CONTINUOUS RANDOM VARIABLES
AND PROBABILITY DISTRIBUTIONS

TABLE OF CONTENTS

Content Developer

Chandra Goswami, Associate Professor, Department of Economics
Dyal Singh College, University of Delhi

Reference

Jay L. Devore: *Probability and Statistics for Engineering and the Sciences*, Cengage Learning, 8th edition [Chapter 4]

CONTINUOUS RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

***Learning objectives*:**

In this chapter you will learn what is meant by a continuous random variable. You will learn how to arrive at the probability distribution of such types of random variables and how to represent these graphically, as well as presentation by summary expressions. You will then learn how to derive cumulative distribution functions from the probability distribution function. You will also be able to derive the probability densities from the cumulative distribution function. If either the probability density function or the cumulative distribution function is known then you will be able to evaluate the probability that the random variable takes on specific values or a range of values. You will also learn how to identify the characteristics of the population distribution like the shape of the distribution.

Chapter Outline

1 CONTINUOUS RANDOM VARIABLES

A random variable is said to be continuous when the outcome of a random experiment can be any real number in a given interval and the number of possibilities is uncountably infinite. The outcomes of experiments are denoted by points on a line or on line segments of the measurement axis.

Example 1.1

Students of a college are given an objective type test. The proportion of correct answers that a student scores in the test is a continuous variable which can range from 0 to 1. Measured as a percentage, the outcome varies from 0 to 100 percent.

Example 1.2

A student travels to college by metro. The frequency of trains in the morning is 4 minutes. If the student reaches the platform as one train is departing she will have to wait for 4 minutes till the next train enters the station. If she reaches just as one train enters the station then she will have to wait 0 minutes to board the train. If she reaches after the earlier train has left and the next train is yet to arrive, she will have to wait for a time period between 0 and 4 minutes. Waiting time is a continuous variable with a minimum of 0 minutes and a maximum of 4 minutes.

Example 1.3

The daily consumption of water (in liters) by an individual at home varies from day to day through any given year. It depends on various factors like amount of time spent at home, weather conditions, time of year, how much of the time spent at home is during waking hours, etc. The unit of measurement is a continuous variable with a minimum value of 0 liters.

***Definition* 1**

A ***random variable is continuous*** if both the following conditions apply

1. Its set of possible values consists either of all numbers in an interval on the number line or all numbers in a disjoint union of such intervals.
2. No possible value of the random variable has a positive probability.

Condition 1 implies that there is no way to create a listing of all the infinite number of possible values of the variable. Condition 2 implies that intervals of values have positive probability. As the width of the interval diminishes, probability of the interval decreases. In the limit, probability of the interval is zero as the width of the interval reduces to zero.

Example 1.4

The university team is scheduled to visit any minute during a three hour long examination starting at 9am. We may want to find the probability that the team visits at a given time or we may be interested in the probability that the visit takes place during a given time interval. The sample space is from 0 to 180 minutes. The probability that the team visits during an interval of length *c* is $\frac{c}{180}$. This assignment of probabilities applies only to intervals on the measurement axis from 0 to 180. The probability decreases as the interval becomes shorter. For an interval of 5 seconds, the probability is computed as $\frac{5}{10800} = 0.0004629$ As the length of the interval approaches zero, the probability that the team will visit also approaches zero. That is why we always assign zero probability for a single point on the number line. This does not mean that the team will not visit. The team will visit at some point in the interval from 0 to 180 minutes even though each point has zero probability.

Variables such as time, height, distance, temperature, area, volume, weight, etc that require measurement are continuous. In practice, however, limitations of measurement instruments often do not allow measurement on a continuous scale. Yet we study models of continuous variables as they often reflect real world situations.

2 PROBABILITY DISTRIBUTIONS FOR CONTINUOUS RANDOM VARIABLES

Whereas the set of possible values of a discrete rv is a sequence, the set of possible values for a continuous rv is an interval. The continuous rv X can take any one of the infinite number of possible values in that interval. In this case random variables can take on values on a continuous scale.

To derive the probability distribution for a continuous rv let us first begin with a discrete rv. Let X be a discrete rv which can take integer values such that $x_1 \leq X \leq x_n$, where x_1 and x_n are the minimum and maximum values respectively of the rv X.
If $x = x_1, x_2, \ldots, x_n$ then we can draw a probability histogram with n rectangles. The area of the rectangle centered at x_j is the proportion of the population that has the value x_j, ie, ${f_j}/{N}$ where N is the population size. Summing over the n values of X we obtain $\sum_{i=1}^{n} \frac{f_i}{N} = 1$

Now we allow X to take one additional value in each interval so that x_1' is midway between x_1 and x_2; x_2' is midway between x_2 and x_3; and so on. Then total number of x values will be 2n – 1 (instead of 2n, as there are n - 1 intervals). With measurements of x taken at smaller intervals, the rectangles become narrower, though the sum of the areas of all rectangles remains one.

