Continuous Probability Distributions

Prerequisites

You should be able to integrate.

Example (1)

Find the integral of $f(x) = 0.5 - 0.125x$ from $x = 0$ to $x = 4$.

Solution

$$I = \int_0^4 0.5 - 0.125x \; dx$$

$$= \left[0.5x - 0.0625x^2\right]_0^4$$

$$= 2 - 1$$

$$= 1$$

You should have studied discrete probability distributions and be aware of the distinction between the terms discrete and continuous. A continuous variable is a random variable whose values are not separated from each other. This means that between any two values, *a* and *b*, there exists a third value.

Example (2)

Classify the following variables as either discrete or continuous.

(*a*) The number of time it takes to obtain a 6 when throwing a fair cubical dice.

(*b*) The length of rivers in Africa.

(*c*) The number of defective components in a box.

(*d*) The mass of planets in the solar system.

Solution

(*a*) Discrete

(*b*) Continuous

(*c*) Discrete

(*d*) Continuous

In this chapter we shall be concerned with variables that take continuous values. They are called *continuous variables*.

Example (3)

What is required for a variable to be a *random* variable?

Solution

The term *random* indicates that to each value that the variable takes there is assigned a probability. The sum of all the probabilities must obey the law of total probability and so must equal 1.

Example (4)

(*a*) Find $I = \int_2^2 0.2x \, dx$.

(*b*) Let

$f(x) = 0.2x \qquad 0 \le x \le 5$

$f(x) = 0 \qquad \text{otherwise}$

(*i*) Find $f(2)$.

(*ii*) Let X be a continuous random variable defined on an interval that contains the value $x = 2$. Explain why the expression $P(X = 2) = 0.4$ must be **false**. Hint: it would be meaningful and possibly true if X were a *discrete* random variable.

Solution

(*a*) $I = \int_2^2 0.2x \, dx = \left[0.01x^2\right]_2^2 = 0$

(*b*) (*i*) $f(2) = 0.4$

(*ii*) Because X is a continuous variable it does not makes sense to assign a probability greater than 0 to a specific value that X can take. This is because the values are not discrete and therefore not cut off from each other. It is not possible to isolate the value $x = 2$ from the neighbourhood of other values that surround it. The value $x = 2$ is continuous with other values. Therefore, $P(X = 2) = 0.4$ is false. It is only possible to assign probabilities to an interval of values. Thus $P(1.9 \le X \le 2.1) \ne 0$ is meaningful and possibly true. On the other hand, it is meaningful and true that $P(X = 2) = 0$ because as part (*a*) shows any integral from 2 to 2 is zero. Strictly speaking for any continuous variable X the probability that $X = a$ where a is any discrete value is 0. Hence $P(X = a) = 0$ is always true for any continuous random variable X.

For continuous probability distributions probabilities shall be assigned to intervals. The probability will take the form $P(a < X < b) = \int_a^b f(x)\,dx$ where $f(x)$ is a function of a specific type that we call a *probability density function*. Because $\int_a^a f(x)\,dx = 0$ for any function f whatsoever, then for a continuous probability distribution it is not possible to draw a distinction between the expressions $<$ and $\leq$ in a statement about probability. Thus

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)\,dx .$$

Probability density function

Definition of a probability density function

Let $f(x)$ be a continuous function. If

(1) $f(x) \geq 0$ for all x

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

then $f(x)$ defines a continuous *probability density function*. The condition $f(x) \geq 0$ for all x is necessary because we cannot have negative probabilities.

Probability

The probability that X takes a value x in the interval $a < x < b$ is $P(a < X < b) = \int_a^b f(x)\,dx$.

Example (5)

Let

$f(x) = 0.5 - 0.125x$ for $0 \leq x \leq 4$

$f(x) = 0$ otherwise

(*a*) Show that $f(x)$ is a probability density function

(*b*) Find the probability that $2 \leq x \leq 3$.

(*c*) Sketch the function $f(x)$.

