Random Variables & Expectation

Random Variables

A **random variable** is a real-valued function whose domain is a sample space.

Random variables are typically denoted by uppercase letters, such as X, Y, and Z. The actual numerical values that a random variable can assume are denoted by lowercase letters, such as x, y, and z.

Mathematically,

$$X : \Omega \longrightarrow \mathbb{R}$$

where Ω represents sample space

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Examples :

Let X denote the random variable that is defined as the sum of two fair dice, then

$P\{X = 2\} = P\{(1, 1)\} = 1/36$

$P\{X = 3\} = P\{(1, 2), (2, 1)\} = 2/36$

$P\{X = 4\} = P\{(1, 3), (2, 2), (3, 1)\} = 3/36$

$P\{X = 5\} = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = 4/36$

$P\{X = 6\} = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = 5/36$

$P\{X = 7\} = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = 6/36$

$P\{X = 8\} = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = 5/36$

$P\{X = 9\} = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = 4/36$

$P\{X = 10\} = P\{(4, 6), (5, 5), (6, 4)\} = 3/36$

$P\{X = 11\} = P\{(5, 6), (6, 5)\} = 2/36$

$P\{X = 12\} = P\{(6, 6)\} = 1/36$

In other words, the random variable X can take on any integral value between 2 and 12.

Note:

$$1 = P(S) = P\left(\bigcup_{i=2}^{12}\{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

Another random variable of possible interest in this experiment is the value of the first die. Letting Y denote this random variable, then Y is equally likely to take on any of the values 1 through 6. That is,

$P\{Y = i\} = 1/6, i = 1, 2, 3, 4, 5, 6.$

Next Example :

Suppose that an individual purchases two electronic components, each of which may be either defective or acceptable. In addition, suppose that the four possible results — (d, d), (d, a), (a, d), (a, a) — have respective probabilities .09, .21, .21, .49 [where (d, d) means that both components are defective, (d, a) that the first component is defective and the second acceptable, and so on]. If we let X denote the number of acceptable components obtained in the purchase, then X is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$P\{X = 0\} = .09$$
$$P\{X = 1\} = .42$$
$$P\{X = 2\} = .49$$

If we were mainly concerned with whether there was at least one acceptable component, we could define the random variable I by

$$I = \begin{cases} 1 & \text{if } X = 1 \text{ or } 2 \\ 0 & \text{if } X = 0 \end{cases}$$

If A denotes the event that at least one acceptable component is obtained, then the random variable I is called the indicator random variable for the event A, since I will equal 1 or 0 depending upon whether A occurs. The probabilities attached to the possible values of I are

$$P\{I = 1\} = .91$$
$$P\{I = 0\} = .09$$

Distribution of a randam variable X

Collection of all the probabilities related to X is the distribution of X.

Let X is the number of 1's in a random binary string of 3 characters then the distribution of X is

x	$P\{X = x\}$
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

Because

$$\Omega = \{000, 001, 010, 100, 111, 110, 101, 011\}$$

Discrete and Continuous Random Variables

In the two foregoing examples, the random variables of interest took on a finite number of possible values. Random variables whose set of possible values can be written either as a finite sequence $x_1, \ldots, x_n$, or as an infinite sequence $x_1, \ldots$ are said to be *discrete*. For instance, a random variable whose set of possible values is the set of nonnegative integers is a discrete random variable. However, there also exist random variables that take on a continuum of possible values i.e., a random variable that can take on any real value in an interval (possibly even the entire real line). These are known as *continuous* random variables. In other words a random variable X is said to be continuous if it can take on the infinite number of possible values associated with intervals of real numbers. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The Cumulative Distribution Function (CDF)

Note: CDF is defined for both Discrete & Continuous r.v.s.

The cumulative distribution function, or more simply the distribution function, F of the random variable X is defined for any real number x by

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

That is, F (x) is the probability that the random variable X takes on a value that is less than or equal to x.

