

Continuous Probability Distributions

Continuous Probability Distributions

Introduction

It is often possible to model real systems by using the same or similar random experiments and their associated random variables. Random variables may be classified in two distinct categories called discrete random variables and continuous random variables. Discrete random variables can take values which are discrete and which can be written in the form of a list. In contrast, continuous random variables can take values anywhere within a specified range. This Section will familiarize you with continuous random variables and their associated probability distributions. This Workbook makes no attempt to cover the whole of this large and important branch of statistics. The most commonly met continuous random variables in engineering are the Uniform, Exponential, Normal and Weibull distributions.

1. Continuous probability distributions

In order to get a good understanding of continuous probability distributions it is advisable to start by answering some fairly obvious questions such as: “What is a continuous random variable?” “Is there any carry over from the work we have already done on discrete random variables and distributions?” We shall start with some basic concepts and definitions.

Continuous random variables

In day-to-day situations met by practising engineers, experiments such as measuring current in a piece of wire or measuring the dimensions of machined components play a part. However closely an engineer tries to control an experiment, there will always be small variations in the results obtained due to many factors: the influence of factors outside the control of the engineer. Such influences include changes in ambient temperature which may affect the accuracy of measuring devices used, slight variation in the chemical composition of the materials used to produce the objects (wire, machined components in this case) under investigation. In the case of machined components, many of the small variations seen in measurements may be due to the influence of vibration, cutting tool wear in the machine producing the component, changes in raw material used and the process used to refine it and even the measurement process itself!

Such variations (current and length for example) can be represented by a **random variable** and it is customary to define an **interval**, finite or infinite, within which variation can take place. Since such a variable (X say) can assume *any* value within an interval we say that the variable is **continuous** rather than discrete. - its values form an entity we can think of as a continuum.

The following definition summarizes the situation.

Definition

A random variable X is said to be continuous if it can assume any value in a given interval. This contrasts with the definition of a discrete random variable which can only assume discrete values.

Practical example

This example will help you to see how continuous random variables arise and will help you to distinguish between continuous and discrete random variables.

Consider a de-magnetised compass needle mounted at its centre so that it can spin freely. Its initial position is shown in Figure 1(a). It is spun clockwise and when it comes to rest the angle θ , from the vertical, is measured. See Figure 1(b).

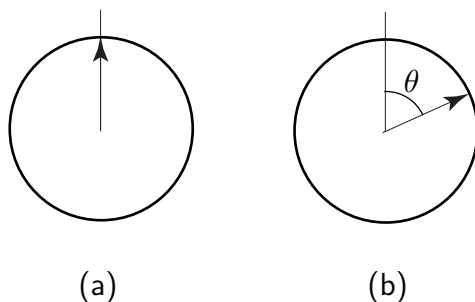


Figure 1

Let X be the random variable

“angle θ measured after each spin”

Firstly, note that X is a random variable since it can take any value in the interval 0 to 2π and we cannot be sure in advance which value it will take. However, after each spin and thinking in probability terms, there are certainly two distinct questions we can ask.

- What is the probability that X lies between two values a, b , i.e. what is $P(a < X < b)$?
- What is the probability that X assumes a *particular* value, say c . We are really asking what is the value of $P(X = c)$?

The first question is easy to answer provided we assume that the probability of the needle coming to rest in a given interval is given by the formula:

$$\text{Probability} = \frac{\text{Given interval in radians}}{\text{Total interval in radians}} = \frac{\text{Given interval in radians}}{2\pi}$$

The following results are easily obtained and they clearly coincide with what we intuitively feel is correct:

$$(a) \ P\left(0 < X < \frac{\pi}{2}\right) = \frac{1}{4} \text{ since the interval } (0, \pi/2) \text{ covers one quarter of a full circle}$$

$$(b) \ P\left(\frac{\pi}{2} < X < 2\pi\right) = \frac{3}{4} \text{ since the interval } (\pi/2, 2\pi) \text{ covers three quarters of a full circle.}$$

It is easy to see the generalization of this result for the interval (a, b) , in which both a, b lie in the interval $(0, 2\pi)$:

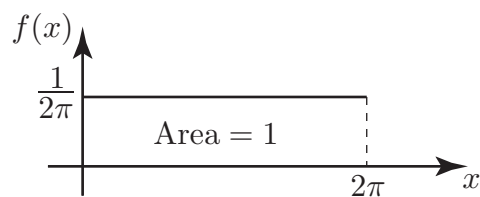
$$P(a < X < b) = \frac{b - a}{2\pi}$$

The second question immediately presents problems! In order to answer a question of this kind would require a measuring device (e.g. a protractor) with infinite precision: no such device exists nor could one ever be constructed. Hence it can **never** be verified that the needle, after spinning, takes any **particular** value; all we can be reasonably sure of is that the needle lies between two particular values.