If we continue this process of measuring x at smaller and smaller intervals, the resulting sequence of probability histograms, of the distributions of corresponding discrete random variables, will approach a smooth curve. Figure 1 illustrates this process in the three panels 1.1, 1.2 and 1.3

Figure 1 Deriving histogram of a continuous random variable

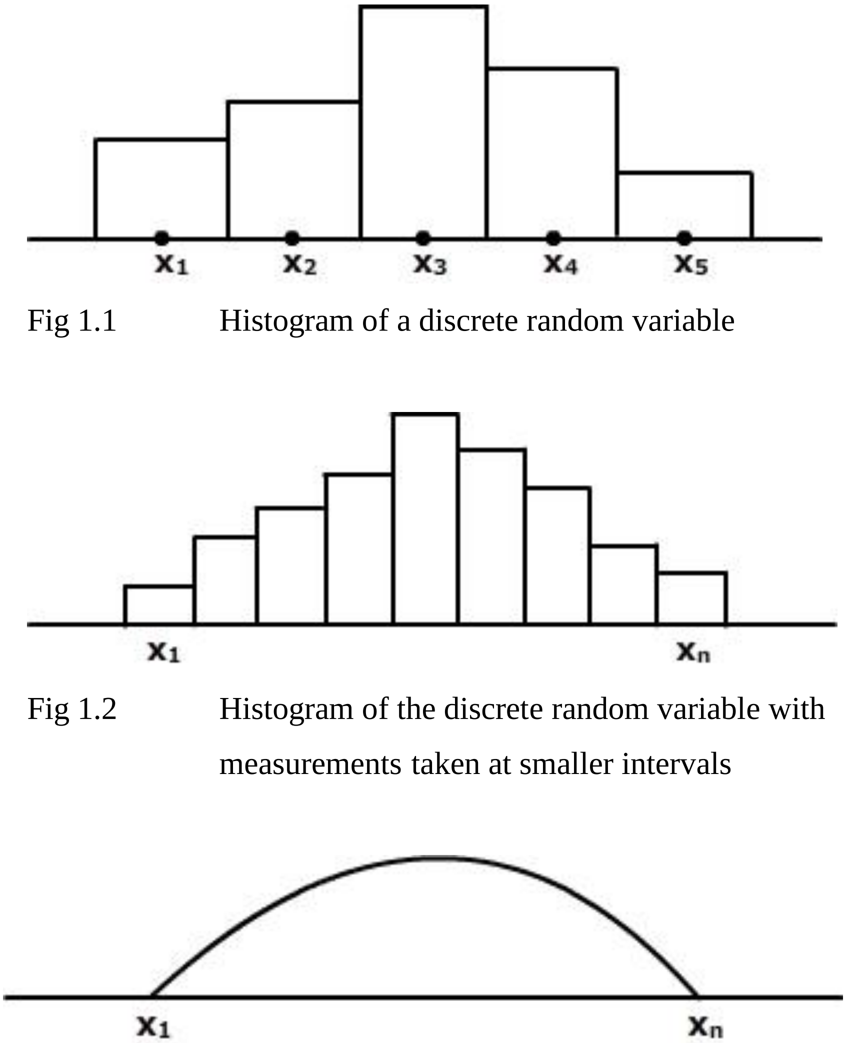


Fig 1.1 Histogram of a discrete random variable

Fig 1.2 Histogram of the discrete random variable with measurements taken at smaller intervals

Fig 1.3 Limit of a sequence of discrete histograms

Since for each histogram the total area of all rectangles equals one, the total area under the continuous curve is also one. The smooth curve represents a continuous probability distribution. The sum of the areas of the rectangles that represent the probability that *X* falls within any specified interval [a, b] approaches the corresponding area under the curve for the interval from *a* to *b*.

***Definition* 2**

Let X be a continuous random variable. Then a probability distribution or ***probability density function*** (pdf) of *X* is a function *f(x)* such that for any two numbers *a* and *b* with a $\leq$ b, P(a $\leq$ X $\leq$ b) = $\int_a^b f(x)\,dx$

Probability density functions are also referred to as density functions.

The probability that *X* takes on a value in the interval [a, b] is the area under the graph of the density function above the interval [a, b] on the number line.

The following two conditions must be satisfied by f(x) to be a pdf:

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

The first condition requires non-negative values of pdf for any x value. The second condition requires that area under the entire curve of f(x) should equal one, ie *X* values are collectively exhaustive. If all possible values of *X* are considered then the second condition will be satisfied. Examples of pdf are the continuous Uniform Distribution, the Normal Distribution, the Exponential Distribution, etc.

Unlike the pmf, where we can obtain P(X = c) as the probability that the discrete rv X takes the value *c*, the probabilities for a continuous rv are always associated with intervals. The pdf yields P(X = c) = 0 for any particular value of the rv *X*. This follows from the definition of a continuous rv as specified in *condition* 2 of *definition* 1.

For the discrete rv *X*, each possible value of *X* is assigned a positive probability. In case of the continuous rv *X*, area under the density curve that lies above any single value of *X* is zero. We have:

P(X=c) = p(c) = $\int_{c}^{c} f(x)\,dx = \lim_{\varepsilon\to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)\,dx = 0$

In view of this property, it does not matter if we include or we exclude the endpoints of the interval from *a* to *b*. Thus, for the continuous rv *X*, if $a \le b$,
$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$.

