MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018

Lecture 4

RANDOM VARIABLES

Contents

1 RANDOM VARIABLES AND MEASURABLE FUNCTIONS

Loosely speaking a random variable provides us with a numerical value, depending on the outcome of an experiment. More precisely, a random variable can be viewed as a function from the sample space to the real numbers, and we will use the notation $X(\omega)$ to denote the numerical value of a random variable X, when the outcome of the experiment is some particular ω. We may be interested in the probability that the outcome of the experiment is such that X is no larger than some c, i.e., that the outcome belongs to the set $\{\omega \mid X(\omega) \leq c\}$. Of course, in order to have a probability assigned to that set, we need to make sure that it is $\mathcal{F}$-measurable. This motivates Definition 1 below.

Example 1. Consider a sequence of five consecutive coin tosses. An appropriate sample space is $\Omega = \{0, 1\}^n$, where "1" stands for heads and "0" for tails. Let $\mathcal{F}$ be the collection of all subsets of Ω, and suppose that a probability measure $\mathbb{P}$ has been assigned to $(\Omega, \mathcal{F})$. We are interested in the number of heads obtained in this experiment. This quantity can be described by the function $X : \Omega \to \mathbb{R}$, defined by

$$X(\omega_1, \ldots, \omega_n) = \omega_1 + \cdots + \omega_n.$$

With this definition, the set $\{\omega \mid X(\omega) < 4\}$ is just the event that there were fewer than 4 heads overall, belongs to the σ-field $\mathcal{F}$, and therefore has a well-defined probability.

Consider the real line, and let $\mathcal{B}$ be the associated Borel σ-field. Sometimes, we will also allow random variables that take values in the extended real line, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. We define the Borel σ-field on $\overline{\mathbb{R}}$, also denoted by $\mathcal{B}$, as the smallest σ-field that contains all Borel subsets of $\mathbb{R}$ and the sets $\{-\infty\}$ and $\{\infty\}$.

1.1 Random variables

Definition 1. *(***Random variables***) Let $(\Omega, \mathcal{F})$ be a measurable space.*

(a) A function $X : \Omega \to \mathbb{R}$ is a **random variable** *if the set $\{\omega \mid X(\omega) \le c\}$ is $\mathcal{F}$-measurable for every $c \in \mathbb{R}$.*

(b) A function $X : \Omega \to \overline{\mathbb{R}}$ is an **extended-valued random variable** *if the set $\{\omega \mid X(\omega) \le c\}$ is $\mathcal{F}$-measurable for every $c \in \overline{\mathbb{R}}$.*

We note here a convention that will be followed throughout: we will always use upper case letters to denote random variables and lower case letters to denote numerical values (elements of $\overline{\mathbb{R}}$). Thus, a statement such as "$X(\omega) = x = 5$" means that when the outcome happens to be ω, then the realized value of the random variable is a particular number x, equal to 5.

Example 2. (**Indicator functions**) Suppose that $A \subset \Omega$, and let $I_A : \Omega \to \{0, 1\}$ be the indicator function of that set; i.e., $I_A(\omega) = 1$, if $\omega \in A$, and $I_A(\omega) = 0$, otherwise. If $A \in \mathcal{F}$, then I_A is a random variable. But if $A \notin \mathcal{F}$, then I_A is not a random variable.

Example 3. (A function of a random variable) Suppose that X is a random variable, and let us define a function $Y : \Omega \to \mathbb{R}$ by letting $Y(\omega) = X^3(\omega)$, for every $\omega \in \Omega$, or $Y = X^3$ for short. We claim that Y is also a random variable. Indeed, for any $c \in \mathbb{R}$, the set $\{\omega \mid Y(\omega) \le c\}$ is the same as the set $\{\omega \mid X(\omega) \le c^{1/3}\}$, which is in $\mathcal{F}$, since X is a random variable.

1.2 The law of a random variable

For a random variable X, the event $\{\omega \mid X(\omega) \le c\}$ is often written as $\{X \le c\}$, and is sometimes just called "the event that $X \le c$." The probability of this event is well defined, since this event belongs to $\mathcal{F}$. Let now B be a more general subset of the real line. We use the notation $X^{-1}(B)$ or $\{X \in B\}$ to denote the set $\{\omega \mid X(\omega) \in B\}$.