We conclude that in experiments of this kind we **never** determine the probability that the random variable assume a particular value but only calculate the probability that it lies within a given range of values. This kind of random variable is called a **continuous random variable** and it is characterised, not by probabilities of the type $P(X = c)$ (as was the case with a discrete random variable), but by a function $f(x)$ called the **probability density function** (pdf for short). In the case of the rotating needle this function takes the simple form given with corresponding plot:

$$f(x) = \begin{cases} \frac{1}{2\pi}, & 0 \leq x < 2\pi \\ 0, & \text{elsewhere} \end{cases}$$

(a) Definition of the pdf



(b) Plot of the pdf

Figure 2

The probability $P(a < X < b)$ is the **area** under the function curve $f(x)$ and so is given by the integral

$$\int_a^b f(x)dx$$

Suppose we wanted to find $P\left(\frac{\pi}{6} < X < \frac{\pi}{4}\right)$. Then using the definition of the pdf for this case:

$$P\left(\frac{\pi}{6} < X < \frac{\pi}{4}\right) = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \frac{1}{2\pi} dx = \left[\frac{x}{2\pi} \right]_{\frac{\pi}{6}}^{\frac{\pi}{4}} = \frac{1}{2\pi} \left[\frac{\pi}{4} - \frac{\pi}{6} \right] = \frac{1}{2\pi} \times \frac{\pi}{12} = \frac{1}{24}$$

This is reasonable since the interval $\left(\frac{\pi}{6}, \frac{\pi}{4}\right)$ is one twenty-fourth of the interval 0 to 2π .

In general terms we have

$$P(a < X < b) = \int_a^b f(t)dt = F(b) - F(a) = \frac{b-a}{2\pi}$$

for the pdf under consideration here. Note also that

(a) $f(x) \geq 0$, for all real x

(b) $\int_{-\infty}^{\infty} f(x) dx = \int_0^{2\pi} \frac{1}{2\pi} dx = 1$, i.e. total probability is 1.

We are now in a position to give a formal definition of a continuous random variable in Key Point 1.

Key Point 1

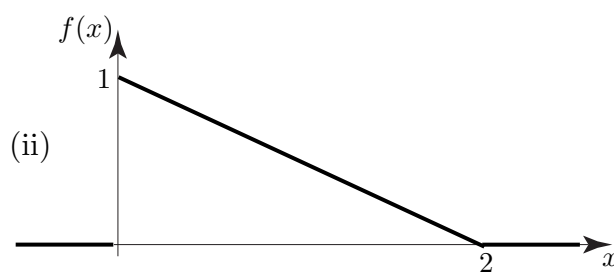
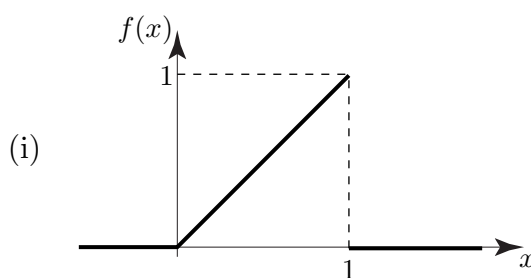
X is said to be a **continuous random variable** if there exists a function $f(x)$ associated with X called the **probability density function** with the properties

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$

The first two bullet points in Key Point 1 are the analogues of the results $P(X = x_i) \geq 0$ and $\sum_i P(X = x_i) = 1$ for discrete random variables.



Which of the following are not probability density functions?