***Notation*:** We will use the notation $X \sim F$ to signify that F is the distribution function of X .

Suppose we wanted to compute $P\{a < X \leq b\}$. This can be accomplished by first noting that the event $\{X \leq b\}$ can be expressed as the union of the two mutually exclusive events $\{X \leq a\}$ and $\{a < X \leq b\}$. Therefore, applying Axiom 3, we obtain that

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

$$\longrightarrow \quad P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\}$$

or

$$P\{a < X \leq b\} = F(b) - F(a)$$

Problem

Suppose the random variable X has distribution function

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\{-x^2\} & x > 0 \end{cases}$$

What is the probability that X exceeds 1?

Solution The desired probability is computed as follows:

$$\begin{aligned} P\{X > 1\} &= 1 - P\{X \leq 1\} \\ &= 1 - F(1) \\ &= e^{-1} \end{aligned}$$

$$\rightarrow \quad P\{X > 1\} = .368$$

Probability Mass Function (PMF)

Note: Only discrete r.v.s have PMF

As was previously mentioned, a random variable whose set of possible values is a sequence is said to be discrete. For a discrete random variable X , we define the probability mass function $p(a)$ of X by

$$p(a) = P\{X = a\}$$

If X assume one of the values $x_1, x_2, \ldots,$ then

$$p(x_i) > 0, \qquad i = 1, 2, \ldots$$

and $$p(x) = 0, \qquad \text{all other values of } x$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Problem

Consider a random variable X that is equal to 1, 2, or 3. If we know that

$$p(1) = \frac{1}{2} \qquad \text{and} \qquad p(2) = \frac{1}{3}$$

Compute $p(3)$.

Solution

$$\text{since } p(1) + p(2) + p(3) = 1$$

$$\Rightarrow p(3) = \frac{1}{6}$$

A graph of $p(x)$ is presented in Figure 4.1.

