Section 1.5. Random Variables

Note. We now assign a numerical value to each element c of the sample space $\mathcal{C}$. We use this assignment (called a "random variable") to associate probabilities with the events (i.e., subsets) in $\mathcal{C}$.

Definition 1.5.1. Consider a random experiment with a sample space $\mathcal{C}$. A function X which assigns to each $c \in \mathcal{C}$ one and only one real number $X(c) = x$ is a *random variable.* The *space* (or *range*) of X is the set of real numbers $\mathcal{D} = \{x \mid x = X(c) \text{ for some } c \in \mathcal{C}\}$. If $\mathcal{D}$ is a countable set then X is a *discrete random variable* and if $\mathcal{D}$ is an interval of real numbers then X is a *continuous random variable.*

Note. We discuss discrete random variables in more detail in the next section (Section 1.6) and discuss continuous random variables in Section 1.7. We illustrate some special cases now.

Note/Definition. Suppose X is a discrete random variable with a finite space (i.e., range) $\mathcal{D} = \{d_1, d_2, \ldots, d_m\}$. Define the function $p_X : \mathcal{D} \to [0, 1]$ as $p_X(d_i) = P(\{c \mid X(c) = d_i\})$ for $i = 1, 2, \ldots, m$. In Section 1.6, p_X is called the *probability mass function* of X. The induced probability distribution, P_X, mapping subsets of $\mathcal{D}$ to $[0, 1]$ is $P_X(D) = \sum_{d_i \in D} p_X(d_i)$ for $D \subset \mathcal{D}$. Exercise 1.5.11 shows that $P_X(D)$ actually *is* a probability on $\mathcal{D}$.

Example 1.5.1. Rolling Two Dice. Let X be the sum of the surfaces on a roll of a pair of fair 6-sided dice with faces labeled 1 through 6. The sample space is the set of ordered pairs (we assume one die can be distinguished from the other) $\mathcal{C} = \{(i,j) \mid 1 \le i,j \le 6\}$. Since the dice are fair, $P(\{(i,j)\}) = 1/36$. The 36 outcomes are:

$$\begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6). \end{array}$$

Define random variable X on the elements of $\mathcal{C}$ as $X((i,j)) = i + j$ (so the space is $\mathcal{D} = \{2,3,\ldots,11,12\}$). that the probability mass function satisfies:

x	2	3	4	5	6	7	8	9	10	11	12
$p_X(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Suppose $B_1 = \{x \mid x = 7, 11\} = \{7, 11\}$ and $B_2 = \{2, 3, 12\}$ (notice that B_1 and B_2 are subset of $\mathcal{D}$). Then

$$P_X(B_1) = \sum_{x \in B_1} p_X(x) = p_X(7) + p_x(11) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36},$$

$$P_X(B_2) = \sum_{x \in B_2} p_X(x) = p_X(2) + p_x(3) + p_X(12) = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36}.$$

Note/Definition. In practice, when X is a continuous random variable, it often represents measurements (the book refers to a person's weight as an example; see page 38). In this case, we usually have a nonnegative function $f_X(x)$ such that for intervals $(a, b) \in \mathcal{D}$, the induced probability distribution of X, P_X, is defined as

$$P_X((a,b)) = P(\{c \in \mathcal{C} \mid a < X(c) < b\}) = \int_a^b f_X(x)\, dx.$$

Such an f_X will be called a *probability density function* (or *pdf*) in X in Section 1.7. Notice that we require $\int_{\mathcal{D}} f_X(x)\, dx = 1$ so that the definition of probability is satisfied by P_X.

Note. So, with $\mathcal{C}$ as the sample space, we have the random variable $X : \mathcal{C} \to \mathbb{R}$, the space (or range) $\mathcal{D} = \text{range}(X) \subset \mathbb{R}$, and the probability distribution P_X mapping subsets of $\mathcal{D}$ to $[0, 1]$ where P_X is defined by summing on investigating the probability denoted mass function f_X or p_X over the subset of $\mathcal{D}$.

