3 Random variables

3.1 Introduction

Random variables constitute an extension of mathematical variables just like complex variables providing an extension to the real variable system. Random variables are mathematical variables with some probability measures attached to them. Before giving a formal definition to random variables, let us examine some random experiments and some variables associated with such random experiments. Let us take the simple experiment of an unbiased coin being tossed twice.

Example 3.1. Tossing an unbiased coin twice. The sample space is

$$S = \{(H,T),(T,H),(H,H),(T,T)\}.$$

There are four outcomes or four elementary events. Let x be the number of heads in the elementary events or in the outcomes. Then x can take the values $0,1,2$, and thus x is a variable here. But we can attach a probability statement to the values taken by this variable x. The probability that x takes the value zero is the probability of getting two tails and it is $\frac{1}{4}$. The probability that x takes the value 1 is the probability of getting exactly one head, which is $\frac{1}{2}$. The probability that x takes the value 2 is $\frac{1}{4}$. The probability that x takes any another value, other than $0,1,2$, is zero because it is an impossible event in this random experiment. Thus the probability function, associated with this variable x, denoted by $f(x)$, can be written as follows:

$$f(x) = \begin{cases} 0.25, & \text{for } x = 0 \\ 0.50, & \text{for } x = 1 \\ 0.25, & \text{for } x = 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Here, x takes individually distinct values with non-zero probabilities. That is, x here takes the specific value zero with probability $\frac{1}{4}$, the value 1 with probability $\frac{1}{2}$ and the value 2 with probability $\frac{1}{4}$. Such random variables are called discrete random variables. We will give a formal definition after giving a definition for a random variable.

We can also compute the following probability in this case. What is the probability that $x \leq a$ for all real values of a? Let us denote this probability by $F(a)$, that is,

$$F(a) = \Pr\{x \leq a\} = \text{probability of the event } \{x \leq a\}.$$

From Figure 3.1, it may be noted that when a is anywhere from $-\infty$ to 0, not including zero, the probability is zero, and hence $F(a) = 0$. At $x = 0$, there is a probability $\frac{1}{4}$ and this remains the same for all values of a from zero to 1 with zero included but 1 ex-

https://doi.org/10.1515/9783110562545-003

cluded, that is, $0 \le a < 1$. Remember that we are computing the sum of all probabilities up to and including point $x = a$, or we are computing the cumulative probabilities in the notation $\Pr\{x \le a\}$. There is a jump at $x = 1$ equal to $\frac{1}{2}$. Thus when $1 \le a < 2$, then all the probabilities cumulated up to a is $0 + \frac{1}{4} + 0 + \frac{1}{2} + 0 = \frac{3}{4}$. When a is anywhere $2 \le a < \infty$, all the probabilities cumulated up to a will be $0 + \frac{1}{4} + 0 + \frac{1}{2} + 0 + \frac{1}{4} + 0 = 1$. Thus the cumulative probability function here, denoted by $F(a) = \Pr\{x \le a\}$, can be written as follows:

$$F(a) = \begin{cases} 0, & -\infty < a < 0 \\ 0.25, & 0 \le a < 1 \\ 0.75, & 1 \le a < 2 \\ 1, & 2 \le a < \infty. \end{cases}$$

Here, for this variable x, we can associate with x a probability function $f(x)$ and a cumulative probability function $F(a) = \Pr\{x \le a\}$.

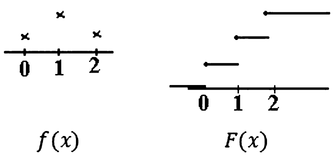

Figure 3.1: Left: Probability function $f(x)$; Right: Cumulative probability function $F(x)$.

Notation 3.1. $\Pr\{c \le x \le d\}$: probability of the event that $c \le x \le d$.

Now let us examine another variable defined over this same sample space. Let y be the number of heads minus the number of tails in the outcomes. Then y will take the value -2 for the sample point (T, T) where the number of heads is zero and the number of tails is 2. The points (H, T) and (T, H) will give a value 0 for y and (H, H) gives a value 2 to y. If $f_y(y)$ denotes the probability function and $F_y(a) = \Pr\{y \le a\}$ the cumulative probability function, then we have the following, which may also be noted from Figure 3.2:

$$f_y(y) = \begin{cases} 0.25, & y = -2 \\ 0.5, & y = 0 \\ 0.25, & y = 2 \\ 0, & \text{elsewhere.} \end{cases}$$

$$F_y(a) = \begin{cases} 0, & -\infty < a < -2 \\ 0.25, & -2 \le a < 0 \\ 0.75, & 0 \le a < 2 \\ 1, & 2 \le a < \infty. \end{cases}$$

Both x and y here are discrete variables in the sense of taking individually distinct values with non-zero probabilities. We may also note one more property that **on a given sample space any number of such variables can be defined**. The above ones, x and y, are only two such variables.

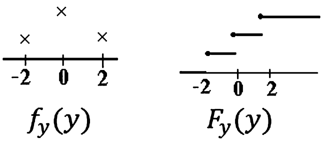

Figure 3.2: Left: Probability function of y; Right: Cumulative probability function of y.

Now, let us consider another example of a variable, which is not discrete. Let us examine the problem of a child playing with scissors and cutting a string of 10 cm into two pieces.