Because the collection of intervals of the form $(-\infty, c]$ generates the Borel σ-field in $\mathbb{R}$, it can be shown that if X is a random variable, then for any Borel set B, the set $X^{-1}(B)$ is $\mathcal{F}$-measurable. It follows that the probability $\mathbb{P}\big(X^{-1}(B)\big) = \mathbb{P}(\{\omega \mid X(\omega) \in B\})$ is well-defined. It is often denoted by $\mathbb{P}(X \in B)$.

Exercise 1. *Suppose X is a random variable. Show that for every Borel subset $B \subset \mathbb{R}$, the set $X^{-1}(B)$ is $\mathcal{F}$-measurable. (Hint: Define the collection $\mathcal{L} = \{B : X^{-1}(B) \in \mathcal{F}\}$.)*

Definition 2. (The probability law of a random variable) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable.*

(a) For every Borel subset B of the real line (i.e., $B \in \mathcal{B}$), we define $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$.

(b) The resulting function $\mathbb{P}_X : \mathcal{B} \to [0, 1]$ is called the **probability law** *of X.*

Sometimes, $\mathbb{P}_X$ is also called the distribution of X, not to be confused with the cumulative distribution function defined in the next section.

According to the next result, the law $\mathbb{P}_X$ of X is also a probability measure. Notice here that $\mathbb{P}_X$ is a measure on $(\mathbb{R}, \mathcal{B})$, as opposed to the original measure $\mathbb{P}$, which is a measure on $(\Omega, \mathcal{F})$. In many instances, the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$ remains in the background, hidden or unused, and one works directly with the much more tangible probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P}_X)$. Indeed, if we are only interested in the statistical properties of the random variable X, the latter space will do.

Proposition 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X be a random variable. Then, the law $\mathbb{P}_X$ of X is a probability measure on $(\mathbb{R}, \mathcal{B})$.*

Proof: Clearly, $\mathbb{P}_X(B) \geq 0$, for every Borel set B. Also, $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$. We now verify countable additivity. Let $\{B_i\}$ be a countable sequence of disjoint Borel subsets of $\mathbb{R}$. Note that the sets $X^{-1}(B_i)$ are also disjoint, and that

$$X^{-1}\big(\cup_{i=1}^{\infty} B_i\big) = \cup_{i=1}^{\infty} X^{-1}(B_i),$$

or, in different notation,

$$\{X \in \cup_{i=1}^{\infty} B_i\} = \cup_{i=1}^{\infty} \{X \in B_i\}.$$

Therefore, using countable additivity on the original probability space, we have

$$\mathbb{P}_X\big(\cup_{i=1}^{\infty} B_i\big) \quad = \quad \mathbb{P}\big(X \in \cup_{i=1}^{\infty} B_i\big) = \sum_{i=1}^{\infty} \mathbb{P}(X \in B_i) = \sum_{i=1}^{\infty} \mathbb{P}_X(B_i).$$

$\square$

1.3 Technical digression: measurable functions

The following generalizes the definition of a random variable.

Definition 3. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable space. A function $f : \Omega_1 \to \Omega_2$ is called $(\mathcal{F}_1, \mathcal{F}_2)$-***measurable** *(or just measurable, if the relevant σ-fields are clear from the context) if $f^{-1}(B) \in \mathcal{F}_1$ for every $B \in \mathcal{F}_2$.*

According to the above definition, given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and taking into account the discussion in Section 1.1, a random variable on a probability space is a function $X : \Omega \to \mathbb{R}$ that is $(\mathcal{F}, \mathcal{B})$-measurable.

As a general rule, functions constructed from other measurable functions using certain simple operations are measurable. We collect, without proof, a number of relevant facts below.

Theorem 1. *Let $(\Omega, \mathcal{F})$ be a measurable space.*

(a) **(Simple random variables)** *If $A \in \mathcal{F}$, the corresponding indicator function I_A is measurable (more, precisely, it is $(\mathcal{F}, \mathcal{B})$-measurable).*

(b) If $A_1, \ldots, A_n$ are $\mathcal{F}$-measurable sets, and $x_1, \ldots, x_n$ are real numbers, the function $X = \sum_{i=1}^n x_i I_{A_i}$, or in more detail,

$$X(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega), \qquad \forall\, \omega \in \Omega,$$

is a random variable (and is called a **simple** *random variable).*

(c) Suppose that $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$, and that $X : \mathbb{R} \to \mathbb{R}$ is continuous. Then, X is a random variable.

