Random Variables and Probability Distributions

Random Variables

Suppose that to each point of a sample space we assign a number. We then have a *function* defined on the sample space. This function is called a *random variable* (or *stochastic variable*) or more precisely a *random function* (*stochastic function*). It is usually denoted by a capital letter such as X or Y. In general, a random variable has some specified physical, geometrical, or other significance.

Discrete Probability Distributions

Let X be a discrete random variable, and suppose that the possible values that it can assume are given by $x_1, x_2, x_3, \ldots$, arranged in some order. Suppose also that these values are assumed with probabilities given by

$$P(X = x_k) = f(x_k) \qquad k = 1, 2, \ldots \tag{1}$$

It is convenient to introduce the *probability function*, also referred to as *probability distribution*, given by

$$P(X = x) = f(x) \tag{2}$$

For $x = x_k$, this reduces to (1) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability function if

1. $f(x) \geq 0$
2. $\sum_{x} f(x) = 1$

where the sum in 2 is taken over all possible values of x.

Distribution Functions for Discrete Random Variables

The distribution function for a discrete random variable X can be obtained from its probability function by noting that, for all x in $(-\infty, \infty)$,

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) \tag{4}$$

where the sum is taken over all values u taken on by X for which $u \leq x$.

If X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, then the distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases} \tag{5}$$

Example

Suppose the range of a discrete random variable is $\{0, 1, 2, 3, 4\}$. If the probability mass function is $f(x) = cx$ for $x = 0, 1, 2, 3, 4$, what is the value of c?

- First of all, $c \geq 0$ as $f(x) \geq 0$.
- $$\begin{aligned} f(0) + f(1) + f(2) + f(3) + f(4) &= 1 \\ c(0 + 1 + 2 + 3 + 4) = 10c &= 1 \end{aligned}$$

  so we must have $c = \frac{1}{10}$.

Example

Suppose the range of a discrete random variable is $\{0, 1, 2, 3, 4\}$ and its probability mass function is $f(x) = \frac{x}{10}$. What is its cumulative distribution function?

Example

First of all, for any $x < 1$, $F(x) = \sum_{x_i \leq 0} f(x_i) = f(0) = 0$.

Next, for $1 \leq x < 2$, $F(x) = \sum_{x_i \leq 1} f(x_i) = f(0) + f(1) = \frac{1}{10}$

For $2 \leq x < 3$, $F(x) = f(0) + f(1) + f(2) = \frac{3}{10}$
Continuing in the same way, we find that:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{10} & 1 \leq x < 2 \\ \frac{3}{10} & 2 \leq x < 3 \\ \frac{6}{10} & 3 \leq x < 4 \\ 1 & 4 \leq x. \end{cases}$$

Example

A discrete random variable X has the cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10} & 0 \leq x < 1 \\ \frac{3}{10} & 1 \leq x < 2 \\ \frac{5}{10} & 2 \leq x < 4 \\ \frac{8}{10} & 4 \leq x < 5 \\ 1 & 5 \leq x. \end{cases}$$

Determine the probability mass function of X

- The cumulative distribution function only changes value at $0, 1, 2, 4, 5$. So the range of X is $\{0, 1, 2, 4, 5\}$.
- $F(0) = \frac{1}{10}$ so $f(0) = \frac{1}{10}$.
- $\frac{3}{10} = F(1) = f(0) + f(1)$ so $f(1) = \frac{2}{10}$.
- Continuing in the same way we see that the probability mass function is

x	0	1	2	4	5
$f(x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

Continuous Random Variables

A nondiscrete random variable X is said to be *absolutely continuous*, or simply *continuous*, if its distribution function may be represented as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \qquad (-\infty < x < \infty) \tag{7}$$

where the function $f(x)$ has the properties

1. $f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

EXAMPLE 2.5 (a) Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function, and (b) compute $P(1 < X < 2)$.