Solution

(*a*) We showed already in example (1) that

$$I = \int_0^4 0.5 - 0.125x \; dx = \left[0.5x - 0.0625x^2\right]_0^4 = 1$$

Since $f(x) > 0$ for all *x*, then *f* is a probability density function.

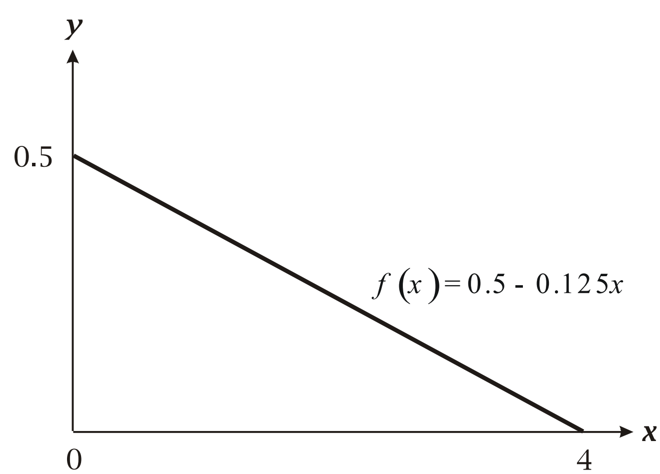
(*b*) $P(2 \le x \le 3) = \int_2^3 0.5 - 0.125x.dx$

$$= \left[0.5x - 0.0625x^2\right]_2^3$$

$$= 1.5 - 0.5625 - (1 - 0.25)$$

$$= 0.1875$$

(*c*) The graph of $y = f(x)$ a straight line. The area of the triangle under the line is 1.



In this example the probability density function $f(x)$ is the *piecewise* addition of three parts.

$f(x) = 0$ if $x < 0$

$f(x) = 0.5 - 0.125x$ if $0 \le x \le 4$

$f(x) = 0$ if $x > 0$

Because *f(x)* has been constructed from “pieces”, we say it has been *defined piecewise.* Probability density functions can be constructed from all sorts of pieces – each piece specified by a different function. This is illustrated by the following example.

Example (6)

Show that the function

$$f(x)=\begin{cases}0 & \text{if } x<0\\ \frac{1}{2}\sin x & \text{if } 0\le x\le \frac{1}{2}+1\\ \frac{1}{2} & \text{if } \frac{\pi}{2}<x\le\frac{\pi}{2}+1\\ 0 & \text{if } x>\frac{\pi}{2}+1\end{cases}$$

is a probability density function, and sketch its curve.

Solution

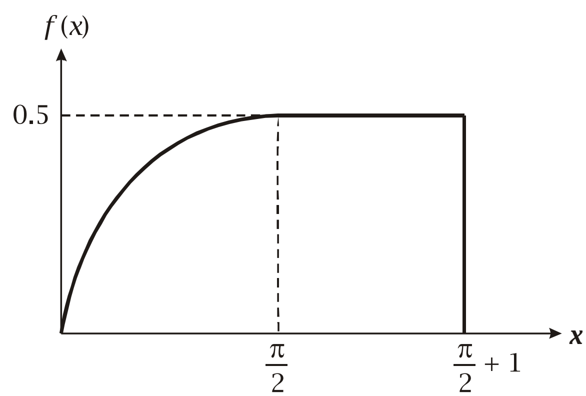
$$\int f(x)\,dx=\int_0^{\pi/2}\frac{1}{2}\sin x\,dx+\int_{\pi/2}^{\pi/2+1}\frac{1}{2}\,dx$$

$$=\left[-\frac{1}{2}\cos x\right]_0^{\pi/2}+\left[\frac{1}{2}x\right]_{\pi/2}^{\pi/2+1}$$

$$=0+\frac{1}{2}+\frac{1}{2}\left(\frac{\pi}{2}+1\right)-\frac{1}{2}\times\frac{\pi}{2}$$

$$=1$$

Since $f(x)>0$ for all x, then f is a pdf (probability density function). Its graph is as follows.