Example 3.2 (Random cut of a string). Let one end of the string be denoted by 0 and the other end by 10 and let the distance from zero to the point of cut be x. Then, of course, x is a variable because we did not know where exactly would be the cut on the string. What is the probability that the cut is anywhere in the interval $2 \leq x \leq 3.5$? In Chapter 1, we have seen that in a situation like this we assign probabilities proportional to the length of the intervals and then

$$\Pr\{2 \leq x \leq 3.5\} = \frac{3.5 - 2.0}{10} = \frac{1.5}{10} = 0.15.$$

What is the probability that the cut is between 2 and 2.001? This is given by

$$\Pr\{2 \leq x \leq 2.001\} = \frac{2.001 - 2.000}{10} = \frac{0.001}{10} = 0.0001.$$

What is the probability that the cut is exactly at 2?

$$\Pr\{x = 2\} = \frac{2 - 2}{10} = 0.$$

Here, x is defined on a continuum of points and the probability that x takes any specific value is zero because here the probabilities are assigned as relative lengths. A point has no length. Such variables, which are defined on continuum of points, will be called continuous random variables. We will give a formal definition after defining a random variable. A probability function which can be associated with this x, denoted by $f_x(x)$, will be of the following form:

$$f_x(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & \text{elsewhere.} \end{cases}$$

Let us see whether we can compute the cumulative probabilities here also. What is the probability that $x \leq a$ for all real values of a? Let us denote this by $F_x(a)$. Then when

$-\infty < a < 0$, the cumulative probability is zero. When $0 \le a < 10$, it is $\frac{a}{10}$, probabilities being relative lengths, and when $10 \le a < \infty$ it is $\frac{10}{10} + 0 = 1$. Thus we have

$$F_x(a) = \begin{cases} 0, & -\infty < a < 0 \\ \frac{a}{10}, & 0 \le a < 10 \\ 1, & 10 \le a < \infty. \end{cases}$$

The probability function in the continuous case is usually called the density function. Some authors do not make a distinction; in both discrete and continuous cases, the probability functions are either called probability functions or density functions. We will use the term "probability function" in the discrete case and mixed cases and "density function" in the continuous case. The density and cumulative density, for the above example, are given in Figure 3.3.

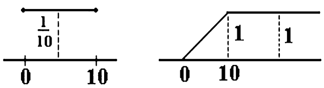

Figure 3.3: Left: Density function of x; Right: Cumulative density function of x.

Here, we may note some interesting properties. The cumulative probability function $F_x(a)$ could have been obtained from the density function by integration. That is,

$$F_x(a) = \int_{-\infty}^{a} f(t)\mathrm{d}t = 0 + \int_0^a \frac{1}{10}\mathrm{d}t = \left[\frac{t}{10}\right]_0^a = \frac{a}{10}.$$

Similarly, the density is available from the cumulative density function by differentiation since here the cumulative density function is differentiable. That is,

$$\left[\frac{\mathrm{d}}{\mathrm{d}a}F_x(a)\right]_{a=x} = \left[\frac{\mathrm{d}}{\mathrm{d}a}\frac{a}{10}\right]_{a=x} = \frac{1}{10} = f_x(x).$$

We have considered two discrete variables associated with the random experiment in Example 3.1 and one continuous random variable in Example 3.2. In all of the three cases, one could have computed the cumulative probabilities, or $\Pr\{x \le a\}$ was defined for all real a, $-\infty < a < \infty$. Such variables will be called random variables. Before giving a formal definition, a few more observations are in order. In the two discrete cases, we had the probability function, which were of the form:

$$f(x_*) = \Pr\{x = x_*\} \tag{3.1}$$

and the cumulative probability function was obtained by adding up the individual probabilities. That is,

$$F(a) = \Pr\{x \le a\} = \sum_{-\infty < x \le a} f(x). \tag{3.2}$$

In Example 3.2, we considered one continuous random variable x where we had the density function

$$f_x(x) = \begin{cases} \frac{1}{10}, & 0 \le x \le 10 \\ 0, & \text{elsewhere,} \end{cases}$$

and the cumulative density function

$$F_x(a) = \Pr\{x \le a\} = \begin{cases} 0, & -\infty < a < 0 \\ \frac{a}{10}, & 0 \le a < 10 \\ 1, & 10 \le a < \infty. \end{cases}$$

$$= \int_{-\infty}^{a} f_x(t)\mathrm{d}t. \tag{3.3}$$

Definition 3.1 (Random variables)**.** Any variable x defined on a sample space S for which the cumulative probabilities $\Pr\{x \le a\}$ can be defined for all real values of a, $-\infty < a < \infty$, is called a real random variable x.

Definition 3.2 (Discrete random variables)**.** Any random variable x which takes individually distinct values with non-zero probabilities is called a discrete random variable and in this case the probability function, denoted by $f_x(x)$, is given by

$$f_x(x_*) = \Pr\{x = x_*\}$$

and obtained by taking successive differences in (3.2).

Definition 3.3 (Continuous random variables)**.** Any random variable x, which is defined on a continuum of points, where the probability that x takes a specific value x_* is zero, is called a continuous random variable and the density function is available from the cumulative density by differentiation, when differentiable, or the cumulative density is available by integration of the density. That is,

$$f_x(x) = \left[\frac{\mathrm{d}}{\mathrm{d}a} F_x(a)\right]_{a=x} \tag{3.4}$$

or

$$F_x(a) = \int_{-\infty}^{a} f_x(t)\mathrm{d}t. \tag{3.5}$$

Definition 3.4 (Distribution function)**.** The cumulative probability/density function of a random variable x is also called the distribution function associated with that random variable x, and it is denoted by $F(x)$:

$$F(a) = [\Pr\{x \le a\}, -\infty < a < \infty]. \tag{3.6}$$

We can also define probability/density function and cumulative function, free of random experiments, by using a few axioms.