(a) Since $f(x)$ satisfies Property 1 if $c \geq 0$, it must satisfy Property 2 in order to be a density function. Now

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{3} cx^2\, dx = \left.\frac{cx^3}{3}\right|_0^3 = 9c$$

and since this must equal 1, we have $c = 1/9$.

(b) $$P(1 < X < 2) = \int_1^2 \frac{1}{9} x^2\, dx = \left.\frac{x^3}{27}\right|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

In case $f(x)$ is continuous, which we shall assume unless otherwise stated, the probability that X is equal to any particular value is zero. In such case we can replace either or both of the signs $<$ in (8) by $\leq$. Thus, in Example 2.5,

$$P(1 \leq X \leq 2) = P(1 \leq X < 2) = P(1 < X \leq 2) = P(1 < X < 2) = \frac{7}{27}$$

<u>Another examples:</u>

1. *Let X be a random variable with probability density function*

$$f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) *What is the value of c?*

We know that for $f(x)$ to be a probability distribution $\int_{-\infty}^{\infty} f(x)dx = 1$. We integrate $f(x)$ with respect to x, set the result equal to 1 and solve for c.

$$
\begin{aligned}
1 &= \int_{-1}^{1} c(1-x^2)dx \\
&= \left[cx - c\frac{x^3}{3}\right]_{-1}^{1} \\
&= \left(c - \frac{c}{3}\right) - \left(-c + \frac{c}{3}\right) \\
&= \frac{2c}{3} - \frac{-2c}{3} \\
&= \frac{4c}{3} \\
c &= \frac{3}{4}
\end{aligned}
$$

Thus, $\boxed{c = \frac{3}{4}}$.

(b) *What is the cumulative distribution function of X?*

We want to find $F(x)$. To do that, integrate $f(x)$ from the lower bound of the domain on which $f(x) \neq 0$ to x so we will get an expression in terms of x.

$$
\begin{aligned}
F(x) &= \int_{-1}^{x} c(1-x^2)dx \\
&= \left[cx - \frac{cx^3}{3}\right]_{-1}^{x} \\
& \text{But recall that } c = \tfrac{3}{4}. \\
&= \frac{3}{4}x - \frac{1}{4}x^3 + \frac{1}{2} \\
&= \boxed{\begin{cases} \frac{3}{4}\left(x - \frac{x^3}{3} + \frac{2}{3}\right) & -1 < x < 1 \\ 0 & \text{elsewhere} \end{cases}}
\end{aligned}
$$

4. *The probability density function of X, the lifetime of a certain type of electronic device (measured in hours), is given by,*

$$f(x) = \begin{cases} \frac{10}{x^2} & x > 10 \\ 0 & x \leq 10 \end{cases}$$

(a) *Find $P(X > 20)$.*

There are two ways to solve this problem, and other problems like it. We note that the area we are interested in is bounded below by 20 and unbounded above. Thus,

$$P(X > c) = \int_c^{\infty} f(x)dx$$

Unlike in the discrete case, there is not really an advantage to using the complement, but you can of course do so. We could consider $P(X > c) = 1 - P(X < c)$,

$$P(X > c) = 1 - P(X < c) = 1 - \int_{-\infty}^{c} f(x)dx$$

$$
\begin{aligned}
P(X > 20) &= \int_{20}^{\infty} \frac{10}{x^2} dx \\
&= \left[-\frac{10}{x} \right]_{20}^{\infty} \\
&= \lim_{x \to \infty} \left(-\frac{10}{x} \right) - \left(-\frac{1}{2} \right) \\
&= \boxed{\frac{1}{2}}
\end{aligned}
$$