This is not the case with discrete random variables. If both *a* and *b* are possible values of the discrete rv *X* then these probabilities will all be different. If a < b, then for the discrete rv X,
$P(a \le X \le b) \ne P(a < X \le b) \ne P(a \le X < b) \ne P(a < X < b)$.

Example 2.1

A milk vendor has a refrigerated storage tank of 1000 liters capacity, which is filled each morning for sale during the day. It is not possible to predict the amount of milk sold on any particular day. The sale of milk on any day can vary from 0 lt. to 1000 lt. Past experience shows that any demand in the interval of 0 and 1000 is equally likely. The rv *X* indicates the sale of milk on a particular day. The pdf of *X* is given by the continuous Uniform Distribution

$$f(x) = \begin{cases} 0.001 & 0 \le x \le 1000 \\ 0 & otherwise \end{cases}$$

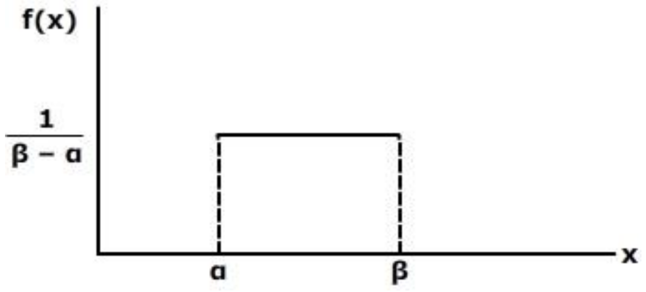
In general, if α and β are the lower and upper limits of the value that the continuous rv *X* can take, then pdf of X is

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta - \alpha} & 0 \le x \le 1000 \\ 0 & otherwise \end{cases}$$

The probability of an interval depends only on the width of the interval in case of the uniform distribution.

The pdf of the uniform distribution is illustrated in Figure 2.

Figure 2 Graph of the continuous Uniform Distribution



In our example, β – α = 1000 so that $\frac{1}{\beta - \alpha}$ = 0.001. We can use this to obtain the probability that sale of milk on a particular day is between 200 and 500 liters as follows:

$P(200 \leq X \leq 500) = (500 - 200)(0.001) = 0.3$

Note that α and β are the parameters of a population of the continuous rv X that is described by a uniform distribution. We have a family of uniform distributions for different values of the two parameters. Each distribution is specified by a particular pair of values of α and β.

Exercise 1

Show that $f(x) = 3x^2$ for $0 < x < 1$ represents a pdf and calculate $P(0.1 < x < 0.5)$.

Solution

f(x) can represent a pdf if both conditions for a pdf are satisfied, ie, $f(x) \geq 0$ and

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

Since $f(x) = 3x^2$ and $x^2 > 0$ always, hence $f(x) \geq 0$ for all x values. Therefore, for $0 < x < 1$, $f(x) \geq 0$ and the first condition is satisfied.

$$\int_0^1 3x^2\,dx = \left.\frac{3x^3}{3}\right|_0^1 = 1 - 0 = 1$$, which satisfies the second condition for pdf.

Since both conditions are satisfied, $f(x) = 3x^2$ represents a pdf for $0 < x < 1$

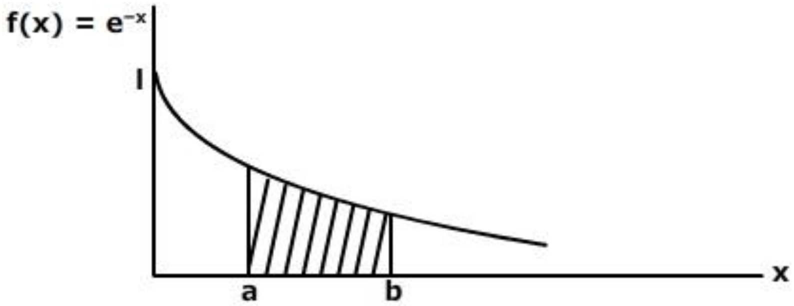
Now, $P(0.1 < x < 0.5) = \int_{0.1}^{0.5} 3x^2\,dx = (0.5)^3 - (0.1)^3 = 0.125 - 0.001 = 0.124$

Example 2.2

The pdf for a continuous rv is given as f(x) = $\begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

So that as x value increases from x = 0, f(x) decreases rapidly or exponentially, as illustrated in Fig 3

Figure 3 pdf of f(x) = e^{-x} for x $\geq$ 0



Now, P(a $\leq$ X $\leq$ b) = $\int_a^b e^{-x}$ dx. This is the shaded area in figure 3.

If a = 2 and b = 5, then

P(2 $\leq$ X $\leq$ 5) = $\int_2^5 e^{-x}$ dx = $\left. -e^{-x} \right|_2^5$ = - (0.006738 – 0.135335) = 0.128597 = 0.13

Therefore, 13 percent of the area under the curve of f(x) = e^{-x} lies above the measurement axis in the interval [2, 5].

Exercise 2

Show that f(x) = e^{-x} for 0 < x < ∞ represents a pdf, and compute the probability that X > 1.