Example 1.5.2. Consider the probability density function

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then for a, b with $0 \le a < b \le 1$ we have

$$P_X((a,b)) = \int_a^b f_X(x)\, dx = \int_a^b 1\, dx = b - a,$$

and, in particular, $P_X(\mathbb{R}) = P_X((0,1)) = 1$. Notice that P_X is defined on all open intervals of real numbers. Since for Riemann integrals, $\int_a^a f_X(x)\, dx = 0$ then P_x is also defined on all single points (where it is zero; $P_X(\{x\}) = 0$ for all $x \in \mathbb{R}$). We can then define P_X for any closed interval $[a, b] \subset [0, 1]$ also as $P_X([a, b]) = \int_a^b f_X(x)\,dx = b - a$. We want the subsets of $\mathbb{R}$ (or subsets of $[0, 1]$) to form a σ-field and, if we restrict ourselves to Riemann integrals, then we have problems. A σ-field which contains all open intervals also contains all Borel sets (see my online notes for Real Analysis 1, MATH 5210, on 1.4. Borel Sets). An example of a Borel set is $S = \{x \in [0, 1] \mid x \in \mathbb{R} \setminus \mathbb{Q}\}$ (the irrational numbers between 0 and 1). We cannot define a Riemann integral of f_X over set S. But the Lebesgue integral of f_X over X doesn't exist. In fact, if S is any Borel set then $P_X(S) = m(S)$ where $m(S)$ denotes the Lebesgue measure of set S. More generally, if S is any measurable set (the Borel sets are a proper subset of the Lebesgue measurable sets) then $P_X(S) = m(S)$). This discussion is evidence as to why measure theory and Lebesgue integration play such an important role in probability and statistics.

Definition 1.5.2. Let X be a random variable. Its *cumulative distribution function* (or *cdf*) mapping $F_X : \mathbb{R} \to [0, 1]$ is

$$F_X(x) = P_X((-\infty, x]) = P(\{x \in \mathcal{C} \mid X(c) \le x\} = P(X \le x).$$

Exercise 1.5.4. In Example 1.5.2 we have the probability density function

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $x \le 0$ then $P(X \le x) = \int_{-\infty}^x f_X(t)\,dt = \int_{-\infty}^x 0\,dt = 0$, if $0 < x < 1$ then $P(X \le x) = \int_{-\infty}^x f_X(t)\,dt = \int_{-\infty}^0 f_X(t)\,dt + \int_0^x f_X(t)\,dt = \int_{-\infty}^0 0\,dt + \int_0^x 1\,dt = x$, and if $x \geq 1$ then $P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{0} f_X(t)\,dt + \int_0^1 f_X(t)\,dt \int_1^x f_X(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^1 1\,dt + \int_1^x 0\,dt = 0+1+0 = 1$. So the cumulative distribution function is

$$F_X(x) = P_X((-\infty, x]) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Notice that $\frac{d}{dx}[F_X(x)] = \frac{d}{dx}\left[\int_{-\infty}^{x} f_X(t)\,dt\right] = f_X(t)$ (except at $x = 0$ and $x = 1$ where f_X is not differentiable) by the Fundamental Theorem of Calculus.

Definition. Let X and Y be random variables. Then X and Y are *equal in distribution*, denoted $X \stackrel{D}{=} Y$, if the cumulative distribution functions are equal, $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

Note. We can have $X \stackrel{D}{=} Y$ even though X and Y are different. With $Y = 1 - X$ where X is based on the probability density function of Example 1.5.4 we have $X \neq Y$ and (1) for $y < 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1 - X \leq y) = P(X \geq 1 - y) \\ &= P(X \geq 1 - y) = 1 - P(X < 1 - y) = 1 - (1) \text{ since } 1 - y > 1 \\ &= 0, \end{aligned}$$

(2) for $y \geq 1$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1 - X \leq y) = P(X \geq 1 - y) \\ &= P(X \geq 1 - y) = 1 - P(X < 1 - y) = 1 - (0) \text{ since } 1 - y \leq 0 \end{aligned}$$

$$= 1,$$

and (3) for $0 \leq y < 1$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1 - X \leq y) = P(X \geq 1 - y) \\ &= P(X \geq 1 - y) = 1 - P(X < 1 - y) = 1 - (1 - y) \text{ since } 0 < 1 - y \leq 1 \\ &= y. \end{aligned}$$

So $F_Y(y) = P(Y \leq y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1. \end{cases}$ and hence $X \stackrel{D}{=} Y$.