Example (9)

A continuous random variable X has probability density function

$f(x)=kx(4-x)$ $\quad 0\le x\le 4$

$f(x)=0$ $\quad$ otherwise

Find the value of k.

Solution

Since f is a probability density function then $\int_{-\infty}^{\infty} f(x)\,dx = 1$. As $f(x)$ takes the value 0 outside the interval $0 \le x \le 4$

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

$$\int_0^4 4kx - kx^2\,dx = 1$$

$$\left[2kx^2 - \frac{kx^3}{3}\right]_0^4 = 1$$

$$32k - \frac{64k}{3} = 1$$

$$k\left(\frac{32}{3}\right) = 1$$

$$k = \frac{3}{32}$$

Expectation of a continuous probability distribution

For a discrete probability distribution, the expectation is defined to be

$E(X) = \text{Sum for all values}(\text{value} \times \text{expected probability})$

$E(X) = \sum x\,P(X = x)$

Where $P(X = x)$ is the probability that the random variable X takes the value x. We will extend this definition to a continuous distribution. If X is a continuous random variable having probability density function $f(x)$ then

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

The expectation $E(x)$ may also be call the mean of the distribution.

Example (10)

The continuous random variable X has probability density function given by

$f(x) = kx^2 \qquad 0 \le x \le 3$

$f(x) = 0 \qquad \text{otherwise}$

(a) Find k. $\qquad (b)$ Show that $E(x) = \frac{9}{4}$

Solution

(a) $$\int_0^3 f(x)\,dx = 1$$

$$\int_0^3 kx^2.dx = 1$$

$$\left[\frac{kx^3}{3}\right]_0^3 = 1$$

$$9k = 1$$

$$k = \frac{1}{9}$$

(b) $$E(x) = \int_0^3 x f(x)\,dx$$

$$= \int_0^3 x\frac{x^2}{9}\,dx$$

$$= \int_0^3 \frac{x^3}{9}\,dx$$

$$= \left[\frac{x^4}{36}\right]_0^3$$

$$= \frac{81}{36}$$

$$= \frac{9}{4}$$

Variance of a continuous probability distribution

For a discrete probability distribution we define the variance by

$$E\left(X^2\right) = \frac{\text{Sum for all values }\left(\text{Square of expected values } \times \text{ expected probability}\right)}{n}$$

$$E(X^2) = \sum\left(x_i\right)^2 P\left(X = x_i\right)$$

$$Var(X) = E(X^2) - \left[E\left(X\right)\right]^2$$

We generalise this to a continuous probability density function so that if X is a continuous random variable having probability density function $f(x)$ then

$$E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)\,dx \qquad Var\left(X\right) = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

Example (11)

X is a continuous random variable with *uniform rectangular distribution* defined on the interval (5,20).

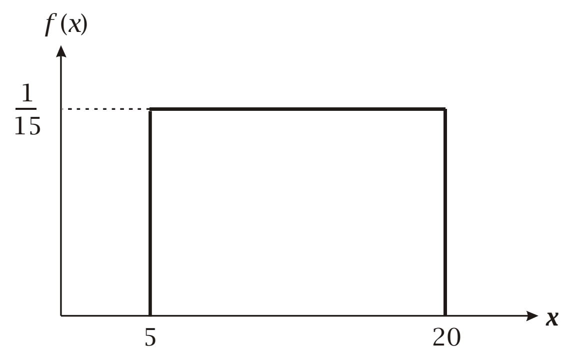
(*a*) Sketch the graph of the distribution.

(*b*) Define the pdf that gives rise to *X*.

(*c*) Find $E(X)$ and $Var(X)$.

Solution

(*a*) A uniform rectangular distribution means that $f(x) = \text{constant}$ on the interval on which it is defined. Since the width of the interval = 15, we have $f(x) = \frac{1}{15}$.