(d) **(Functions of a random variable)** *Let X be a random variable, and suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous (or more generally, $(\mathcal{B}, \mathcal{B})$-measurable). Then, $f(X)$ is a random variable.*

(e) **(Functions of multiple random variables)** *Let $X_1, \ldots, X_n$ be random variables, and suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuous. Then, $f(X_1, \ldots, X_n)$ is a random variable. In particular, $X_1 + X_2$ and $X_1 X_2$ are random variables.*

Another way that we can form a random variable is by taking the limit of a sequence of random variables. Let us first introduce some terminology. Let each f_n be a function from some set Ω into $\mathbb{R}$. Consider a new function $f = \inf_n f_n$ defined by $f(\omega) = \inf_n f_n(\omega)$, for every $\omega \in \Omega$. The functions $\sup_n f_n$, $\liminf_{n\to\infty} f_n$, and $\limsup_{n\to\infty} f_n$ are defined similarly. (Note that even if the f_n are everywhere finite, the above defined functions may turn out to be extended-valued.) If the limit $\lim_{n\to\infty} f_n(\omega)$ exists for every ω, we say that the sequence of functions $\{f_n\}$ **converges pointwise**, and define its **pointwise limit** to be the function f defined by $f(\omega) = \lim_{n\to\infty} f_n(\omega)$. For example, suppose that $\Omega = [0, 1]$ and that $f_n(\omega) = \omega^n$. Then, the pointwise limit $f = \lim_{n\to\infty} f_n$ exists, and satisfies $f(1) = 1$, and $f(\omega) = 0$ for $\omega \in [0, 1)$.

Theorem 2. *Let $(\Omega, \mathcal{F})$ be a measurable space. If X_n is a random variable for every n, then $\inf_n X_n$, $\sup_n X_n$, $\liminf_{n\to\infty} X_n$, and $\limsup_{n\to\infty} X_n$ are random variables. Furthermore, if the sequence $\{X_n\}$ converges pointwise, and $X = \lim_{n\to\infty} X_n$, then X is also a random variable.*

As a special case of Theorem 2, we have that a pointwise limit of a sequence of simple random variables (with "simple" defined in the statement of Theorem 1) is measurable. On the other hand, we note that measurable functions can be highly discontinuous. For example the function which equals 1 at every rational number, and equals zero otherwise, is measurable, because it is the indicator function of a measurable set.

2 CUMULATIVE DISTRIBUTION FUNCTIONS

A simple way of describing the probabilistic properties of a random variable is in terms of the cumulative distribution function, which we now define.

Definition 4. (Cumulative distribution function) *Let X be a random variable. The function $F_X : \mathbb{R} \to [0,1]$, defined by*

$$F_X(x) = \mathbb{P}(X \leq x),$$

*is called the cumulative distribution function (***CDF***) of X.*

Example 4. Let X be the number of heads in two independent tosses of a fair coin. In particular, $\mathbb{P}(X=0) = \mathbb{P}(X=2) = 1/4$, and $\mathbb{P}(X=1) = 1/2$. Then,

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

Example 5. (A uniform random variable and its square) Consider a probability space $(\Omega, \mathcal{B}, \mathbb{P})$, where $\Omega = [0,1]$, $\mathcal{B}$ is the Borel σ-field $\mathcal{B}$, and $\mathbb{P}$ is the Lebesgue measure. The random variable U defined by $U(\omega) = \omega$ is said to be uniformly distributed. Its CDF is given by

$$F_U(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

Consider now the random variable $X = U^2$. We have $\mathbb{P}(U^2 \leq x) = 0$, when $x < 0$, and $\mathbb{P}(U^2 \leq x) = 1$, when $x \geq 1$. For $x \in [0,1)$, we have

$$\mathbb{P}(U^2 \leq x) = \mathbb{P}\big(\{\omega \in [0,1] \mid \omega^2 \leq x\}\big) = \mathbb{P}\big(\{\omega \in [0,1] : \omega \leq \sqrt{x}\}\big) = \sqrt{x}.$$

Thus,

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

2.1 CDF properties

The cumulative distribution function of a random variable always possesses certain properties.