3.2 Axioms for probability/density function and distribution functions

Definition 3.5 (Density/Probability function). Any function $f(x)$ satisfying the following two axioms is called the probability/density function of a real random variable x:
(i) $f(x) \geq 0$ for all real x, $-\infty < x < \infty$;
(ii) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$ if x is continuous; and $\sum_{-\infty<x<\infty} f(x) = 1$ if x is discrete.

Example 3.3. Check whether the following can be probability functions for discrete random variables:

$$f_1(x) = \begin{cases} 2/3, & x = -2 \\ 1/3, & x = 5 \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_2(x) = \begin{cases} 3/4, & x = -3 \\ 2/4, & x = 0, \\ -1/4, & x = 2 \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_3(x) = \begin{cases} 3/5, & x = 0 \\ 3/5, & x = 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Solution 3.3. Consider $f_1(x)$. Here, $f_1(x)$ takes the non-zero values $\frac{2}{3}$ and $\frac{1}{3}$ at the points $x = -2$ and $x = 5$, respectively, and x takes all other values with zero probabilities. Condition (i) is satisfied, $f(x) \geq 0$ for all values of x. Condition (ii) is also satisfied because $\frac{2}{3} + \frac{1}{3} + 0 = 1$. Hence $f_1(x)$ here can represent a probability function for a discrete random variable x. We could have also stated $f_1(x)$ as follows:

$$f_1(-2) = \frac{2}{3}; \quad f_1(5) = \frac{1}{3}; \quad f(x) = 0 \quad \text{elsewhere}$$

where, for example, $f_1(-2)$ means $f_1(x)$ at $x = -2$.

$f_2(x)$ is such that $\sum_x f_2(x) = 1$, and thus the second condition is satisfied. But $f_2(x)$ at $x = 2$ or $f_2(2) = -\frac{1}{4}$ which is negative, and hence condition (i) is violated. Hence $f_2(x)$ here cannot be the probability function of any random variable.

$f_3(x)$ is non-negative for all values of x because $f_3(x)$ takes the values $0, \frac{3}{5}, \frac{3}{5}$ but

$$\sum_x f_3(x) = 0 + \frac{3}{5} + \frac{3}{5} = \frac{6}{5} > 1.$$

Here, condition (ii) is violated, and hence $f_3(x)$ cannot be the probability function of any random variable.

Example 3.4. Check whether the following can be density functions of some random variables:

$$f_1(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, b > a \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_2(x) = \begin{cases} cx^4, & 0 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_3(x) = \begin{cases} \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & 0 \le x < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_4(x) = \begin{cases} x, & 0 \le x < 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Solution 3.4. $f_1(x)$ is non-negative since it is either 0 or $\frac{1}{b-a}$ where $b-a>0$. Hence condition (i) is satisfied. Now, check the second condition:

$$\int_{-\infty}^{\infty} f_1(x)\mathrm{d}x = 0 + \int_a^b \frac{1}{b-a}\mathrm{d}x = \left[\frac{x}{b-a}\right]_a^b = \frac{b-a}{b-a} = 1.$$

Hence the second condition is also satisfied. It is a density function of a continuous random variable. The graph is given in Figure 3.4.

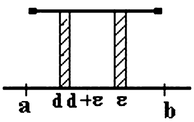

Figure 3.4: Uniform or rectangular density.

This density looks like a rectangle, and hence it is called a *rectangular density*. Since the probabilities are available as integrals or areas under the curve if we take any interval of length ϵ (epsilon) units, say from d to $d+\epsilon$, then the probability that x falls in the interval d to $d+\epsilon$ or $d \le x \le d+\epsilon$ is given by the integral:

$$\int_d^{d+\epsilon} \frac{1}{b-a}\mathrm{d}x = \frac{\epsilon}{b-a}.$$

Since it is a rectangle, if we take an interval of length ϵ anywhere in the interval $a \le x \le b$, then the area will be the same as $\frac{\epsilon}{b-a}$ or we can say that the total area 1 is uniformly distributed over the interval $[a,b]$. In this sense, this density $f_1(x)$ is also called *uniform density*. Also we may observe here that these unknown quantities a and b could be any constants, free of x. As long as $b>a$, $f_1(x)$ is a density.

$f_2(x) \geq 0$ for all values of x if $c > 0$ since either it is zero or x^4 in the interval $[0, 1]$ which is positive. Thus condition (i) is satisfied if $c > 0$. Now, let us check condition (ii):

$$\int_{-\infty}^{\infty} f_2(x)\mathrm{d}x = 0 + \int_0^1 cx^4\mathrm{d}x$$
$$= \left[c\frac{x^5}{5}\right]_0^1 = \frac{c}{5}.$$

Hence condition (ii) is satisfied if $c = 5$. For $c = 5$, $f_2(x)$ is a density function.

$f_3(x)$ satisfies condition (i) when θ (theta) is positive because an exponential function can never be negative. Hence $f_3(x)$ takes zero or a positive value only. Now let us check the second condition:

$$\int_{-\infty}^{\infty} \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}\mathrm{d}x = 0 + \int_0^{\infty} \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}\mathrm{d}x = [-\mathrm{e}^{-\frac{x}{\theta}}]_0^{\infty} = 1.$$

Hence it is a density. Note that whatever be the value of θ as long as it is positive, $f_3(x)$ is a density, see Figure 3.5.