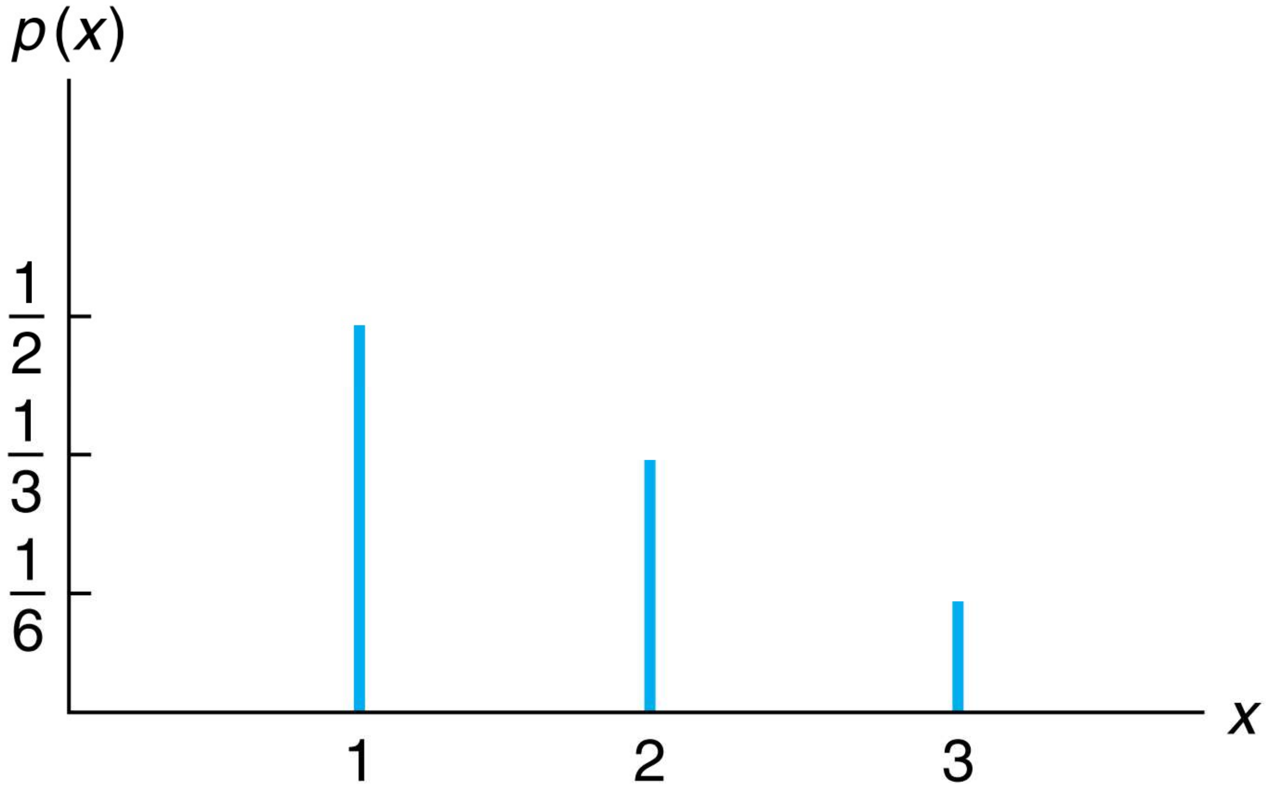

FIGURE 4.1 *Graph of* $p(x)$

Relation of CDF with PMF

The cumulative distribution function F can be expressed in terms of $p(x)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$

If X is a discrete random variable whose set of possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then its distribution function F is a step function. That is, the value of F is constant in the intervals $[x_{i-1}, x_i)$ and then takes a step (or jump) of size $p(x_i)$ at x_i.

For instance, suppose X has a probability mass function given by

$$p(1) = \tfrac{1}{2}, \qquad p(2) = \tfrac{1}{3}, \qquad p(3) = \tfrac{1}{6}$$

Then the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \leq a < 2 \\ \frac{5}{6} & 2 \leq a < 3 \\ 1 & 3 \leq a \end{cases}$$

This is graphically presented in Figure 4.2.