(*b*) $f(x) = \frac{1}{15}$ $\quad 5 \le x \le 20$

$f(x) = 0$ $\quad$ otherwise

(*c*) $$E(x) = \int_5^{20} x f(x) dx = \int_5^{20} \frac{1}{15} x \, dx = \left[\frac{x^2}{30}\right]_5^{20} = \frac{20^2}{30} - \frac{5^2}{30} = \frac{375}{30} = \frac{25}{2}$$

$$E(x^2) = \int_5^{20} x^2 f(x).dx = \int_5^{20} \frac{x^2}{15}.dx = \left[\frac{x^3}{45}\right]_5^{20} = \frac{20^3}{45} - \frac{5^3}{45} = \frac{7875}{45} = 175$$

$$Var(x) = E(x^2) - [E(x)]^2 = 175 - \left(\frac{25}{2}\right)^2 = \frac{760 - 625}{4} = \frac{75}{4}$$

Cumulative distribution function

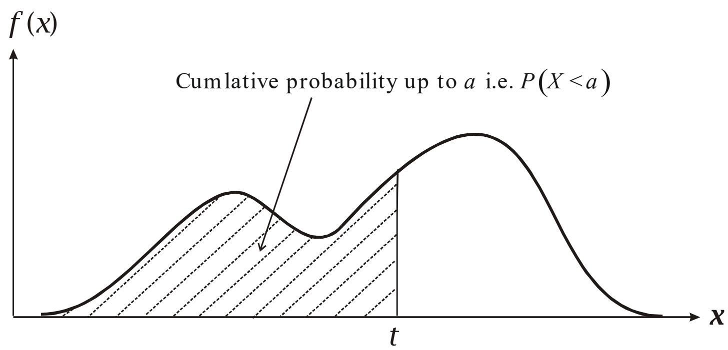
For a discrete random variable cumulative frequencies and probabilities are found by adding together frequencies or probabilities up to a given value. For instance, given the discrete probability distribution

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

The cumulative probabilities are

x	0	1	2	3	4
cumulative probability	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{7}{10}$	$\frac{9}{10}$	1

For a continuous probability distribution with probability density function $f(x)$ the probabilities are given by areas under the curve $y=f(x)$. Thus, the *cumulative probability* to a given value t is the area under curve $y=f(x)$ from $-\infty$ to t.



Hence if X is a random variable with probability density function $f(x)$ then the *cumulative distribution function* of X is given by

$$F(t)=P(X\le t)=\int_0^t f(x)\,dx$$

Example (12)

A continuous random variable X has probability density function

$f(x)=\frac{1}{8}(6-x) \qquad 1\le x\le 3$

$f(x)=0 \qquad$ otherwise

Find the cumulative distribution function for X.

Solution

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

$$= \int_{1}^{t} \frac{1}{8}(6 - x)\ dx$$

$$= \left[\frac{3}{4}x - \frac{x^2}{16}\right]_1^t$$

$$= \frac{3}{4}t - \frac{t^2}{16} - \left(\frac{3}{4} - \frac{1}{16}\right)$$

$$= \frac{3t}{4} - \frac{t^2}{16} - \frac{11}{16}$$

$$F(t) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{3}{4}t - \frac{t^2}{16} - \frac{11}{16} & \text{for } 1 \le t \le 3 \\ 1 & 1 \text{ if } x > 3 \end{cases}$$

Medians, quartiles and percentiles

The median of a probability distribution is that value that divides the distribution into two halves. That is, it is the value of the variable *X* such that the cumulative probability of *X* up to that value is $\frac{1}{2}$. Let *m* denote the median value. Then, if $F(t)$ is the cumulative distribution function for X, $F(m) = \frac{1}{2}$

If $f(x)$ is the probability density function for *X* then $\int_0^m f(x)dx = \frac{1}{2}$.