Theorem 3. *Let X be a random variable, and let F be its CDF.*

(a) **(Monotonicity)** *If $x \leq y$, then $F_X(x) \leq F_X(y)$.*

(b) **(Limiting values)** *We have $\lim_{x\to-\infty} F_X(x) = 0$, and $\lim_{x\to\infty} F(x) = 1$.*

(c) **(Right-continuity)** *For every x, we have $\lim_{y\downarrow x} F_X(y) = F_X(x)$.*

Proof:

(a) Suppose that $x \leq y$. Then, $\{X \leq x\} \subset \{X \leq y\}$, which implies that

$$F(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F(y).$$

(b) Since $F_X(x)$ is monotonic in x and bounded below by zero, it converges as $x \to -\infty$, and the limit is the same for every sequence $\{x_n\}$ converging to $-\infty$. So, let $x_n = -n$, and note that the sequence of events $\cap_{n=1}^{\infty}\{X \leq -n\}$ converges to the empty set. Using the continuity of probabilities, we obtain

$$\lim_{x\to-\infty} F_X(x) = \lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}(X \leq -n) = \mathbb{P}(\emptyset) = 0.$$

The proof of $\lim_{x\to\infty} F_X(x) = 1$ is similar, and is omitted.

(c) Consider a decreasing sequence $\{x_n\}$ that converges to x. The sequence of events $\{X \leq x_n\}$ is decreasing and $\cap_{n=1}^{\infty}\{X \leq x_n\} = \{X \leq x\}$. Using the continuity of probabilities, we obtain

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(X \leq x_n) = \mathbb{P}(X \leq x) = F_X(x).$$

Since this is true for every such sequence $\{x_n\}$, we conclude that $\lim_{y\downarrow x} F_X(y) = F_X(x)$. □

We note that CDFs need not be left-continuous. For instance, in Example 4, we have $\lim_{x\uparrow 0} F_X(x) = 0$, but $F_X(0) = 1/4$.

2.2 From a CDF to a probability law

Consider a function $F : \mathbb{R} \to [0, 1]$ that satisfies the three properties in Theorem 3; we call such a function a **distribution function**. Given a distribution function F, does there exist a random variable X, defined on some probability space, whose CDF, F_X, is equal to the given distribution F? This is certainly the case for the distribution function F function that satisfies $F(x) = x$, for $x \in (0, 1)$: the uniform random variable U in Example 5 will do. More generally, the objective $F_X = F$ can be accomplished by letting $X = g(U)$, for a suitable function $g : (0, 1) \to \mathbb{R}$.

Theorem 4. *Let F be a given distribution function. Consider the probability space $([0, 1], \mathcal{B}, \mathbb{P})$, where $\mathcal{B}$ is the Borel σ-field, and $\mathbb{P}$ is the Lebesgue measure. There exists a measurable function $X : \Omega \to \mathbb{R}$ whose CDF F_X satisfies $F_X = F$.*

Proof: We first present the proof under an additional simplifying assumption that F is continuous and strictly increasing. Then, the range of F is the entire interval $(0, 1)$. Furthermore, F is invertible: for every $y \in (0, 1)$, there exists a unique x, denoted $F^{-1}(y)$, such that $F(x) = y$. We define $U(\omega) = \omega$ and $X(\omega) = F^{-1}(\omega)$, for every $\omega \in (0, 1)$, so that $X = F^{-1}(U)$. Note that $F(F^{-1}(\omega)) = \omega$ for every $\omega \in (0, 1)$, so that $F(X) = U$. Since F is strictly increasing, we have $X \le x$ if and only $F(X) \le F(x)$, or $U \le F(x)$. (Note that this also establishes that the event $\{X \le x\}$ is measurable, so that X is indeed a random variable.) Thus, for every $x \in \mathbb{R}$, we have

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\big(F(X) \le F(x)\big) = \mathbb{P}(U \le F(x)) = F(x),$$

as desired.

The general case is treated by defining a right-continuous inverse of F, called a *quantile function* of F:

$$q(y) = \inf\{x : F(x) > y\} \qquad \forall 0 < y < 1\,.$$

It can be shown that for every y

$$\{x : y < F(x)\} \subseteq \{x : q(y) \le x\} \subseteq \{x : y \le F(x)\}$$

and therefore $F(U) \sim \mathbb{P}$. $\square$

Note that the probability law of X assigns probabilities to all Borel sets, whereas the CDF only specifies the probabilities of certain intervals. Nevertheless, the CDF contains enough information to recover the law of X.