Solution

f(x) = e^{-x} would represent a pdf if f(x) $\geq$ 0 and $\int_0^\infty f(x)\,dx = 1$ for 0 < x < ∞

Since e > 0, for all positive x values $e^{-x} > 0$.

f(x) = 1 for x = 0. If x> 0, f(x) < 1. As x → ∞, f(x) → 0

$\int_0^\infty f(x)\,dx$ = $\int_0^\infty e^{-x}\,dx$ = $\left. -e^{-x} \right|_0^\infty$ = [0 – 1] = 1.

Thus both conditions are satisfied and f(x) is a pdf.

P(X > 1) = $\int_1^\infty e^{-x}\,dx$ = - [0 – e^{-1}] = e^{-1} = 0.368

Exercise 3

The pdf of the rv X is given by

$$f(x) = \begin{cases} \dfrac{k}{\sqrt{x}} & 0 < x < 4 \\ 0 & otherwise \end{cases}$$

Find (a) the value of *k*, and (b) P(X > 1)

Solution

(a) Given that f(x) is a pdf we have $\int_0^4 \frac{k}{\sqrt{x}}\,dx = 1$

Now $\int_0^4 \frac{4}{\sqrt{x}}\,dx = \left.\frac{k\sqrt{x}}{1/2}\right|_0^4$ = 2k [2 – 0] = 4k. Equating 4k and 1 we get k = $\frac{1}{4}$ so that

f(x) = $\frac{1}{4\sqrt{x}}$

(b) P(X > 1) = $\int_1^4 \frac{1}{4\sqrt{x}}\,dx = \left.\frac{2\sqrt{x}}{4}\right|_1^4$ = 1 - $\frac{1}{2}$ = $\frac{1}{2}$ = 0.5

Exercise 4

If the continuous random variable *X* can take only non-negative values and has the density function f(x) = e^{2x} for x $\geq$ 0, and 0 otherwise, what is the maximum value of *X*?

Solution

If f(x) is a density function then $\int_0^\infty e^{2x} dx = 1$ for x $\geq$ 0, and 0 otherwise.

$\int_0^x e^{2y} dy = \left.\frac{e^{2y}}{2}\right|_0^x = \frac{e^{2x}}{2} - \frac{1}{2} = 1 \Rightarrow$ e^{2x} = 3

Therefore, 2x = ln 3 = 1.0986, so that x = 0.549

Hence, f(x) will be a density function for 0 $\leq$ x $\leq$ 0.549. Maximum value of X is 0.549

3 CUMULATIVE DISTRIBUTION FUNCTIONS FOR CONTINUOUS RANDOM VARIABLES

Similar to the case of discrete random variables, there are many problems where we need to know the probability that a continuous rv *X* takes a value that does not exceed a specified value *x*. For this we need the cumulative distribution function (cdf) of *X*.

***Definition* 3**

If X is a continuous random variable then ***the cumulative distribution function*** F(x) for X is defined for every number *x* by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\ dx$$

For each x, F(x) is the area under the density curve to the left of x. As x value increases, F(x) also increases smoothly until F(x) =1 and then it continues as a flat line parallel to the measurement axis.

The cdf gives the probability P(X $\leq$ x) obtained by integrating the pdf f(y) between -∞ and *x.* As in the case of the discrete rv, here too F(- ∞) = 0, F(∞) = 1, and
F(a) $\leq$ F(b) when a < b.
Also P(a $\leq$ X < b) = F(b) – F(a) where a $\leq$ b.
Since *X* is a continuous rv,
P(a $\leq$ X < b) = P(a < X < b) = P(a < X $\leq$ b) = F(b) – F(a) where a $\leq$ b.

Example 3.1

Given the uniform distribution f(x; A, B) = $\begin{cases} \dfrac{1}{B-A} & A \leq x \leq B \\ 0 & otherwise \end{cases}$,

the cdf, $F(x) = \int_{-\infty}^{x} f(y)\ dy$

Since minimum value of the rv is *A*, we have

$$F(x) = \int_{A}^{x} \frac{1}{B-A}\ dy \quad = \quad \frac{1}{B-A}\, y \Bigg|_{y=A}^{y=x} \quad = \frac{x-A}{B-A}$$

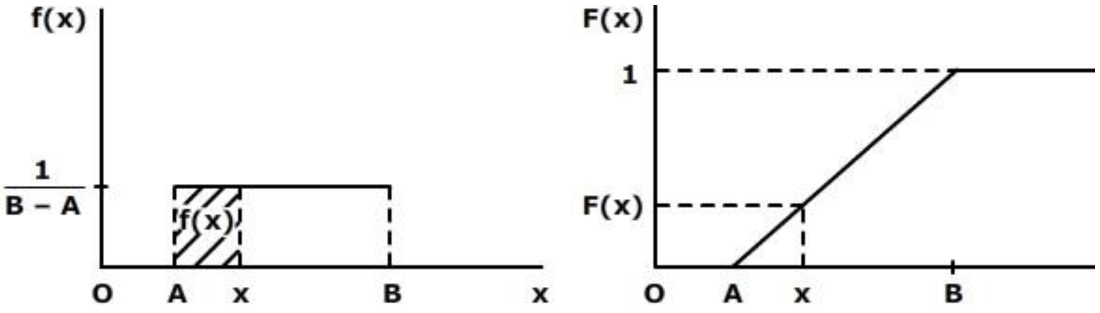
Since $\int_A^B \frac{1}{B-A}\, dx \quad = \quad 1$ therefore F(x) = 0 for x < A and F(x) = 1 for x ≥ B.