Example (13)

A random variable *X* has probability density function given by

$$f(x) = \begin{cases} \frac{2}{3}x & 0 \le x \le 1 \\ \frac{2}{3} & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the cumulative distribution function for *X* and find the median of *x*.

Solution

$$F(t) = \int_0^t \frac{2}{3} x dx \qquad \text{if } 0 \le t \le 1$$

$$= \frac{t^2}{3}$$

$$F(t) = \frac{1}{3} + \int_1^t \frac{2}{3}\, dx \qquad \text{if } 1 \le t \le 2$$

$$= \frac{1}{3} + \frac{2t}{3} - \frac{2}{3}$$

$$= \frac{2t}{3} - \frac{1}{3}$$

$$F(t) = 1 \qquad \text{if } t > 2$$

$$F(t) = \begin{cases} \frac{t}{3} & \text{for } 0 \le t \le 1 \\ \frac{2t}{3} - \frac{1}{3} & \text{for } 1 \le t \le 2 \\ 1 & \text{for } t > 2 \end{cases}$$

For the median $F(m) = \frac{1}{2}$

$$\frac{2m}{3} - \frac{1}{3} = \frac{1}{2}$$

$$\frac{2m}{3} = \frac{5}{6}$$

$$m = \frac{5}{4}$$

The first (lower) and third (upper) quartiles are the values t_1 and t_2 such that

$F(t_1) = \frac{1}{4}$ and $F(t_3) = \frac{3}{4}$

where $F(t)$ is the cumulative probability function.

A percentile is a value of t such that $F(t)$ is equal to a given percentage of the cumulative distribution.

Example (14)

The continuous random variable *X* has cumulative distribution function *F* given by

$$F(x)=\begin{cases}0 & \text{for } x<0\\ \frac{1}{10}\left(\frac{x^2}{2}+x^3\right) & \text{for } 0\le x\le 2\\ 1 & \text{otherwise}\end{cases}$$

(*a*) Evaluate $P(0.5\le X\le 1.2)$.

(*b*) Given that the lower quartile of *X*. lies between 1.20 and 1.21 find by trial and improvement the value of the lower quartile correct to 3 decimal places.

(*c*) Find an expression for $f(x)$ valid for $0\le x\le 2$, where *f* denotes the probability density function of *X*..

(*d*) Evaluate $E(X)$.

Solution

(*a*) $F(0.5)=\frac{1}{10}\left(\frac{0.5^2}{2}+0.5^3\right)=0.025$

$F(1.2)=\frac{1}{10}\left(\frac{1.2^2}{2}+1.2^3\right)=0.2448$

$P(0.5\le X\le 1.2)=F(1.2)-F(0.5)=0.2448-0.025=0.2198$

(*b*) The lower quartile if the value t_1 such that $F(t_1)=\frac{1}{4}$

$F(1.2)=0.2448$

$F(1.21)=0.2503...$

$F(1.205)=0.2475...$

$F(1.208)=0.2492...$

$F(1.209)=0.2498...$

$F(1.2095)=0.2500...$

$t_1=1.209\quad(3\text{ d.p.})$

(*c*) Since $F(t)=\int_0^t f(x)\,dx$

$f(x)$ is the derivative of $F(x)$

$f(x)=\frac{d}{dx}\left(\frac{1}{10}\left(\frac{x^2}{2}+x^3\right)\right)=\frac{1}{10}x+3x^2=\frac{1}{10}x(1+3x)$ for $0\le x\le 2$

(d) $$E(X) = \int_0^2 x f(x) dx$$

$$= \int_0^2 x\left(\frac{1}{10}x(1+3x)\right)dx$$

$$= \frac{1}{10}\int_0^2 x^2 + 3x^4 \; dx$$

$$= \frac{1}{10}\left[\frac{1}{3}x^3 + \frac{3}{4}x^4\right]_0^2$$

$$= \frac{1}{10}\left(\frac{8}{3} + 12\right)$$

$$= \frac{22}{15}$$