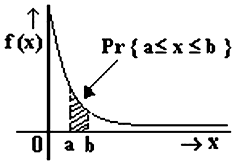

Figure 3.5: Exponential or negative exponential density.

Since this density is associated with an exponential function it is called an *exponential density*. Note that if θ is negative, then $\frac{1}{\theta} < 0$ even though the exponential function remains positive. Thus condition (i) will be violated. If θ is negative, then the exponent $-\frac{x}{\theta} > 0$ thereby the integral from 0 to ∞ will be ∞. Thus condition (ii) will also be violated. For $\theta \leq 0$ here, $f_3(x)$ cannot be a density. When integration is from 0 to ∞, the exponential function with a positive exponent cannot create a density we need not say "negative exponential density" and we simply say that it is an exponential density, and it is implied that the exponent is negative.

$f_4(x)$ is zero or x in $[0, 1)$ and $2 - x$ in $[1, 2]$, and hence $f_4(x) \geq 0$ for all x and condition (i) is satisfied. The total integral is available from the integrals over the several intervals:

$$\int_{-\infty}^{\infty} f_4(x)\mathrm{d}x = 0 + \int_0^1 x\mathrm{d}x + \int_1^2 (2 - x)\mathrm{d}x + 0$$
$$= \left[\frac{x^2}{2}\right]_0^1 + \left[2x - \frac{x^2}{2}\right]_1^2 = \frac{1}{2} + \frac{1}{2} = 1.$$

Thus, condition (ii) is also satisfied and $f_4(x)$ here is a density.

The graph of this density looks like a triangle, and hence it is called a *triangular density* as shown in Figure 3.6.

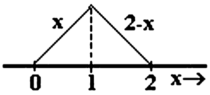

Figure 3.6: Triangular density.

Definition 3.6 (Parameters). Arbitrary constants sitting in a density or probability function are called parameters.

In $f_1(x)$ of Example 3.4, there are two unknown quantities a and b. Irrespective of the values of a and b, as long as $b > a$ then we found that $f_1(x)$ was a density. Hence there are two parameters in that density. In $f_3(x)$ of Example 3.4, we had one unknown quantity θ. As long as θ was positive, $f_3(x)$ remained as a density. Hence there is one parameter here in this density, and that is $\theta > 0$.

Definition 3.7 (Normalizing constant). If a constant sitting in a function is such that for a specific value of this constant the function becomes a density or probability function then that constant is called the normalizing constant.

In $f_2(x)$ of Example 3.4, there was a constant c but for $c = 5$, $f_2(x)$ became a density. This c is the normalizing constant there.

Definition 3.8 (Degenerate random variable). If the whole probability mass is concentrated at one point, then the random variable is called a degenerate random variable or a mathematical variable. Consider the following density/probability function:

$$f(x) = \begin{cases} 1, & x = b \\ 0, & \text{elsewhere.} \end{cases}$$

Here, at $x = b$ the whole probability mass 1 is there and everywhere else the function is zero. The random variable here is called a *degenerate random variable* or with probability 1 the variable x takes the value b or it is a mathematical variable. If there are two points such that at $x = c$ we have probability 0.9999 and at $x = d \neq c$ we have probability 0.0001, then it is not a degenerate random variable even though most of the probability is at one point $x = c$.

Thus, statistics or statistical science is a systematic study of random phenomena and random variables, extending the study of mathematical variables, and as such mathematical variables become special cases of random variables or as degenerate random variables. This author had coined the name “Statistical Science” when he launched the Statistical Science Association of Canada, which became the present Statistical Society of Canada. Thus in this author’s definition, statistical sciences has a wider coverage compared to mathematical sciences. But nowadays the term mathematical sciences is used to cover all aspects of mathematics and statistics.

Example 3.5. Compute the distribution function for the following probability functions:

$$f_1(x) = \begin{cases} 0.3, & x = -2 \\ 0.2, & x = 0, \\ 0.5, & x = 3 \\ 0, & \text{otherwise;} \end{cases}$$

$$f_2(x) = \begin{cases} c(\frac{1}{2})^x, & x = 0, 1, \\ 0, & \text{otherwise.} \end{cases}$$

Solution 3.5. The distribution function in the discrete case is

$$F(a) = \Pr\{x \le a\} = \sum_{-\infty < x \le a} f(x).$$

Hence for $f_1(x)$, it is zero for $-\infty < x < -2$, then there is a jump of 0.3 at $x = -2$, and so on. Therefore,

$$F(a) = \begin{cases} 0, & -\infty < a < -2 \\ 0.3, & -2 \le a < 0 \\ 0.5\,(= 0.3 + 0.2), & 0 \le a < 3 \\ 1, & 3 \le a < \infty. \end{cases}$$

It is a *step function*. In general, for a discrete case we get a step function as the distribution function.

For $f_2(x)$, the normalizing constant c is to be determined to make it a probability function. If it is a probability function, then the total probability is

$$0 + \sum_{x=0}^{2} \left(c\frac{1}{2}\right)^x = 0 + c\left(1 + \frac{1}{2} + \frac{1}{4}\right) = c\frac{7}{4}.$$

Hence for $c = \frac{4}{7}$, $f_2(x)$ is a probability function and it is given by

$$f_2(x) = \begin{cases} 4/7, & x = 0 \\ 2/7, & x = 1 \\ 1/7, & x = 2 \\ 0, & \text{otherwise.} \end{cases}$$

Hence the distribution function is given by

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ 4/7, & 0 \le x < 1 \\ 6/7, & 1 \le x < 2 \\ 1, & 2 \le x < \infty. \end{cases}$$

Again, note that it is a step function. The student may draw the graphs for the distribution function for these two cases.