Corollary 1. *There is a one-to-one correspondence between distribution functions F and probability measures $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B})$.*

Proof: Indeed, for any CDF F Theorem 4 constructs a random variable X whose CDF $F_X = F$. On the other hand, for a random variable X, by Proposition 1, the induced set-function $\mathbb{P}_X(\cdot)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$, and given a probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B})$ we obtain a CDF $F(c) = \mathbb{P}_X((-\infty, c])$.

It remains to check that different probability measures $\mathbb{P}_X$ and $\mathbb{P}'_X$ necessarily yield different CDFs. Indeed, if $\mathbb{P}_X$ and $\mathbb{P}'_X$ coincide on all intervals $(-\infty, c]$ then $\mathbb{P}_X = \mathbb{P}'_X$ by Proposition 1 of Lecture 2 (see Remark there) since the collection of sets $\{(-\infty, c], c \in \mathbb{R}\}$ is a generating p-system for $\mathcal{B}$. $\square$

3 DISCRETE RANDOM VARIABLES

Discrete random variables take values in a countable set. We need some notation. Given a function $f : \Omega \to \mathbb{R}$, its **range** is the set

$$f(\Omega) = \{x \in \mathbb{R} \mid \exists\, \omega \in \Omega \text{ such that } f(\omega) = x\}.$$

Definition 5. Discrete random variables and PMFs)

(a) *A random variable X, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is said to be* **discrete** *if its range $X(\Omega)$ is finite or countable.*

(b) *If X is a discrete random variable, the function $p_X : \mathbb{R} \to [0, 1]$ defined by $p_X(x) = \mathbb{P}(X = x)$, for every x, is called the* **(probability) mass function** *of X, or* **PMF** *for short.*

Consider a discrete random variable X whose range is a finite set C. In that case, for any Borel set A, countable additivity yields

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A \cap C) = \sum_{x \in A \cap C} \mathbb{P}(X = x).$$

In particular, the CDF is given by

$$F_X(x) = \sum_{\{y \in A \cap C \,|\, y \leq x\}} \mathbb{P}(X = y).$$

A random variable that takes only integer values is discrete. For instance, the random variable in Example 4 (number of heads in two coin tosses) is discrete.

Also, every simple random variable is discrete, since it takes a finite number of values. However, more complicated discrete random variables are also possible.

Example 6. Let the sample space be the set $\mathbb{N}$ of natural numbers, and consider a measure that satisfies $\mathbb{P}(n) = 1/2^n$, for every $n \in \mathbb{N}$. The random variable X defined by $X(n) = n$ is discrete.

Suppose now that the rational numbers have been arranged in a sequence, and that x_n is the nth rational number, according to this sequence. Consider the random variable Y defined by $Y(n) = x_n$. The range of this random variable is countable, so Y is a discrete random variable. Its range is the set of rational numbers, every rational number has positive probability, and the set of irrational numbers has zero probability.

We close by noting that discrete random variables can be represented in terms of indicator functions. Indeed, given a discrete random variable X, with range $\{x_1, x_2, \ldots\}$, we define $A_n = \{X = x_n\}$, for every $n \in \mathbb{N}$. Observe that each set A_n is measurable (why?). Furthermore, the sets A_n, $n \in \mathbb{N}$, form a partition of the sample space. Using indicator functions, we can write

$$X(\omega) = \sum_{n=1}^{\infty} x_n I_{A_n}(\omega).$$

Conversely, suppose we are given a sequence $\{A_n\}$ of disjoint events, and a real sequence $\{x_n\}$. Define $X : \Omega \to \mathbb{R}$ by letting $X(\omega) = x_n$ if and only if $\omega \in A_n$. Then X is a discrete random variable, and $\mathbb{P}(X = x_n) = \mathbb{P}(A_n)$, for every n.

4 CONTINUOUS RANDOM VARIABLES

The definition of a continuous random variable is more subtle. It is not enough for a random variable to have a "continuous range."

Definition 6. *A random variable X, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is said to be* **continuous** *if there exists a nonnegative measurable function $f : \mathbb{R} \to [0, \infty)$ such that*

$$F_X(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad \forall\, x \in \mathbb{R}. \tag{1}$$

The function f is called a **(probability) density function** *(or* **PDF***, for short) for X,* There is some ambiguity in the above definition, because the meaning of the integral of a measurable function may be unclear. We will see later in this course how such an integral is defined. For now, we just note that the integral is well-defined, and agrees with the Riemann integral encountered in calculus, if the function is continuous, or more generally, if it has a finite number of discontinuities.