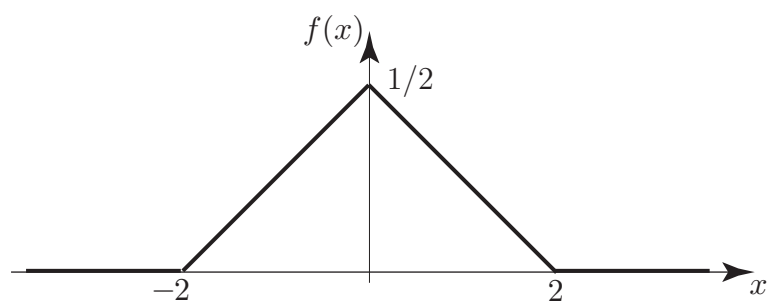


(iii)
$$f(x) = \begin{cases} x^2 - 4x + \frac{10}{3}, & 0 \leq x \leq 3 \\ 0, & \text{elsewhere} \end{cases}$$

Check whether the first two statements in Key Point 1 are satisfied for each pdf above:



Find the probability that X takes a value between -1 and 1 when the pdf is given by the following figure.



The cumulative distribution function

Analogous to the formula for the cumulative distribution function:

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

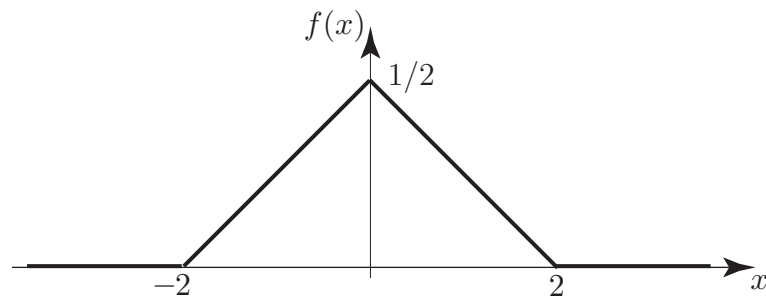
used in the case of a discrete random variable X with associated probabilities $P(X = x_i)$, we define a **cumulative probability distribution function** $F(x)$ by means of the integral (being a form of a sum):

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

The cdf represents the probability of observing a value less than or equal to x .