Example 3.6. Evaluate the distribution function for the following densities:

$$f_1(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & 0 \le x < \infty \\ 0, & \text{otherwise;} \end{cases}$$

$$f_2(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 \le x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

Solution 3.6. The distribution function, by definition, in the continuous case is

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x.$$

Hence in $f_1(x)$,

$$\begin{aligned}\int_{-\infty}^{t} f_1(x)\mathrm{d}x &= 0 + \int_0^t \frac{1}{\theta}e^{-\frac{x}{\theta}}\mathrm{d}x \\ &= [-e^{-\frac{x}{\theta}}]_0^t = 1 - e^{-\frac{t}{\theta}}, \quad 0 \le t < \infty,\end{aligned}$$

and zero from $-\infty < x < 0$. For $f_2(x)$, one has to integrate in different pieces. Evidently, $F(t) = 0$ for $-\infty < t < 0$. When t is in the interval 0 to 1, the function is x and its integral is $\frac{x^2}{2}$. Therefore,

$$\left[\frac{x^2}{2}\right]_0^t = \frac{t^2}{2}.$$

When t is in the interval 1 to 2 the integral up to 1, available from $\frac{t^2}{2}$ at $t = 1$ which is $\frac{1}{2}$, plus the integral of the function $(2 - x)$ from 1 to t is to be computed. That is,

$$\frac{1}{2} + \int_1^t (2 - x)\mathrm{d}x = \frac{1}{2} + \left[2x - \frac{x^2}{2}\right]_1^t = -1 + 2t - \frac{t^2}{2}.$$

When t is above 2, the total integral is one. Hence we have

$$F(t) = \begin{cases} 0, & -\infty < t < 0 \\ \frac{t^2}{2}, & 0 \le t < 1 \\ -1 + 2t - \frac{t^2}{2}, & 1 \le t < 2 \\ 1, & t \ge 2. \end{cases}$$

The student is asked to draw the graphs of the distribution function in these two density functions.

3.2.1 Axioms for a distribution function

If we have a discrete or continuous random variable, the distribution function is $F(t) = \Pr\{x \le t\}$. Without reference to a random variable x, one can define $F(t)$ by using the following axioms:
(i) $F(-\infty) = 0$;
(ii) $F(\infty) = 1$;
(iii) $F(a) \le F(b)$ for all $a < b$;
(iv) $F(t)$ is right continuous.

Thus $F(t)$ is a monotonically non-decreasing (either it increases steadily or it remains steady for some time) function from zero to 1 when t varies from $-\infty$ to ∞. The student may verify that conditions (i) to (iv) above are satisfied by all the distribution functions that we considered so far.

3.2.2 Mixed cases

Sometime we may have a random variable where part of the probability mass is distributed on some individually distinct points (discrete case) but the remaining probability is distributed over a continuum of points (continuous case). Such random variables are called *mixed cases*. We will list one example here, from where it will be clear how to handle such cases.

Example 3.7. Compute the distribution function for the following probability function for a mixed case:

$$f(x) = \begin{cases} \frac{1}{2}, & x = -2 \\ x, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Solution 3.7. The definition for the distribution function remains the same whether the variable is discrete, continuous or mixed:

$$F(t) = \Pr\{x \le t\}.$$

For $-\infty < t < -2$, obviously $F(t) = 0$. There is a jump of $\frac{1}{2}$ at $t = -2$ and then it remains the same until 1. In the interval $[0, 1]$, the function is x and its integral is

$$\int_0^t x\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^t = \frac{t^2}{2}.$$

For t greater than 1, the total probability 1 is attained. Therefore, we have

$$F(t)=\begin{cases}0, & -\infty<t<-2\\ \frac{1}{2}, & -2\le t<0\\ \frac{1}{2}+\frac{t^2}{2}, & 0\le t<1\\ 1, & t\ge 1.\end{cases}$$

The graph will look like that in Figure 3.7.

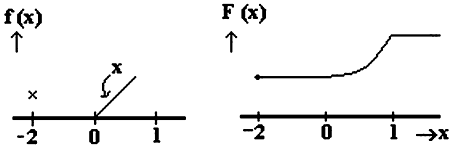

Figure 3.7: The distribution function for a mixed case.

Note that for t up to 0 it is a step function then the remaining part is a continuous curve until 1 and then it remains steady at the final value 1.

Example 3.8. Compute the probabilities (i) $\Pr\{-2\le x\le 1\}$, (ii) $\Pr\{0\le x\le 1.7\}$ for the probability function

$$f(x)=\begin{cases}0.2, & x=-1,\\ 0.3, & x=0,\\ 0.3, & x=1.5,\\ 0.2, & x=2,\\ 0, & \text{otherwise}.\end{cases}$$

Solution 3.8. In the discrete case, the probabilities are added up from those at individual points. When $-2\le x\le 1$, the probabilities in this interval are 0, 0.2 at $x=-1$ and 0.3 at $x=0$. Therefore, the answer to (i) is $0+0.2+0.3=0.5$. When $0\le x\le 1.7$, the probabilities are 0, 0.3 at $x=0$ and 0.3 at $x=1.5$. Hence the answer to (ii) is $0+0.3+0.3=0.6$.

In the discrete case, the probability that x falls in a certain interval is the sum of the probabilities from the corresponding distinct points with non-zero probabilities falling in that interval.

