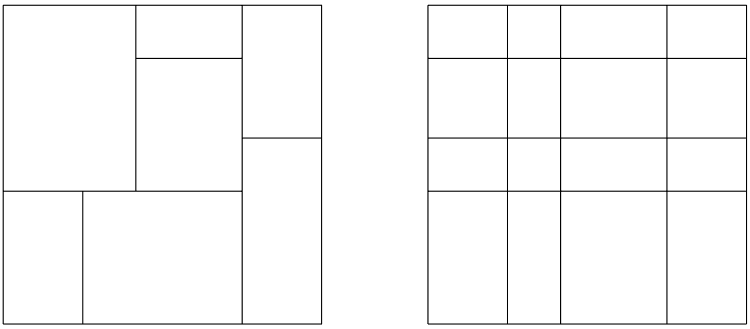

Figure 1.2. Conversion of a subdivision to a regular one.

It is easy to see that for regular subdivisions $\lambda(A) = \sum_k \lambda(B_k)$. (First consider the case in which all the endpoints are finite, and then take limits to get the general case.) To extend this result to a general finite subdivision $A = +_j A_j$, subdivide further to get a regular one see Figure 1.2.

The proof of (ii) is almost identical to that in Theorem 1.1.2. To make things easier to write and to bring out the analogies with Theorem 1.1.2, we let

$$(x, y) = (x_1, y_1) \times \cdots \times (x_d, y_d)$$
$$(x, y] = (x_1, y_1] \times \cdots \times (x_d, y_d]$$
$$[x, y] = [x_1, y_1] \times \cdots \times [x_d, y_d]$$

for $x, y \in \mathbf{R}^d$. Suppose first that $-\infty < a < b < \infty$, where the inequalities mean that each component is finite, and suppose $(a, b] \subset \cup_{i\ge 1}(a^i, b^i]$, where (without loss of generality) $-\infty < a^i < b^i < \infty$. Let $\bar{1} = (1, \ldots, 1)$, pick $\delta > 0$ so that

$$\mu((a + \delta\bar{1}, b]) < \mu((a, b]) + \epsilon$$

and pick η_i so that

$$\mu((a, b^i + \eta_i\bar{1}]) < \mu((a^i, b^i]) + \epsilon 2^{-i}$$

The open rectangles $(a^i, b^i + \eta_i\bar{1})$ cover $[a + \delta\bar{1}, b]$, so there is a finite subcover (α^j, β^j), $1 \le j \le J$. Since $(a + \delta\bar{1}, b] \subset \cup_{j=1}^J (\alpha^j, \beta^j]$, (b) in Lemma 1.1.5 implies

$$\mu([a + \delta\bar{1}, b]) \le \sum_{j=1}^J \mu((\alpha^j, \beta^j]) \le \sum_{i=1}^\infty \mu((a^i, b^i + \eta_i\bar{1}])$$

So, by the choice of δ and η_i,

$$\mu((a, b]) \le 2\epsilon + \sum_{i=1}^\infty \mu((a^i, b^i])$$

and since ϵ is arbitrary, we have proved the result in the case $-\infty < a < b < \infty$. The proof can now be completed exactly as before. ■

Exercises

1.1.2. Let $\Omega = \mathbf{R}$, $\mathcal{F}$ = all subsets so that A or A^c is countable, $P(A) = 0$ in the first case and $= 1$ in the second. Show that $(\Omega, \mathcal{F}, P)$ is a probability space.

1.1.3. Recall the definition of $\mathcal{S}_d$ from Example 1.1.3. Show that $\sigma(\mathcal{S}_d) = \mathcal{R}^d$, the Borel subsets of $\mathbf{R}^d$.

1.1.4. A σ-field $\mathcal{F}$ is said to be **countably generated** if there is a countable collection $\mathcal{C} \subset \mathcal{F}$ so that $\sigma(\mathcal{C}) = \mathcal{F}$. Show that $\mathcal{R}^d$ is countably generated.

1.1.5. (i) Show that if $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ are σ-algebras, then $\cup_i \mathcal{F}_i$ is an algebra. (ii) Give an example to show that $\cup_i \mathcal{F}_i$ need not be a σ-algebra.

1.1.6. A set $A \subset \{1, 2, \ldots\}$ is said to have **asymptotic density** θ if

$$\lim_{n\to\infty} |A \cap \{1, 2, \ldots, n\}|/n = \theta$$

Let $\mathcal{A}$ be the collection of sets for which the asymptotic density exists. Is $\mathcal{A}$ a σ-algebra? an algebra?

