CHAPTER 2

Random Variables and their Distributions

We shall not always be interested in an experiment itself, but rather in some consequence of its random outcome. For example, many gamblers are more concerned with their losses than with the games which give rise to them. Such consequences, when real valued, may be thought of as functions which map Ω into the real line $\mathbb{R}$, and these functions are called ‘random variables’.

EXAMPLE 1. A fair coin is tossed twice: $\Omega =$ {HH, HT, TH, TT}. For $\omega \in \Omega$, let $X(\omega)$ be the number of heads, so that

$$X(\mathrm{HH}) = 2,\ X(\mathrm{HT}) = X(\mathrm{TH}) = 1,\ X(\mathrm{TT}) = 0.$$

Now suppose that a gambler wagers his fortune of \$$1$ on the result of this experiment. He gambles cumulatively so that his fortune is doubled each time a head appears, and is annihilated on the appearance of a tail. His subsequent fortune W is a random variable given by

$$W(\mathrm{HH}) = 4,\ W(\mathrm{HT}) = W(\mathrm{TH}) = W(\mathrm{TT}) = 0.$$

After the experiment is done and the outcome $\omega \in \Omega$ is known, a random variable X: $\Omega \to \mathbb{R}$ takes some value. In general this numerical value is more likely to lie in certain subsets of $\mathbb{R}$ than in certain others, depending on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and the function X itself. We wish to be able to describe the distribution of the likelihoods of possible values of *X*. Example 1 above suggests that we might do this through the function $f : \mathbb{R} \to [0, 1]$ defined by

$$f(x) = \text{ probability that } X \text{ is equal to } x,$$

but this turns out to be inappropriate in general. Rather, we use the *distribution function* F: $\mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = \text{ probability that } X \text{ does not exceed } x.$$

More rigorously, this is

$$F(x) = \mathbb{P}(A(x)) \tag{2.1}$$

where $A(x) \subset \Omega$ is given by $A(x) = \{\omega \in \Omega : X(\omega) \leq x\}$. However, $\mathbb{P}$ is a function on the collection $\mathcal{A}$ of events; we cannot discuss $\mathbb{P}(A(x))$ unless $A(x)$ belongs to $\mathcal{A}$, and so we are led to the following definition.

Definition 2. **A random variable** is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}$ for each $x \in \mathbb{R}$. Such a function is said to be **$\mathcal{A}$-measurable**.

Events written as $\{\omega \in \Omega : X(\omega) \in C\}$ for some subset C of $\mathbb{R}$ are commonly abbreviated to $\{\omega : X(\omega) \in C\}$ or $\{X \in C\}$.

We quote without proof the following properties of random variables.

Proposition 3. *(a) If X is a random variable, then for $a, b \in \mathbb{R}$ with $a < b$, the following sets*

$$\{X > a\}, \{a < X \leq b\}, \{X = a\}, \{X < a\}, \{X \geq a\}$$

are all in $\mathcal{A}$.

(b) If X, Y are random variables and c is a constant, then cX and $X + Y$ are also random variables. This means that set of all random variables constitutes a vector space.

(c) If $X_1, X_2, \ldots, X_d$ are random variables and $f : \mathbb{R}^d \to \mathbb{R}$ is continuous, then $f(X_1, \ldots, X_d)$ is again a random variable.

REMARK 4. We shall always use upper-case letters, such as X, Y, and Z, to represent generic random variables, whilst lower-case letters, such as x, y, and z, will be used to represent possible numerical values of these variables.

Every random variable has a distribution function, given by (2.1); distribution functions are *very* important and useful.

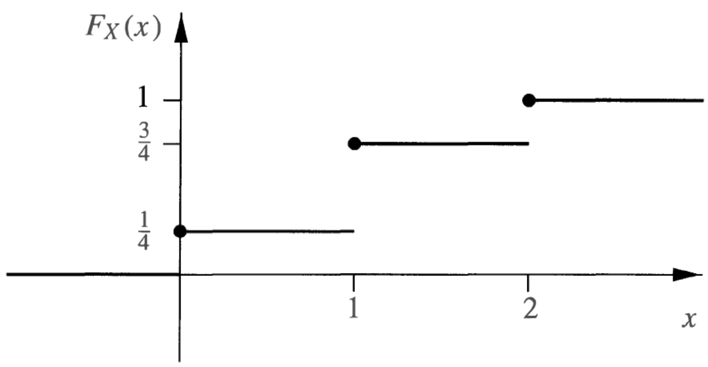
DEFINITION 5. The **distribution function** of a random variable X is the function $F : \mathbb{R} \to [0, 1]$ given by $F(x) = \mathbb{P}(X \leq x)$.

This is the obvious abbreviation of equation (2.1). We write F_X where it is necessary to emphasize the role of X.