Example 3.9. Compute the following probabilities on the waiting time t, (i) $\Pr\{0\le t\le 2\}$, (ii) $\Pr\{3\le t\le 10\}$ if the waiting time has an exponential density with the parameter $\theta=5$.

Solution 3.9. The waiting time having an exponential density with parameter $\theta=5$ means that the density of t is given by

$$f(t)=\begin{cases}\frac{1}{5}\mathrm{e}^{-\frac{t}{5}}, & 0\le t<\infty\\ 0, & \text{elsewhere}.\end{cases}$$

Probabilities are the areas under the density curve between the corresponding ordinates or the integral of the density over the given interval. Hence for (i) the probability is given by

$$\int_0^2 \frac{1}{5} e^{-\frac{t}{5}} dt = [-e^{-\frac{t}{5}}]_0^2 = 1 - e^{-\frac{2}{5}}.$$

In a similar manner, the probability for (ii) is given by

$$\int_3^{10} f(t) dt = [-e^{-\frac{t}{5}}]_3^{10} = e^{-\frac{3}{5}} - e^{-\frac{10}{5}}.$$

The following shaded areas in Figure 3.8 are the probabilities.

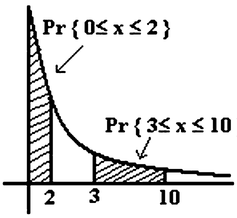

Figure 3.8: Probabilities in the exponential density.

In a continuous case, the probability of the variable x falling in a certain interval $[a, b]$ is the area under the density curve over the interval $[a, b]$ or between the ordinates at $x = a$ and $x = b$.

Exercises 3.2

3.2.1. Check whether the following are probability functions for some discrete random variables:

$$f_1(x) = \begin{cases} \frac{1}{2}, & x = -1 \\ \frac{1}{2}, & x = 1 \\ 0, & \text{elsewhere;} \end{cases} \qquad f_2(x) = \begin{cases} 2, & x = \frac{2}{3} \\ 1, & x = \frac{1}{3} \\ 0, & \text{elsewhere.} \end{cases}$$

$$f_3(x) = \begin{cases} 1.2, & x = 0 \\ -0.2, & x = 1 \\ 0, & \text{elsewhere;} \end{cases} \qquad f_4(x) = \begin{cases} 0.8, & x = 1 \\ 0.3, & x = 2 \\ 0, & \text{otherwise.} \end{cases}$$

3.2.2. Check whether the following are density functions for some continuous random variables:

$$f_1(x) = \begin{cases} c(x^2 + 3x + 1), & 0 \le x \le 2 \\ 0, & \text{otherwise;} \end{cases}$$

$$f_2(x) = \begin{cases} \frac{c}{x^2}, & 1 \le x < \infty \\ 0, & \text{otherwise;} \end{cases}$$

$$f_3(x) = c\mathrm{e}^{-\beta|x|}, \quad -\infty < x < \infty;$$

$$f_4(x) = \begin{cases} cx^2, & 0 < x < 2 \\ 6 - x, & 2 \le x \le 6 \\ 0, & \text{otherwise.} \end{cases}$$

3.2.3. An unbiased coin is tossed several times. If x denotes the number of heads in the outcomes, construct the probability function of x when the coin is tossed (i) once; (ii) two times; (iii) five times.

3.2.4. In a multiple choice examination, there are 8 questions and each question is supplied with 3 possible answers of which one is the correct answer to the question. A student, who does not know any of the correct answers, is answering the questions by picking the answers at random. Let x be the number of correct answers. Construct the probability function of x.

3.2.5. In Exercise 3.2.4, let x be the number of trials (answering the questions) at which the first correct answer is obtained, such as the third ($x = 3$) question answered is the first correct answer. Construct the probability function of x.

3.2.6. In Exercise 3.2.4, let the x-th trial resulted in the 3rd correct answer. Construct the probability function of x.

3.2.7. Compute the distribution function for each probability function in Exercise 3.2.1 and draw the corresponding graphs.

3.2.8. Compute the distribution function for each probability function in Exercise 3.2.2 and draw the corresponding graphs also.

3.2.9. Compute the distribution functions and draw the graphs in Exercises 3.2.3–3.2.6.

3.2.10. For the following mixed case, compute the distribution function:

$$f(x) = \begin{cases} \frac{1}{4}, & x = -5 \\ x, & 0 < x < 1, \\ \frac{1}{4}, & x = 5 \\ 0, & \text{otherwise.} \end{cases}$$

3.2.11. In Exercise 3.2.2, compute the following probabilities: (i) $\Pr\{1 \le x \le 1.5\}$ for $f_1(x)$; (ii) $\Pr\{2 \le x \le 5\}$ for $f_2(x)$; (iii) $\Pr\{-2 \le x \le 2\}$ for $f_3(x)$; (iv) $\Pr\{1.5 \le x \le 3\}$ for $f_4(x)$.

3.2.12. In Exercises 3.2.4 and 3.2.5, compute the probability for $2 \le x \le 5$, and in Exercise 3.2.6 compute the probability for $4 \le x \le 7$.

Note 3.1. For a full discussion of statistical densities and probability functions in common use, we need some standard series such as binomial series, logarithmic series, exponential series, etc. We will mention these briefly here. Those who are familiar with these may skip this section and go directly to the next chapter.