1.2 Distributions

Probability spaces become a little more interesting when we define random variables on them. A real-valued function X defined on Ω is said to be a **random variable** if for every Borel set $B \subset \mathbf{R}$ we have $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$. When we need to emphasize the σ-field, we will say that X is $\mathcal{F}$-**measurable** or write $X \in \mathcal{F}$. If Ω is a discrete probability space (see Example 1.1.1), then any function $X : \Omega \to \mathbf{R}$ is a random variable. A second trivial, but useful, type of example of a random variable is the **indicator function** of a set $A \in \mathcal{F}$:

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

The notation is supposed to remind you that this function is 1 on A. Analysts call this object the characteristic function of A. In probability, that term is used for something quite different. (See Section 3.3.)

If X is a random variable, then X induces a probability measure on $\mathbf{R}$ called its **distribution** by setting $\mu(A) = P(X \in A)$ for Borel sets A. Using the notation introduced above, the right-hand side can be written as $P(X^{-1}(A))$. In words, we pull $A \in \mathcal{R}$ back to $X^{-1}(A) \in \mathcal{F}$ and then take P of that set.

To check that μ is a probability measure we observe that if the A_i are disjoint, then using the definition of μ; the fact that X lands in the union if and only if it lands in one of the A_i; the fact that if the sets $A_i \in \mathcal{R}$ are disjoint then the events $\{X \in A_i\}$ are disjoint; and the definition of μ again, we have:

$$\mu(\cup_i A_i) = P(X \in \cup_i A_i) = P(\cup_i \{X \in A_i\}) = \sum_i P(X \in A_i) = \sum_i \mu(A_i)$$

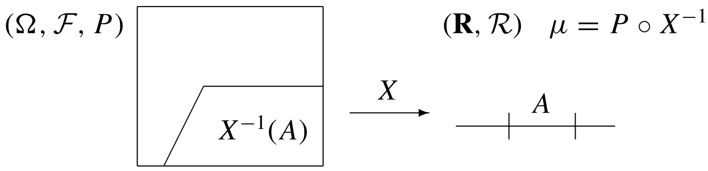

Figure 1.3. Definition of the distribution of X.

The distribution of a random variable X is usually described by giving its **distribution function**, $F(x) = P(X \le x)$.

Theorem 1.2.1. *Any distribution function F has the following properties:*
(i) F is nondecreasing.
(ii) $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.
(iii) F is right continuous, that is, $\lim_{y\downarrow x} F(y) = F(x)$.
(iv) If $F(x-) = \lim_{y\uparrow x} F(y)$ then $F(x-) = P(X < x)$.
(v) $P(X = x) = F(x) - F(x-)$.

Proof. To prove (i), note that if $x \le y$ then $\{X \le x\} \subset \{X \le y\}$, and then use (i) in Theorem 1.1.1 to conclude that $P(X \le x) \le P(X \le y)$.

To prove (ii), we observe that if $x \uparrow \infty$, then $\{X \le x\} \uparrow \Omega$, while if $x \downarrow -\infty$, then $\{X \le x\} \downarrow \emptyset$, and then use (iii) and (iv) of Theorem 1.1.1.

To prove (iii), we observe that if $y \downarrow x$, then $\{X \le y\} \downarrow \{X \le x\}$.

To prove (iv), we observe that if $y \uparrow x$, then $\{X \le y\} \uparrow \{X < x\}$.

For (v), note $P(X = x) = P(X \le x) - P(X < x)$ and use (iii) and (iv). ■

The next result shows that we have found more than enough properties to characterize distribution functions.

Theorem 1.2.2. *If F satisfies (i), (ii), and (iii) in Theroem 1.2.1, then it is the distribution function of some random variable.*

Proof. Let $\Omega = (0, 1)$, $\mathcal{F}$ = the Borel sets, and P = Lebesgue measure. If $\omega \in (0, 1)$, let

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

Once we show that

$$(*) \qquad \{\omega : X(\omega) \le x\} = \{\omega : \omega \le F(x)\}$$

the desired result follows immediately since $P(\omega : \omega \le F(x)) = F(x)$. (Recall P is Lebesgue measure.) To check ($\star$), we observe that if $\omega \le F(x)$ then $X(\omega) \le x$, since $x \notin \{y : F(y) < \omega\}$. On the other hand if $\omega > F(x)$, then since F is right continuous, there is an $\epsilon > 0$ so that $F(x + \epsilon) < \omega$ and $X(\omega) \ge x + \epsilon > x$. ■