EXAMPLE 6. **(Example** 1 **revisited).** The distribution function F_X of X is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{4} & \text{if } 0 \leq x < 1, \\ \frac{3}{4} & \text{if } 1 \leq x < 2, \\ 1 & \text{if } x \geq 2. \end{cases}$$

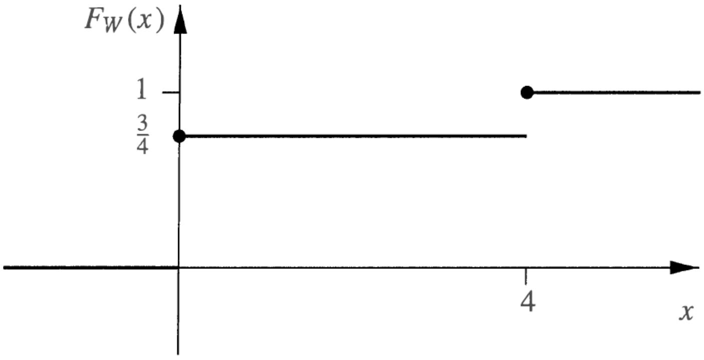
and is sketched as follows:



The distribution function F_W of W is given by

$$F_W(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{3}{4} & \text{if } 0 \leq x < 4, \\ 1 & \text{if } x \geq 4, \end{cases}$$

and is sketched as below.



This illustrates the important point that the distribution function of a random variable X tells us about the values taken by X and their relative likelihoods, rather than about the sample space and the collection of events.

LEMMA 7. *A distribution function F has the following properties:*

(a) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$,
(b) *if $x < y$, then $F(x) \leq F(y)$,*
(c) *F is right-continuous, that is, $F(x+h) \to F(x)$ as $h \downarrow 0$.*

PROOF. (a) Let $B_n = \{\omega \in \Omega \ : \ X(\omega) \leq -n\} = \{X \leq -n\}$. The sequence $B_1, B_2, \ldots$ is decreasing with the empty set as limit. Thus, by Proposition 1.14 (vii), $\mathbb{P}(B_n) \to \mathbb{P}(\emptyset) = 0$. The other part is similar.

(b) Let $A(x) = \{X \leq x\}, A(x, y) = \{x < X \leq y\}$. Then $A(y) = A(x) \cup A(x, y)$ is a disjoint union, and so by Definition 5,

$$\mathbb{P}(A(y)) = \mathbb{P}(A(x)) + \mathbb{P}(A(x,\ y))$$

giving

$$F(y) = F(x) + \mathbb{P}(x < X \le y) \ge F(x).$$

(c) Let $B_h = \{X \le x + h\}$. Then $B_h \downarrow \{X \le x\}$ as $h \downarrow 0$. By continuity of probability measures, $F(x+h) = \mathbb{P}(B_h) \downarrow \mathbb{P}(X \le x) = F(x)$ as $h \downarrow 0$. □

Actually, this lemma characterizes distribution functions. That is to say, F is the distribution function of some random variable if and only if it satisfies (a), (b), and (c) of Lemma 7.

For the time being we can forget all about probability spaces and concentrate on random variables and their distribution functions. The distribution function F of X contains a great deal of information about X.

EXAMPLE 8. **(Constant variables).** The simplest random variable takes a constant value on the whole domain Ω. Let $c \in \mathbb{R}$ and define $X\colon \Omega \to \mathbb{R}$ by

$$X(\omega) = c \text{ for all } \omega \in \Omega.$$

The distribution function $F(x) = \mathbb{P}(X \le x)$ is the step function

$$F(x) = \begin{cases} 0 & x < c, \\ 1 & x \ge c. \end{cases}$$

Slightly more generally, we call X is *almost surely constant* if there exists $c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$.

EXAMPLE 9. **(Indicator functions).** Let A be an event and let $I_A\colon \Omega \to \mathbb{R}$ be the *indicator function* of A; that is,

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \in A^{\mathrm{c}} \end{cases}$$

Then I_A is a random variable taking the values 1 and 0 with probabilities $\mathbb{P}(A)$ and $\mathbb{P}(A^{\mathrm{c}})$ respectively. Its distribution function $F(x) = \mathbb{P}(X \le x)$ is

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - \mathbb{P}(A) & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases}$$

X is said to have the *Bernoulli distribution* which is sometimes denoted by Bern(p) with $p := \mathbb{P}(A)$.

LEMMA 10. *Let F be the distribution function of X. Then*

(a) $\mathbb{P}(X > x) = 1 - F(x)$,

(b) $\mathbb{P}(x < X \le y) = F(y) - F(x)$,

(c) $\mathbb{P}(X = x) = F(x) - \lim_{y \uparrow x} F(y)$.

PROOF. (a) and (b) are *exercises*.

(c) Let $B_n = \{x - 1/n < X \le x\}$. Then $B_n \downarrow \{x\}$. By continuity of probability measures, we have

$$\begin{aligned} \mathbb{P}(X = x) &= \lim_{n\to\infty} \mathbb{P}(x - 1/n < X \le x) \\ &= \lim_{n\to\infty} [F(x) - F(x - 1/n)] \\ &= F(x) - \lim_{n\to\infty} F(x - 1/n) = F(x) - \lim_{y\uparrow x} F(y). \end{aligned}$$

□