3.3 Some commonly used series

The following power series can be obtained by using the following procedure when the function is differentiable. Let $f(x)$ be differentiable countably infinite number of times and let it admit a power series expansion

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + \cdots$$

then the coefficient

$$a_n = \frac{[\frac{\mathrm{d}^n}{\mathrm{d}x^n}f(x)|_{x=0}]}{n!}$$

or the series is

$$f(x) = f(0) + \frac{f^{(1)}(0)}{1!}x + \frac{f^{(2)}(0)}{2!}x^2 + \cdots \tag{3.7}$$

where $f^{(r)}(0)$ means to differentiate $f(x)$, r times and then evaluate at $x = 0$. All of the following series are derived by using the same procedure.

3.3.1 Exponential series

$$\mathrm{e}^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^r}{r!} + \cdots \quad \text{for all } x. \tag{3.8}$$

$$\mathrm{e}^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \cdots + (-1)^r\frac{x^r}{r!} + \cdots \quad \text{for all } x. \tag{3.9}$$

3.3.2 Logarithmic series

Logarithm to the base e is called the natural logarithms and it is denoted by ln.

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \quad \text{for } |x| < 1. \tag{3.10}$$

For the convergence of the series, we need the condition $|x| < 1$:

$$\ln(1 - x) = -\left[x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots\right], \quad |x| < 1. \tag{3.11}$$

3.3.3 Binomial series

The students are familiar with the binomial expansions for positive integer values, which can also be obtained by direct repeated multiplications, and the general result can be established by the method of induction:

$$(1+x)^2 = 1+2x+x^2; \quad (a+b)^2 = a^2+2ab+b^2;$$
$$(1+x)^3 = 1+3x+3x^2+x^3; \quad (a+b)^3 = a^3+3a^2b+3ab^2+b^3;$$
$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \cdots + \binom{n}{n}x^n, \quad n = 1,2,\ldots; \tag{3.12}$$
$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{n}a^0 b^n,$$
$$n = 1,2,\ldots.$$

What happens if the exponent is not a positive integer, if the exponent is something like $\frac{1}{2}, -20, -\frac{3}{2}$ or some general rational number α (alpha)? We can derive an expansion by using (3.7). Various forms of these are given below:

$$(1-x)^{-\alpha} = 1 + \frac{(\alpha)_1}{1!}x + \frac{(\alpha)_2}{2!}x^2 + \cdots + \frac{(\alpha)_r}{r!}x^r + \cdots, \quad |x| < 1. \tag{3.13}$$

If α is not a negative integer, then we need the condition $|x| < 1$ for the convergence of the series. The Pochhammer symbol is

$$(\alpha)_r = \alpha(\alpha+1)\cdots(\alpha+r-1), \quad \alpha \neq 0, \ (\alpha)_0 = 1. \tag{3.14}$$

Various forms of (3.13) can be obtained by replacing x by $-x$ and α by $-\alpha$. For the sake of completeness, these will be listed here for ready reference:

$$(1+x)^{-\alpha} = [1-(-x)]^{-\alpha} = 1 - \frac{(\alpha)_1}{1!}x + \frac{(\alpha)_2}{2!}x^2 - \cdots, \quad |x| < 1. \tag{3.15}$$
$$(1-x)^{\alpha} = (1-x)^{-(-\alpha)} = 1 + \frac{(-\alpha)_1}{1!}x + \frac{(-\alpha)_2}{2!}x^2 + \cdots,$$
$$\text{for } |x| < 1. \tag{3.16}$$
$$(1+x)^{\alpha} = [1-(-x)]^{-(-\alpha)} = 1 - \frac{(-\alpha)_1}{1!}x + \frac{(-\alpha)_2}{2!}x^2 - \cdots, \tag{3.17}$$

for $|x| < 1$. In all cases, the condition $|x| < 1$ is needed for the convergence of the series except in the case when the exponent is a positive integer. When the exponent is $\alpha > 0$, then the coefficient of $\frac{x^r}{r!}$ is $(-\alpha)_r$. If α is a positive integer, then this Pochhammer symbol will be zero for some r and the series will terminate into a polynomial, and hence the question of convergence does not arise. We have used the form $(1 \pm x)^{\pm\alpha}$. This is general enough because if we have a form

$$(a \pm b)^{\pm\alpha} = a^{\pm\alpha}\left(1 \pm \frac{b}{a}\right)^{\pm\alpha}$$

and thus we can convert to the form $(1 \pm x)$ by taking out a or b to make the resulting series convergent.

3.3.4 Trigonometric series

$$\begin{aligned}
\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \\
\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \\
\mathrm{e}^{ix} &= \cos x + i \sin x, \quad i = \sqrt{-1}.
\end{aligned}$$

3.3.5 A note on logarithms

The mathematical statement

$$a^x = b$$

can be stated as the exponent x is the logarithm of b to the base a. For example, $2^3 = 8$ can be written as 3 (the exponent) is the logarithm of 8 to the base 2. The definition is restricted to b being strictly a positive quantity when real or logarithm of negative quantities or zero is not defined in the real case. The standard notations used are the following:

$\log b \equiv \log_{10} b$ or common logarithm or logarithm to the base 10. When we say "$\log y$", it is a logarithm of y to be base 10.

$\ln b \equiv \log_{\mathrm{e}} b$ or natural logarithm or logarithm to the base e. When we say "$\ln y$", it is a logarithm of y to be base e.