Since $\lim_{x\to\infty} F_X(x) = 1$, we must have $\lim_{x\to\infty} \int_{-\infty}^{x} f(t)dt = 1$, or

$$\int_{-\infty}^{\infty} f_X(t)\, dt = 1. \tag{2}$$

Any nonnegative measurable function that satisfies Eq. (2) is called a density function. Conversely, given a density function f, we can define $F(x) = \int_{-\infty}^{x} f(t)\, dt$, and verify that F is a distribution function. It follows that given a density function, there always exists a random variable whose PDF is the given density.

If a CDF F_X is differentiable at some x, the corresponding value $f_X(x)$ can be found by taking the derivative of F_X at that point. However, CDFs need not be differentiable, so this will not always work. Let us also note that a PDF of a continuous random variable is not uniquely defined. We can always change the PDF at a finite set of points, without affecting its integral, hence multiple PDFs can be associated to the same CDF. However, this nonuniqueness rarely becomes an issue. In the sequel, we will often refer to "the PDF" of X, ignoring the fact that it is nonunique.

Example 7. For a uniform random variable, we have $F_X(x) = \mathbb{P}(X \leq x) = x$, for every $x \in (0, 1)$. By differentiating, we find $f_X(x) = 1$, for $x \in (0, 1)$. For $x < 0$ we have $F_X(x) = 0$, and for $x > 1$ we have $F_X(x) = 1$; in both cases, we obtain $f_X(x) = 0$. At $x = 0$, the CDF is not differentiable. We are free to define $f_X(0)$ to be 0, or 1, or in fact any real number; the value of the integral of f_X will remain unaffected.

Example 8. Consider the random variable $X(\omega) = \omega^2$ on $[0, 1]$ which we discussed in Section 1. Let $f(t) = \frac{1}{2\sqrt{t}}$ when $t \in [0, 1]$ and $f(t) = 0$ for all other t. Then for every $x \in [0, 1]$ we have $\int_{-\infty}^{x} f(t)dt = \int_{0}^{x} f(t)dt = \sqrt{x} = F(x)$. We can check trivially that the equality holds for all other values of x. Thus $f(t) = \frac{1}{2\sqrt{t}}$ is the density function corresponding to this probability distribution.

Using the PDF of a continuous random variable, we can calculate the probability of various subsets of the real line. For example, we have $\mathbb{P}(X = x) = 0$, for all x, and if $a < b$,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f_X(t)\, dt.$$

More generally, for any Borel set B, it turns out that

$$\mathbb{P}(X \in B) = \int_B f(t)\,dt = \int I_B(t) f_X(t)\,dt,$$

and that $\mathbb{P}(X \in B) = 0$ whenever B has Lebesgue measure zero. However, more detail on this subject must wait until we develop the theory of integration of measurable functions.

We close by pointing out that not all random variables are continuous or discrete. For example, suppose that X is a discrete random variable, and that Y is a continuous random variable. Fix some $\lambda \in (0, 1)$, and define

$$F(z) = \lambda F_X(x) + (1 - \lambda) F_Y(y). \tag{3}$$

It can be verified that F is a distribution function, and therefore can be viewed as the CDF of some new random variable Z. However, the random variable Z is neither discrete, nor continuous. For an interpretation, we can visualize Z being generated as follows: we first generate the random variables X and Y; then, with probability λ, we set $Z = X$, and with probability $1 - \lambda$, we set $Z = Y$.

Even more pathological random variables are possible. Appendix B discusses a particularly interesting one.

5 APPENDIX A – CONTINUOUS FUNCTIONS

We introduce here some standard notation and terminology regarding convergence of function values, and continuity.

(a) Consider a function $f : \mathbb{R}^m \to \mathbb{R}$, and fix some $x \in \mathbb{R}^m$. We say that $f(y)$ **converges** to a value c, as y tends to x, if we have $\lim_{n\to\infty} f(x_n) = c$, for every sequence $\{x_n\}$ of elements of $\mathbb{R}^m$ such that $x_n \neq x$ for all n, and $\lim_{n\to\infty} x_n = x$. In this case, we write $\lim_{y\to x} f(y) = c$.

(b) If $f : \mathbb{R}^m \to \mathbb{R}$ and $\lim_{y\to x} f(y) = f(x)$, we say that f is **continuous** at x. If this holds for every x, we say that f is continuous.