(b) *What is the cumulative distribution function of X?*

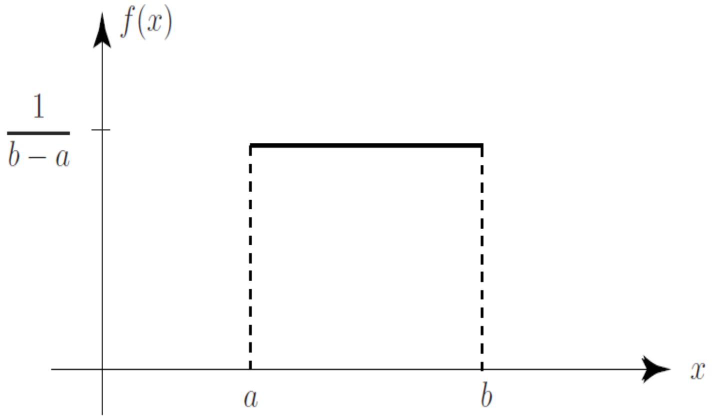
We want to find $F(x)$. To do that, integrate $f(x)$ from the lower bound of the domain on which $f(x) \neq 0$ to x so we will get an expression in terms of x.

$$
\begin{aligned}
F(x) &= \int_{10}^{y} \frac{10}{x^2} dx \\
&= \left[-\frac{10}{x} \right]_{10}^{y} \\
&= -\frac{10}{y} - (-1) \\
&= \boxed{\begin{cases} 1 - \frac{10}{y} & y > 10 \\ 0 & y \leq 10 \end{cases}}
\end{aligned}
$$

The Uniform Distribution

1. The Uniform Distribution

The Uniform or Rectangular distribution has random variable X restricted to a finite interval $[a, b]$ and has $f(x)$ has constant density over the interval. An illustration is



The function $f(x)$ is defined by:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Using the definitions of expectation and variance leads to the following calculations. As you might expect, for a uniform distribution, the calculations are not difficult.
Using the basic definition of expectation we may write:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_a^b x\frac{1}{b-a}\mathrm{d}x = \frac{1}{2(b-a)}\left[x^2\right]_a^b \\ &= \frac{b^2-a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$

Using the formula for the variance, we may write:

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= \int_a^b x^2.\frac{1}{b-a}\mathrm{d}x - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)}\left[x^3\right]_a^b - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{b^3-a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 \\ &= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Key Point

The Uniform random variable X whose density function $f(x)$ is defined by

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

has expectation and variance given by the formulae

$$E(X) = \frac{b+a}{2} \qquad \text{and} \qquad V(X) = \frac{(b-a)^2}{12}$$

Example The current (in mA) measured in a piece of copper wire is known to follow a uniform distribution over the interval $[0, 25]$. Write down the formula for the probability density function $f(x)$ of the random variable X representing the current. Calculate the mean and variance of the distribution and find the cumulative distribution function $F(x)$.

Solution

Over the interval $[0, 25]$ the probability density function $f(x)$ is given by the formula

$$f(x) = \begin{cases} \dfrac{1}{25-0} = 0.04, & 0 \le x \le 25 \\ \\ 0 & \text{otherwise} \end{cases}$$

Using the formulae developed for the mean and variance gives

$$E(X) = \frac{25+0}{2} = 12.5 \text{ mA} \qquad \text{and} \qquad V(X) = \frac{(25-0)^2}{12} = 52.08 \text{ mA}^2$$

The cumulative distribution function is obtained by integrating the probability density function as shown below.

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

Hence, choosing the three distinct regions $x < 0$, $0 \le x \le 25$ and $x > 25$ in turn gives:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{25} & 0 \le x \le 25 \\ 1 & x > 25 \end{cases}$$

Example

Suppose the amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.

What is the probability that daily sales will fall between 2,500 gallons and 3,000 gallons? Answer:

$$P(2500 < X \leq 3000) = \frac{1}{5000 - 2000}(3000 - 2500)$$

$$= 0.1667\,.$$

What is the probability that the service station will sell *at least* 4,000 gallons? Answer:

$$P(X > 4000) = \frac{1}{5000 - 2000}(5000 - 4000)$$

$$= 0.3333\,.$$

What is the probability that the service station will sell *exactly* 2,500 gallons? Answer: $P(X = 2500) = 0$,