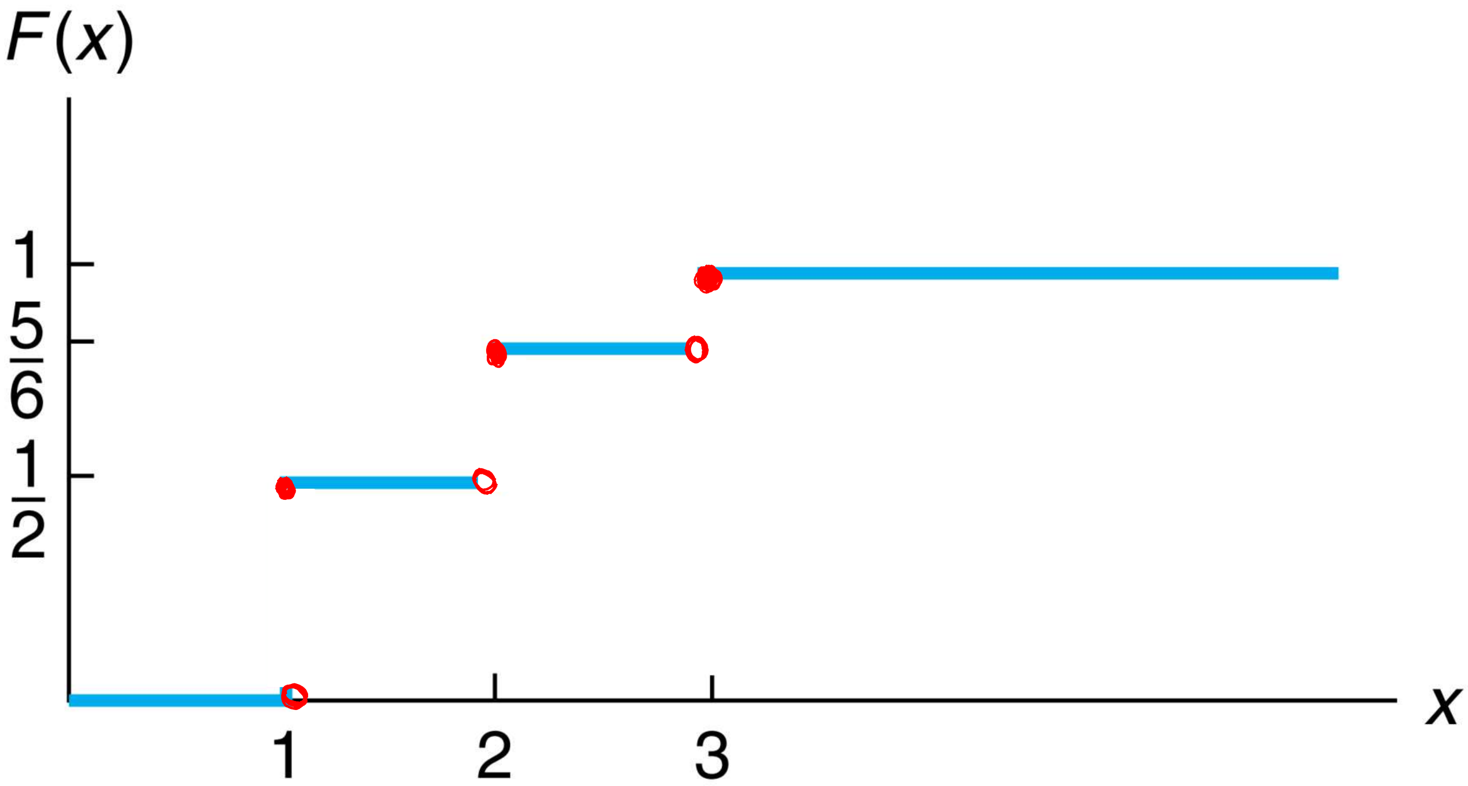

FIGURE 4.2 *Graph of $F(x)$*

Note:

If we are given a CDF and it's a step function, then it is the CDF of a discrete random variable.

(A step function is that one which is made up of disconnected constant functions. This type of function is also referred to as a stair function because its graph is like stairs.)

Method of finding PMF from a CDF of a discrete random variable

We can get an idea about the possible values of the random variable from the CDF by looking to the points where the jumps occurs in the graph of CDF.

If $x_1, x_2, \ldots$ are possible values of the random variable then

$$F(x_1) = P\{X \leq x_1\} = P\{X = x_1\} = p(x_1)$$

$$\Rightarrow F(x_1) = p(x_1) \text{ or } p(x_1) = F(x_1)$$

Next

$$F(x_2) = P\{X \leq x_2\} = P\{X \leq x_1\} + P\{X = x_2\}$$

$$\Rightarrow F(x_2) = F(x_1) + p(x_2) \Rightarrow p(x_2) = F(x_2) - F(x_1)$$

Similarly

$$p(x_3) = F(x_3) - F(x_2) \text{ and so on.}$$

Practice Problem

Find the PMF of the random variable whose CDF is given by

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{2} & 1 \leq a < 2 \\ \frac{5}{6} & 2 \leq a < 3 \\ 1 & 3 \leq a \end{cases}$$

Probability Density Function (PDF)

Note: Only continuous r.v.s have PDF

We say that a random variable X has a *continuous distribution* or that X is a *continuous random variable* if there exists a nonnegative function f , defined on the real line, such that for every interval of real numbers (bounded or unbounded), the probability that X takes a value in the interval is the integral of f over the interval.

$$P\{a \leq X \leq b\} = \int_a^b f(x)\, dx \quad \bigstar$$

The function *f(x)* is called the *probability density function* of the random variable *X.*

Note that a probability density function $f(x)$ must satisfy the following

$$f(x) \geq 0 \text{ for all } x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

If we let $a = b$ in ★ then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any particular value is zero. The fact that $P(X = a) = 0$ does not imply that $X = a$ is impossible. If it did, all values of X would be impossible and X couldn't assume any value. What happens is that the probability in the distribution of X is spread so thinly that we can only see it on sets like nondegenerate intervals.