Hence, the cdf of the uniform distribution is

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \le x < B \\ 1 & x \ge B \end{cases}$$

The pdf and cdf of the uniform distribution of a continuous rv are illustrated in Fig 4.

Figure 4 pdf & cdf of a uniform distribution

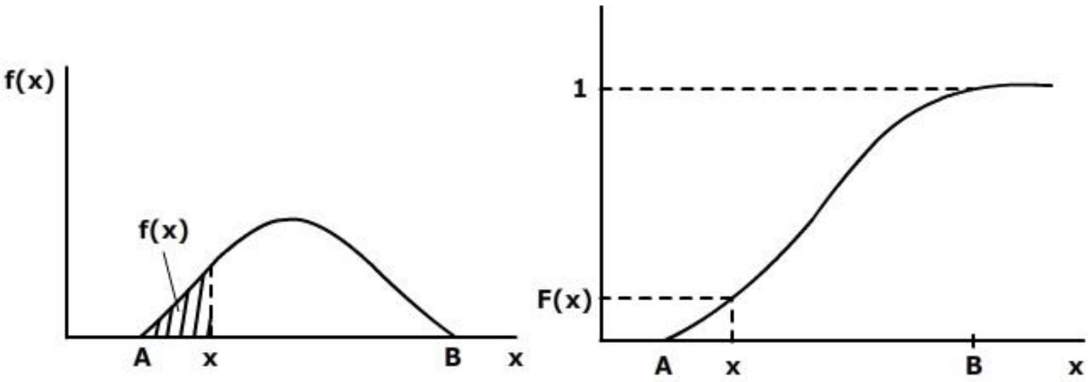


(a) pdf

(b) cdf

If the graph of the pdf is bell-shaped as in case of the Normal Distribution [fig 5 (a)], then the cdf will be as in Figure 5 (b)

Figure 5 pdf & cdf of normal distribution



(a) pdf

(b) cdf

Exercise 5

The density function of the rv *X* is given by

$$f(x) = \begin{cases} 6x(1-x) & 0<x<1 \\ 0 & otherwise \end{cases}$$

Obtain the cdf and compute P(X < ½).

Solution

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \quad = \quad \int_{0}^{x} 6y(1-y)\,dy \quad = \quad 6\int_{0}^{x} (y-y^2)\,dy \quad = \quad 6\left[\frac{y^2}{2}-\frac{y^3}{3}\right]_{y=0}^{y=x}$$

If x ≤ 0, F(x) = 0

If 0 < x < 1, F(x) = 3x^2 – 2x^3

If x = 1, F(x) = 3 – 2 = 1

If x > 1 F(x) = 1 since f(x) = 0

Therefore the cdf can be represented as follows

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

To compute P(X < ½), we substitute x = ½ in F(x) since P(X < ½) = P (X ≤ ½) for a continuous rv.

F(1/2) = 3(1/4) – 2(1/8) = $\frac{3}{4}-\frac{1}{4}=\frac{1}{2}$ = 0.5

Exercise 6

Show that the expression $g(x)=\dfrac{x+1}{2}$ can serve as a cdf for -1 ≤ x < 1.

Solution

If g(x) is to represent a cdf we must show that g(x) = 0 for x ≤ -1, g(x) = 1 for x ≥ 1, and 0 < g(x) < 1 for the interval -1 ≤ x < 1.

Now, $g(-1)=\dfrac{-1+1}{2}=0$, and $g(1)=\dfrac{1+1}{2}=1$. Let us select a value x = 0 in the given interval.

Then g(0) = $\frac{1}{2}$ where 0 < $\frac{1}{2}$ < 1.

Since all three requirements are satisfied, g(x) can serve as a cdf for -1 ≤ x < 1

4 DERIVING PROBABILITY DENSITIES FROM CUMULATIVE DISTRIBUTION FUNCTIONS

The cdf, if given, can be used to obtain the corresponding pdf. The cdf is also useful for computing the probabilities of various intervals.

Let *X* be a continuous random variable with the pdf f(x) and the cdf F(x).