(c) Consider a function $f : \mathbb{R} \to \mathbb{R}$, and fix some $x \in \mathbb{R} \cup \{-\infty\}$. We say that $f(y)$ **converges** to a value c, as y decreases to x, if we have $\lim_{n\to\infty} f(x_n) = c$, for every decreasing sequence $\{x_n\}$ of elements of $\mathbb{R}^m$ such that $x_n > x$ for all n, and $\lim_{n\to\infty} x_n = x$. In this case, we write $\lim_{y\downarrow x} f(y) = c$.

(d) If $f : \mathbb{R} \to \mathbb{R}$ and $\lim_{y\downarrow x} f(y) = f(x)$, we say that the function f is **right-continuous** at x. If this holds for every $x \in \mathbb{R}$, we say that f is right-continuous.

(e) Consider a function $f : \mathbb{R} \to \mathbb{R}$, and fix some $x \in \mathbb{R} \cup \{\infty\}$. We say that $f(y)$ **converges** to a value c, as y increases to x, if we have $\lim_{n\to\infty} f(x_n) = c$, for every increasing sequence $\{x_n\}$ of elements of $\mathbb{R}$ such that $x_n < x$ for all n, and $\lim_{n\to\infty} x_n = x$. In this case, we write $\lim_{y\uparrow x} f(y) = c$.

(f) If $f : \mathbb{R} \to \mathbb{R}$ and $\lim_{y\uparrow x} f(y) = f(x)$, we say that the function f is **left-continuous** at x. If this holds for every $x \in \mathbb{R}$, we say that f is left-continuous.

For a function $f : \mathbb{R} \to \mathbb{R}$, and some $x \in \mathbb{R}$, it is not hard to show, starting from the above definitions, that f is continuous at x if and only if it is both left-continuous and right-continuous.

6 APPENDIX B – THE CANTOR SET, AN UNUSUAL RANDOM VARIABLE, AND A SINGULAR MEASURE

Every number $x \in [0, 1]$ has a ternary expansion of the form

$$x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \qquad \text{with } x_i \in \{0, 1, 2\}. \tag{4}$$

This expansion is not unique. For example, $1/3$ admits two expansions, namely $.10000\cdots$ and $.022222\cdots$. Nonuniqueness occurs only for those x that admit an expansion ending with an infinite sequence of 2s. The set of such unusual x is countable, and therefore has Lebesgue measure zero.

The **Cantor set** C is defined as the set of all $x \in [0, 1]$ that have a ternary expansion that uses only 0s and 2s (no 1s allowed). The set C can be constructed as follows. Start with the interval $[0, 1]$ and remove the "middle third" $(1/3, 2/3)$. Then, from each of the remaining closed intervals, $[0, 1/3]$ and $[2/3, 1]$, remove their middle thirds, $(1/9, 2/9)$ and $(7/9, 8/9)$, resulting in four closed intervals, and continue this process indefinitely. Note that C is measurable, since it is constructed by removing a countable sequence of intervals. Also, the length (Lebesgue measure) of C is 0, since at each stage its length is mutliplied by a factor of 2/3. On the other hand, the set C has the same cardinality as the set $\{0, 2\}^\infty$, and is uncountable.

Consider now an infinite sequence of independent rolls of a 3-sided die, whose faces are labeled 0, 1, and 2. Assume that at each roll, each of the three possible results has the same probability, 1/3. If we use the sequence of these rolls to form a number x, then the probability law of the resulting random variable is the Lebesgue measure (i.e., picking a ternary expansion "at random" leads to a uniform random variable).

The Cantor set can be identified with the event consisting of all roll sequences in which a 1 never occurs. (This event has zero probability, which is consistent with the fact that C has zero Lebesgue measure.)

Consider now an infinite sequence of independent tosses of a fair coin. If the ith toss results in tails, record $x_i = 0$; if it results in heads, record $x_i = 2$. Use the x_is to form a number x, using Eq. (4). This defines a random variable X on $([0,1], \mathcal{B})$, whose range is the set C. The probability law of this random variable is therefore concentrated on the "zero-length" set C. At the same time, $\mathbb{P}(X = x) = 0$ for every x, because any particular sequence of heads and tails has zero probability. A measure with this property is called **singular**.

The random variable X that we have constructed here is neither discrete nor continuous. Moreover, the CDF of X cannot be written as a mixture of the kind considered in Eq. (3).