Note. The following theorem gives some properties of the cumulative density function of a random variable.

Theorem 1.5.1. Let X be a random variable with cumulative distribution function $F(x)$. Then

(a) For all a and b, if $a < b$ then $F(a) \leq F(b)$ (i.e., F is nondecreasing).

(b) $\lim_{x \to -\infty} F(x) = 0$.

(c) $\lim_{x \to \infty} F(x) = 1$.

(d) $\lim_{x \downarrow x_0} F(x) = \lim_{x \to x_0^+} F(x) = F(x_0)$ (i.e., F is right continuous).

Note. Theorem 1.5.1(d) does not hold for $x \uparrow x_0$ or $x \to x_0^-$. That is, the cumulative distribution function need not be continuous from the left. In fact, the cumulative distribution function of the experiment of rolling a fair 6-sided die is not continuous from the right, as shown in Figure 1.5.1.

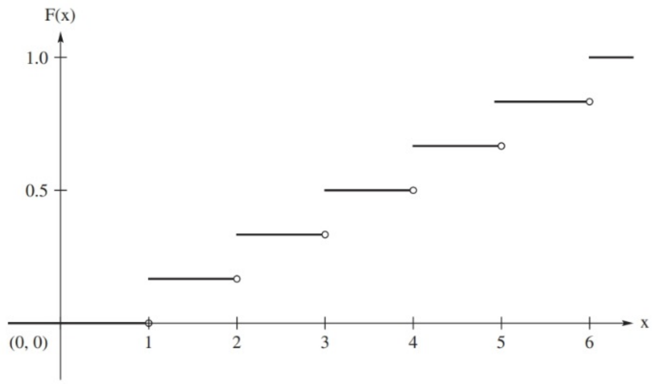

Figure 1.5.1. The cdf for the random variable associated with rolling a fair 6-sided die.

Note. The next formula relates certain probabilities to the cumulative distribution function.

Theorem 1.5.2. Let X be a random variable with cumulative distribution function F_X. Then for $a < b$ we have $P(a < X \le b) = F_X(b) - F_X(a)$.

Note. By Theorem 1.5.1(a), we see that a cumulative distribution function is non-decreasing (i.e., "monotone increasing") and by definition a cdf function is defined on $\mathbb{R}$. Such a function can have at most a countable number of discontinuities (see my online Analysis 1, MATH 4127/5127, notes on 4.2. Monotone and Inverse Functions; see Theorem 4-14). The next theorem shows that the discontinuities of the cdf "has mass"; that is, if F_X is discontinuous at x then $P(X = x) > 0$.

Theorem 1.5.3. For random variable X, $P(X = x) = F_X(x) - F_X(x^-)$ for all $x \in \mathbb{R}$, where $F_X(x^-) = \lim_{z \to x^-} F_X(z)$.

Example 1.5.5. Let X be the lifetime in years of a mechanical part. Assume the cdf for X is

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-x} & \text{for } x \geq 0. \end{cases}$$

If we calculate the probability density function by differentiating F_X then we get

$$\frac{d}{dx}[F_X(x)] = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{for } x < 0 \end{cases}$$

since the derivative is modified at $x = 0$. The textbook was Theorem 1.5.3 to justify defining the pdf f_X at $x = 0$ as $f_X(0) = 0$. However, since the cfd and pdf are related as $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ then we can let $f_X(0)$ be any value because a Riemann integral is unaffected by the value of the integrand at a single point (a Lebesgue integral is also unaffected; however Riemann-Stieltjes and Lebesgue-Stieltjes integrals can be affected by single values of the integrand; see 6.3. The Riemann-Stieltjes Integral from Analysis 2, MATH 4127/5127, and 20.3. Cumulative Distribution Functions and Borel Measures on $\mathbb{R}$ from Real Analysis 2, MATH 5220).

Note. We further explore random variables, pdf's, and cdf's in the next two sections.