Then for any two numbers *a* and *b* such that a < b,

P(X < a) = F(a). Hence,

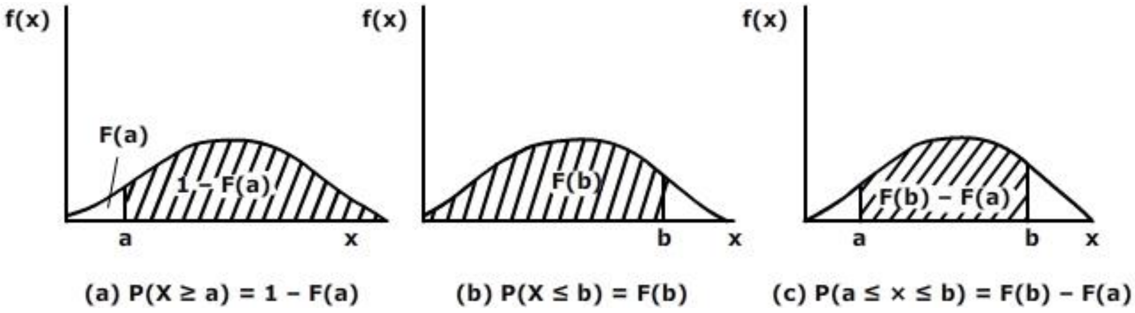
P(X > a) = 1 – F(a) and P(X < b) = F(b), so that

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

$= P(X < b) – P(X < a)$

= F(b) – F(a), as illustrated in Figure 6

Figure 6 Probability of an interval



For given cdf we can obtain the pdf by taking the derivative of F(x). By definition 3, if *X* is a continuous rv and the value of its probability density at *y* is *f(y)* then the cdf is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\, dy \quad \text{where } -\infty < x < \infty$$

Hence, $f(x) = \dfrac{dF(x)}{dx} = F'(x)$ at every x at which the derivative $F'(x)$ exists.

Example 4.1

In example 3.1, for the uniform distribution the cdf is

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \le x < B \\ 1 & x \ge B \end{cases}$$

The graph of F(x) is given in Fig 4(b).

It can be seen that F(x) is differentiable for A < x < B.

At x = A and x = B, F(x) cannot be differentiated.

For x < A, F(x) = 0 and for x > B, F(x) = 1

Hence, $F^{'}(x) = f(x) = 0$ if x < A, or, if x > B.

For A < x < B, $F^{'}(x) = \frac{d}{dx}\left(\frac{x-A}{B-A}\right) = \frac{1}{B-A}$ = f(x).

Thus we obtain the pdf of the uniform distribution as

$$f(x) = \begin{cases} 0 & x < A \\ \dfrac{1}{B-A} & A < x < B \\ 0 & x > B \end{cases}$$

Since x is continuous, f(x) = $\frac{1}{B-A}$ = P(A < x < B) = P(A $\leq$ x $\leq$ B)

Exercise 7

A continuous rv *Y* has a cdf given by

$$F(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

Compute P($\frac{1}{2}$ < Y ≤ $\frac{3}{4}$) in the two ways by using (a) the cdf, and (b) the pdf

Solution

(a) P($\frac{1}{2}$ < Y ≤ $\frac{3}{4}$) = F($\frac{3}{4}$) – F($\frac{1}{2}$) = $\frac{9}{16}$ - $\frac{1}{4}$ = $\frac{5}{16}$ = 0.3125

(b) First we obtain the pdf by differentiating F(y)

F'(y) = 0 for y < 0 and for y $\geq$ 1. If 0 $\leq$ y < 1, then F'(y) = f(y) = 2y so that the pdf is as follows:

$$f(y) = \begin{cases} 2y & 0 \leq y < 1 \\ 0 & otherwise \end{cases}$$

Then P($\frac{1}{2}$ < Y ≤ $\frac{3}{4}$) = $\int_{1/2}^{3/4} 2y\, dy = y^2 \Big|_{1/2}^{3/4}$

$$= \frac{9}{16} - \frac{1}{4} = \frac{5}{16}$$

= 0.3125

5 PERCENTILES OF A CONTINUOUS DISTRIBUTION

We know that the entire area under the graph of the pdf f(x), above the measurement axis, is 1. Therefore 100 percent of the probability distribution for all possible values of the continuous rv *X* lies to the left of the maximum value that *X* can take.

We may require to find two possible values *a* and *b* of *X* such that:

(i) a < b, and

(ii) a certain percentage of the area under the graph of f(x) lies between *a* and *b.*

For example, given the distribution of marks obtained by students in an examination, we may need to find the minimum marks scored by the top 5 percent of the students.

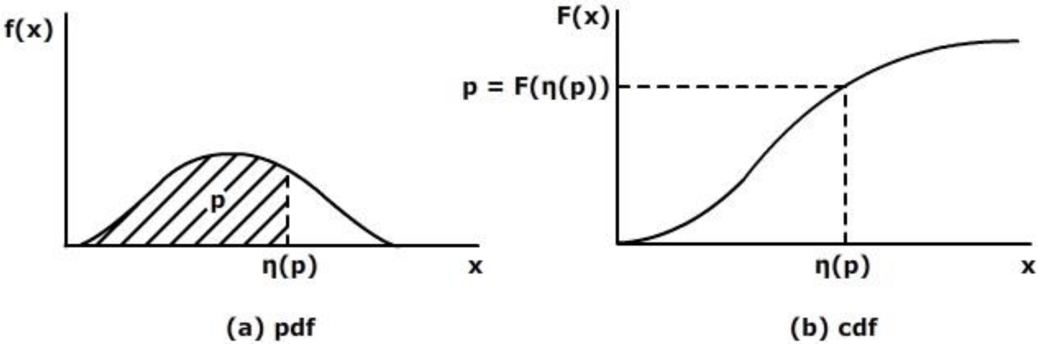
***Definition* 4**

Let *p* be a number between 0 and 1. The (100p)th ***percentile*** of the distribution of a continuous random variable X, denoted by η(p), is defined by

$$p = F[\eta(p)] = \int_{-\infty}^{\eta(p)} f(y)\,dy$$

Then η(p) is that value on the measurement axis such that 100p percent of the area under the graph of f(x) lies to the left of η(p) and 100(1-p) percent lies to the right. This is illustrated in Figure 7