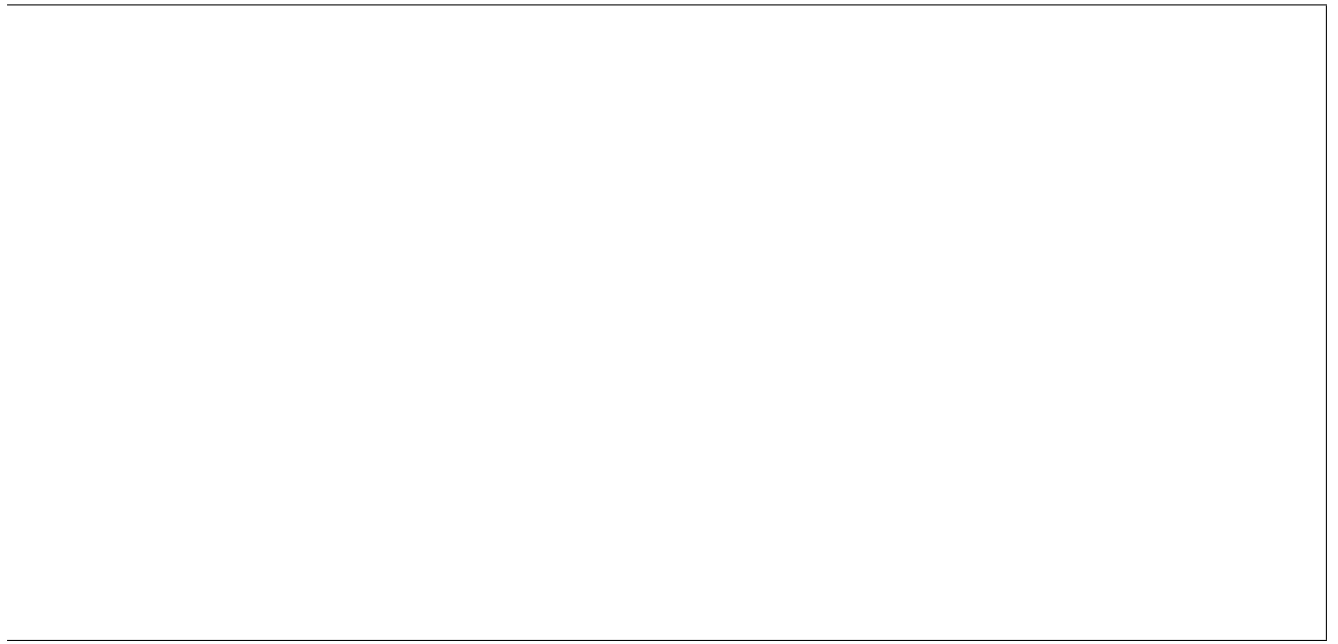


For the pdf in the diagram below



obtain the cdf and verify the result obtained in the previous Task for $P(-1 \leq X \leq 1)$.

Your solution





Example 1

Traditional electric light bulbs are known to have a mean lifetime to failure of 2000 hours. It is also known that the distribution function $p(t)$ of the time to failure takes the form

$$p(t) = 1 - e^{-t/\mu}$$

where μ is the mean time to failure. You will see if you study the topic of reliability in more detail that this is a realistic distribution function. The reliability function $r(t)$, giving the probability that the light bulb is still working at time t , is defined as

$$r(t) = 1 - p(t) = e^{-t/\mu}$$

Find the proportion of light bulbs that you would expect to fail before 1500 hours and the proportion you would expect to last longer than 2500 hours.

Mean and variance of a continuous distribution

You will probably have realised by now that, essentially, the definitions of discrete and continuous random variables are virtually the same provided we use the analogues given in the following table:

Quantity	Discrete Variable	Continuous Variable
Probability	$P(X = x)$	$f(x)dx$
Allowed Values	$P(X = x) \geq 0$	$f(x) \geq 0$
Summation	$\sum P(X = x)$	$\int f(x)dx$
Expectation	$E(X) = \sum xP(X = x)$?
Variance	$V(X) = \sum (x - \mu)^2 P(X = x)$?

Completing the above table of analogues to write down the mean and variance of a continuous variable leads to the *obvious* definitions given in Key Point 2:

Key Point 2

Let X be a continuous random variable with associated pdf $f(x)$. Then its expectation and variance denoted by $E(X)$ (or μ) and $V(X)$ (or σ^2) respectively are given by:

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

and

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

As with discrete random variables the variance $V(X)$ can be written in an alternative form, more amenable to calculation:

$$V(X) = E(X^2) - [E(X)]^2$$

where $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$.



For the variable X with pdf

$$f(x) = \begin{cases} \frac{1}{2}x, & 0 \leq x \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

find $E(X)$ and then $V(X)$.

First find $E(X)$:



The mileage (in 1000s of miles) for which a certain type of tyre will last is a random variable with pdf

$$f(x) = \begin{cases} \frac{1}{20}e^{-x/20}, & \text{for all } x > 0 \\ 0 & \text{for all } x < 0 \end{cases}$$

Find the probability that the tyre will last

- (a) at most 10,000 miles;
- (b) between 16,000 and 24,000 miles;
- (c) at least 30,000 miles.

Your solution

Important continuous distributions

There are a number of continuous distributions which have important applications in engineering and science. The areas of application and a little of the history (where appropriate) of the more important and useful distributions will be discussed in the later Sections and other Workbooks devoted to each of the distributions. Among the most important continuous probability distributions are:

- (a) the Uniform or Rectangular distribution, where the random variable X is restricted to a finite interval $[a, b]$ and $f(x)$ has constant density often defined by a function of the form:

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

(38.2)

- (b) the Exponential distribution defined by a probability density function of the form:

$$f(t) = \lambda e^{-\lambda t} \quad \lambda \text{ is a given constant}$$

(38.3)

- (c) the Normal distribution (often called the Gaussian distribution) where the random variable X is defined by a probability density function of the form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \mu, \sigma \text{ are given constants}$$

(39)

- (d) the Weibull distribution where the random variable X is defined by a probability density function of the form:

$$f(x) = \alpha\beta(\alpha x)^{\beta-1} e^{-(\alpha x)^\beta} \quad \alpha, \beta \text{ are given constants}$$

(46.1)