For all other bases, other than 10 or e, write the base and write it as $\log_a b$. This note is given here because the students usually do not know the distinction between the notations "log" and "ln". For example,

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln x = \frac{1}{x}, \quad \frac{\mathrm{d}}{\mathrm{d}x} \log x = \frac{1}{x} \log_{10} \mathrm{e} \neq \frac{1}{x}.$$

Note 3.2. In Section 3.2.1, we have given an axiomatic definition of a distribution function and we defined a random variable with the help of the distribution function. Let us denote the distribution function associated with a random variable x by $F(x)$. If $F(x)$ is differentiable at an arbitrary point x, then let us denote the derivative by $f(x)$. That is, $\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$, which will also indicate that

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

In this situation, we call $F(x)$ an *absolutely continuous* distribution function. Absolute continuity is related to more general measures and integrals known as Lebesgue integrals. For the time being, if you come across the phrase "absolutely continuous distribution function", then assume that $F(x)$ is differentiable and its derivative is the density $f(x)$.

Note 3.3. Suppose that a density function $f(x)$ has non-zero part over the interval $[a,b]$ and zero outside this interval. When x is continuous, then the probability that $x = a$, that is, $\Pr\{x = a\} = 0$ and $\Pr\{x = b\} = 0$. Then the students have the confusion whether $f(x)$ should be written as non-zero in $a \le x \le b$ or $a < x \le b$ or $a \le x < b$ or $a < x < b$. Should we include the boundary points $x = a$ and $x = b$ with the non-zero part of the density or with the zero part? For example, if we write an exponential density:

$$f(x) = \begin{cases} \frac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & \theta > 0,\ 0 \le x < \infty \\ 0, & \text{elsewhere} \end{cases}$$

should we write $0 < x < \infty$ or $0 \le x < \infty$. Note that if we are computing only probabilities then it will not make any difference. But if we are looking for a mode, then the function has a mode at $x = 0$ and if $x = 0$ is not included in the non-zero part of the density, then naturally we cannot evaluate the mode. For estimation of the parameters also, we may have similar problems. For example, if we consider a uniform density

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

then what is known as maximum likelihood estimates [discussed in Module 7] for the parameters a and b do not exist if the end points are not included. That is, if the non-zero part of the density is written as $a < x < b$, then the maximum likelihood estimates for a and b do not exist. Hence when writing the non-zero part of the density include the end points of the interval where the function is non-zero.

Note 3.4. Note that when a random variables x is continuous, then the following probability statements are equivalent:

$$\begin{aligned} \Pr\{a < x < b\} &= \Pr\{a \le x < b\} = \Pr\{a < x \le b\} = \Pr\{a \le x \le b\} \\ &= F(b) - F(a) \end{aligned}$$

where $F(x)$ is the distribution function. Also when $F(x)$ is absolutely continuous

$$F(b) - F(a) = \int_a^b f(t)\mathrm{d}t \quad \text{or} \quad \frac{\mathrm{d}}{\mathrm{d}x}F(x) = f(x)$$

where $f(x)$ is the density function.

Exercises 3.3

3.3.1. By using a binomial expansion show that, for $n = 1, 2, \ldots$

$$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n}$$
$$0 = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} + \cdots \pm \binom{n}{n}$$

3.3.2. By using the identity,

$$(1+x)^m(1+x)^n \equiv (1+x)^{m+n}$$

and comparing the coefficient of x^r on both sides show that

$$\sum_{s=0}^{r}\binom{m}{s}\binom{n}{r-s} = \binom{m+n}{r}, \quad m,n = 1,2,\ldots.$$

3.3.3. By using the identity,

$$(1+x)^{n_1}(1+x)^{n_2}\cdots(1+x)^{n_k} \equiv (1+x)^{n_1+\cdots+n_k}$$

and comparing the coefficient of x^r on both sides show that

$$\sum_{r_1}\cdots\sum_{r_k}\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_k}{r_k} = \binom{n}{r}$$

where $r = r_1 + \cdots + r_k$, $n = n_1 + \cdots + n_k$, $n_j = 1,2,\ldots$, $j = 1,\ldots,k$.

3.3.4. Show that

$$\sum_{m=1}^{n} m = \frac{n(n+1)}{2}; \quad \sum_{m=1}^{n} m^2 = \frac{n(n+1)(2n+1)}{6};$$
$$\sum_{m=1}^{n} m^3 = \left[\frac{n(n+1)}{2}\right]^2.$$

3.3.5. Compute the sums $\sum_{m=1}^{n} m^4$; $\sum_{m=1}^{n} m^5$; $\sum_{m=1}^{n} m^p$, $p = 6,7,\ldots$.

3.3.6. Show that

$$a + ar + ar^2 + \cdots + ar^{n-1} = a\frac{(1-r^n)}{1-r}, \quad r \neq 1;$$
$$a + ar + ar^2 + \cdots = a\sum_{n=0}^{\infty} r^n = \frac{a}{1-r}, \quad \text{for } |r| < 1.$$

3.3.7. What is the infinite sum in Exercise 3.3.6 for (i) $r = 1$; (ii) $r = -1$; (iii) $r > 1$; (iv) $r < -1$.

3.3.8. Evaluate the sum $\sum_{x=k}^{\infty}\binom{x-1}{k-1}p^k q^{x-k}$, $q = 1 - p$, $0 < p < 1$.

3.3.9. Evaluate the sum $\sum_{x=0}^{n}\binom{n}{x}p^x \mathrm{e}^{tx} q^{n-x}$, $q = 1 - p$, $0 < p < 1$.

3.3.10. Compute the sum $\sum_{x=k}^{\infty}\binom{x-1}{k-1}p^k \mathrm{e}^{tx} q^{x-k}$, $q = 1 - p$, $0 < p < 1$.