Figure 7 Percentiles



If p = 0.3 then 30% of the area under the graph of f(x) lies to the left of η(0.3) and 70% to the right of η(0.3). The 30th percentile is denoted by η(0.3) since p = 0.3

Example 5.1

For the rv *X* with following pdf

$$f(x) = \begin{cases} \frac{1}{8}(x+1) & 2 < x < 4 \\ 0 & otherwise \end{cases}$$

To find the 75th percentile, η(0.75), we need to first obtain the cdf from the given pdf.

$$\text{F(x)} = \int_{-\infty}^{\infty} \frac{1}{8}(x+1)\,dx$$

$$\text{Therefore, F[}\eta\text{(p)] = p} = \int_{2}^{\eta(p)} \left(\frac{x}{8}+\frac{1}{8}\right) dx \;=\; \frac{x^2}{16}+\frac{x}{8}\Bigg|_{2}^{\eta(p)}$$

Substituting p = 0.75 = $\frac{3}{4}$ we obtain

$$\frac{3}{4} = \frac{[\eta(p)]^2}{16} + \frac{\eta(p)}{8} - \frac{4}{16} - \frac{2}{8} \;=\; \frac{1}{16}[\eta(p)]^2 + 2\eta(p) - 8$$

Rearranging the terms we get

$[\eta(p)]^2 + 2\eta(p) - 20 = 0$

Factorising, η(p) = $\dfrac{-2 \pm \sqrt{4+80}}{2} = -1 \pm 4.58 = 3.58\; or\; -5.58$

Since minimum value of *X* is 2 and the maximum is 4, the 75th percentile is 3.58 because that is the only possible value that *X* can take. The alternative value -5.58 does not fall in the range of possible values.

Hence, η(p) = 3.58

6 SHAPE OF THE PROBABILITY DISTRIBUTION

One of the applications of percentiles is to find the median of the distribution of a continuous random variable. The median is that value of the random variable where half of the distribution lies to the left of that value and the remaining 50 percent of the distribution is to the right of the value.

***Definition* 5**

The ***median*** of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile so that $\tilde{\mu}$ satisfies the condition F($\tilde{\mu}$) = 0.5

Example 6.1

The median of the pdf given in example 5.1 is computed by letting p = ½ so that

$$F[\tilde{\mu}] = \frac{1}{2} = \int_{2}^{\tilde{\mu}} \left(\frac{x}{8} + \frac{1}{8} \right) dx \ = \ \frac{x^2}{16} + \frac{x}{8}\bigg|_{2}^{\tilde{\mu}} = \frac{\tilde{\mu}^2}{16} + \frac{\tilde{\mu}}{8} - \frac{1}{2}$$

Therefore, $\frac{\tilde{\mu}^2}{16} + \frac{\tilde{\mu}}{8} - 1 = 0 \ \Rightarrow \tilde{\mu}^2 + 2\tilde{\mu} - 1600 = 0$

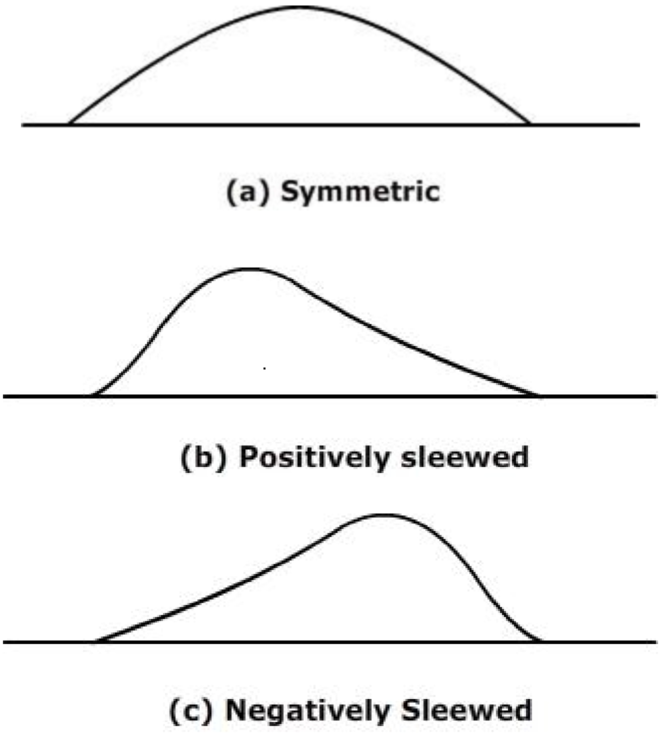
So that $\tilde{\mu} = \frac{-2 \pm \sqrt{4 + 64}}{2} = -1 \pm 4.123$

Since 2 < x < 4, $\tilde{\mu}$ = 3.123.