As an immediate consequence of the fact that in the continuous case probabilities associated with individual points are always zero, we find that if we speak of the probability associated with the interval from a to b, it does not matter whether either endpoint is included. Symbolically,

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b)$$

Problem

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

Solution

(a) Since f is a probability density function, we must have that $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that

$$C\int_0^2 (4x - 2x^2)\,dx = 1$$

$$\Rightarrow \quad C\left[2x^2 - \frac{2x^3}{3}\right]\Bigg|_{x=0}^{x=2} = 1$$

or

$$C = \frac{3}{8}$$

(b)

$$P\{X > 1\} = \int_1^{\infty} f(x)\,dx = \frac{3}{8}\int_1^{2} (4x - 2x^2)\,dx = \frac{1}{2}$$

Problem

The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down?
(b) it will function for fewer than 100 hours?

Solution (a) Since

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_{0}^{\infty} e^{-x/100}\,dx$$

we obtain

$$1 = -\lambda(100)e^{-x/100}\Big|_0^\infty = 100\lambda \quad \text{or} \quad \lambda = \frac{1}{100}$$

Hence, the probability that a computer will function between 50 and 150 hours before breaking down is given by

$$P\{50 < X < 150\} = \int_{50}^{150} \frac{1}{100} e^{-x/100}\, dx = -e^{-x/100}\Big|_{50}^{150}$$
$$= e^{-1/2} - e^{-3/2} \approx .383$$

(b) Similarly,

$$P\{X < 100\} = \int_{0}^{100} \frac{1}{100} e^{-x/100}\, dx = -e^{-x/100}\Big|_{0}^{100} = 1 - e^{-1} \approx .632$$

In other words, approximately 63.2 percent of the time, a computer will fail before registering 100 hours of use. ■

Relation b/w CDF and PDF

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)\, dx$$

Differentiating both sides yields

$$\frac{d}{da}F(a) = f(a) \qquad \text{(By Fundamental Theorem of Calculus)}$$

Problem

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Solution

For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\ dt = \int_{-1}^{x} \frac{t^2}{3} dt = \frac{t^3}{9}\bigg|_{-1}^{x} = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

Practice Problems

1.

The distribution function of the random variable X is given

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{2} & 0 \leq x < 1 \\ \dfrac{2}{3} & 1 \leq x < 2 \\ \dfrac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

(a) Plot this distribution function.
(b) What is $P\{X > \frac{1}{2}\}$?
(c) What is $P\{2 < X \leq 4\}$?

2.

Five men and 5 women are ranked according to their scores on an examination. Assume that no two scores are alike and all 10! possible rankings are equally likely. Let X denote the highest ranking achieved by a woman (for instance, $X = 2$ if the top-ranked person was male and the next-ranked person was female). Find $P\{X = i\}, i = 1, 2, 3, \ldots, 8, 9, 10$.

3. If a random variable has the probability density

$$f(x) = \begin{cases} 2\,e^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

find the probabilities that it will take on a value

(a) between 1 and 3;

(b) greater than 0.5.

4.

Find k so that the following can serve as the probability density of a random variable:

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ kxe^{-4x^2} & \text{for } x > 0 \end{cases}$$

5. The length of satisfactory service (years) provided by a certain model of laptop computer is a random variable having the probability density

$$f(x) = \begin{cases} \dfrac{1}{4.5}\, e^{-x/4.5} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

Find the probabilities that one of these laptops will provide satisfactory service for

(a) at most 2.5 years;

(b) anywhere from 4 to 6 years;

(c) at least 6.75 years.

6.

Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 \le x < 4, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Evaluate k.

(b) Find $F(x)$ and use it to evaluate $P(3 < X < 4)$.