Half the area of the density curve is to the left of 3.123 and the other half is to the right.

If a random variable has a symmetric pdf then the median will coincide with the point of symmetry since half the area under the density curve lies on either side of the point. A positively skewed distribution has a long right-hand tail. Similarly, a negatively skewed distribution has a long left-hand tail. Figure 8 illustrates the three kinds of distributions.

Figure 8 Examples of symmetric and asymmetric distributions



(a) Symmetric

(b) Positively sleewed

(c) Negatively Sleewed

Example 6.2

The incomes of employees of a company will usually be positively skewed as there are a large number of low income workers and fewer employees with high income.

Example 6.3

A well known manufacturing company assures that its product will last a minimum period of three years. However, due to a defective component sourced from one of the suppliers, the lifetime of a batch of the product is likely to be drastically reduced. The distribution will then be negatively skewed.

It can be shown that for a symmetric pdf the median coincides with the mean of the distribution. If the mean and median have different values then the distribution is asymmetric, ie, skewed. If mean is less than median the distribution is skewed to the left or negatively skewed. On the other hand a distribution is positively skewed or skewed to the right when the mean is greater than the median.

The ***mode*** of the distribution is that value of the random variable at which the graph of the probability distribution reaches its highest point. If there is only one peak or "high point" it is a unimodal distribution. If there are two modes it is called a bimodal distribution. A distribution having more than two modes is said to be multimodal.

The mode of a unimodal distribution of a random variable is obtained by differentiating the probability density function. For the rv *X* with pdf *f(x)*, the mode is the value of *X* at which f'(x) = 0 and f''(x) < 0

Example 6.4

Suppose that the rv *X* has pdf

$$f(x) = \begin{cases} \frac{1}{9}\left(4 - x^2\right) & -1 \le x \le 2 \\ 0 & otherwise \end{cases}$$

Differentiating f(x) with respect to x, we get

$f'(x) = 0 - \frac{2x}{9}$

Setting $f'(x) = 0$ we get x = 0

Taking the second derivative,

$f''(x) = -\frac{2}{9}$ so that $f''(x) < 0$

Therefore, the mode of this pdf is at x = 0

Comparison of the mode and median can also be used to indicate the shape of the distribution. For a symmetric distribution mode = median. In case of a positively skewed distribution, medium > mode, whereas medium < mode for a negatively skewed distribution.

The other characteristics of the distribution like mean and variance can be computed with the help of mathematical expectations.

PRACTICE QUESTIONS

1. Suppose the rv Y has the pdf f(y = $4y^3$ *for* $0 \leq y \leq 1$ *and* 0 otherwise. Find $P(0 \leq Y \leq ½)$.

2. If Y is an exponential rv $f(y) = \lambda e^{-\lambda y}$ *for* $y \geq 0$ *and* 0 *otherwise*, find the cdf F(y).

3. Suppose the cdf of the rv Y is $F(y) = \frac{1}{12}(y^2 + y^3)$ *for* $0 \leq y \leq 2$ *and* 0 otherwise. Find the pdf f(y)

4. The amount of coffee (in grams) in a 230-gm jar filled by a certain machine is a random variable whose probability density is given by

$$f(x) = \begin{cases} 0 & x \leq 227.5 \\ \frac{1}{5} & 227.5 < x < 232.5 \\ 0 & x \geq 232.5 \end{cases}$$

Find the probabilities that a 230-gram jar filled by this machine will contain

(a) at most 228.65 gm of coffee

(b) anywhere from 229.34 to 231.66 gm of coffee

(c) at least 229.85 gm of coffee

5. Suppose the cdf for the continuous rv X is

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x^2}{4} & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Use the cdf to obtain the following:

(a) P(X $\leq$ 1)

(b) P(0.5 $\leq$ X $\leq$ 1)

(c) P(X > 1.5)

(d) Median

(e) pdf of X

6. The time taken by employees of a company to complete a task is a rv that has a uniform distribution. Let X= time taken in minutes. The minimum and maximum times are 10 minutes and 50 minutes respectively. For a new task, those taking less time will be considered efficient and given a bonus while those taking too much time are inefficient and will be sent for additional training. To qualify for bonus an employee must belong to the best 20 percent of all employees. To require training the employee must belong to the worst 30 percent category. What is the range of time for which an employee would neither get a bonus and nor be required to go for additional training?

7. In certain experiments, the error made in determining the velocity of a projectile is a random variable having a uniform density with minimum value α = - 0.015 and maximum value β = 0.015. Find the probabilities that such an error will

(i) be between – 0.002 and 0.003

(ii) exceed 0.005 in absolute value

8. If the continuous random variable *X* can take only non-negative values and has the density function $f(x) = 2e^{-2x}$ for $x \geq 0$, what is the maximum value of *X*? (*Hint*: Use conditions required to be satisfied by a pdf)

.