Contents 38

Continuous Probability Distributions

Learning outcomes

In this Workbook you will learn what a continuous random variable is. You wll find out how to determine the expectation and variance of a continuous random variable which are measures of the centre and spread of the distribution. You will learn about two distributions important in engineering - uniform and exponential.

Continuous Probability Distributions

38.1

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. The Uniform and Exponential distributions are introduced in Sections 38.2 and 38.3 while the Normal distribution and the Weibull distribution are covered in HELM 39 and HELM 46 respectively.

Prerequisites

Before starting this Section you should ...

- understand the concepts of probability
- be familiar with the concepts of expectation and variance

Learning Outcomes

On completion you should be able to ...

- explain what is meant by the term continuous random variable
- explain what is meant by the term continuous probability distribution
- use two continuous distributions which are important to engineers

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).

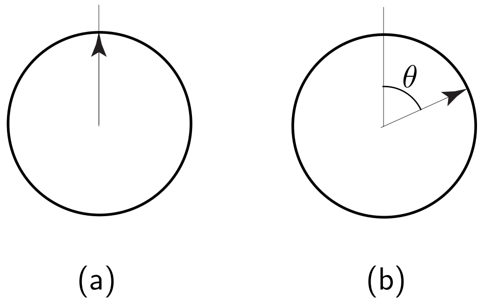


Figure 1

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 $\text{P}(a < X < b)$?
- What is the probability that X assumes a *particular* value, say c. We are really asking what is the value of $\text{P}(X = c)$?

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

(b) $\text{P}\left(\frac{\pi}{2} < X < 2\pi\right) = \frac{3}{4}$ since the interval $(\pi/2, 2\pi)$ 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 $(0, 2\pi)$:

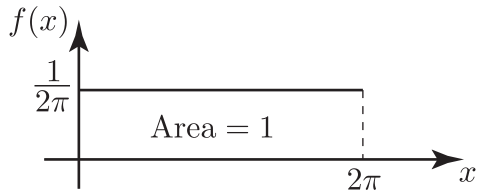
$$\text{P}(a < X < b) = \frac{b-a}{2\pi}$$

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 $\text{P}(X = c)$ (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:

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

(a) Definition of the pdf



(b) Plot of the pdf

Figure 2

The probability $\text{P}(a < X < b)$ 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 $\text{P}\left(\frac{\pi}{6} < X < \frac{\pi}{4}\right)$. Then using the definition of the pdf for this case:

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

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

In general terms we have

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

for the pdf under consideration here. Note also that

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

(b) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{2\pi} \frac{1}{2\pi}\,dx = 1,$ 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

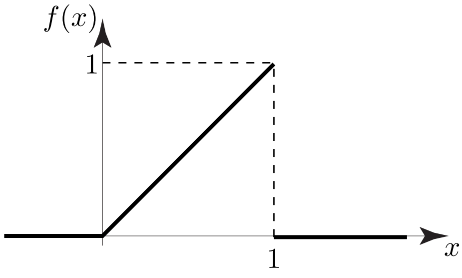
- $f(x) \geq 0$ for all x
- $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$
- $\displaystyle\text{P}(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$

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

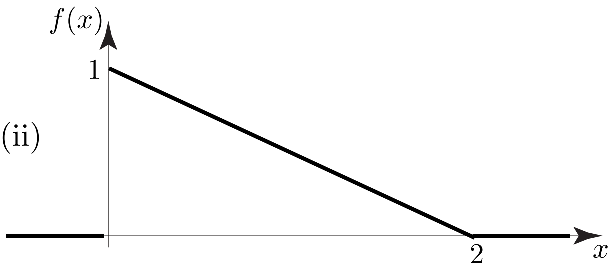
Task

Which of the following are not probability density functions?

(i)



(ii)



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

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

Your solution

For (i)

Answer

(i) We can write $f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 0, & \text{elsewhere} \end{cases}$

$f(x) \geq 0$ for all x but $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 x\,dx = \frac{1}{2} \neq 1.$

Thus this function is not a valid probability density function because the integral's value is not 1.

Your solution

For (ii)

Answer
(ii)

Note that $f(x) = \begin{cases} 1 - \frac{1}{2}x, & 0 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$ $\qquad f(x) \ge 0$ for all x

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^2 \left(1 - \frac{1}{2}x\right)\,dx = \left[x - \frac{x^2}{4}\right]_0^2 = 2 - 1 = 1$$

(Alternatively, the area of the triangle is $\frac{1}{2} \times 1 \times 2 = 1$)

This implies that $f(x)$ is a valid probability density function.

Your solution
For (iii)

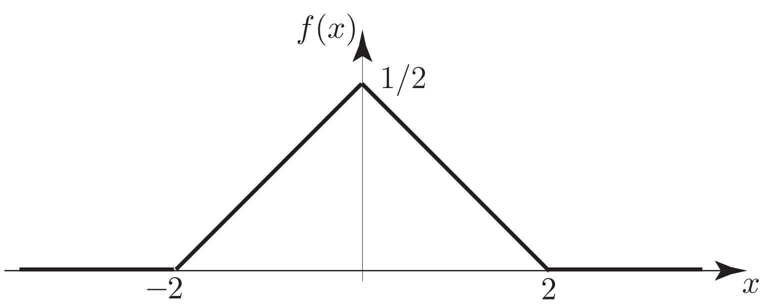
Answer
(iii)

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 \left(x^2 - 4x + \frac{10}{3}\right)\,dx = \left[\frac{x^3}{3} - 2x^2 + \frac{10}{3}x\right]_0^3 = (9 - 18 + 10) = 1$$

but $f(x) < 0$ for $1 \le x \le 3$. Hence (iii) is not a pdf.

Task

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



First find k:

Your solution

Answer

$$\int_{-\infty}^{\infty} f(x)\,dx = \text{area under curve} = \text{area of triangle} = \frac{1}{2}\times 4\times k = 2k$$

Also $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1,$ so $2k = 1$ hence $k = \dfrac{1}{2}$

State the formula for $f(x)$:

Your solution

Answer

$$f(x) = \begin{cases} \dfrac{1}{2} - \dfrac{1}{4}x, & 0 \le x \le 2 \\[2ex] \dfrac{1}{2} + \dfrac{1}{4}x, & -2 \le x < 0 \\[2ex] 0, & \text{elsewhere.} \end{cases}$$

Write down an integral to represent $\text{P}(-1 < X < 1)$. Use symmetry to evaluate the integral.

Your solution

Answer

$$\int_{-1}^{1} f(x)\,dx = 2\int_0^1 \left(\frac{1}{2} - \frac{1}{4}x\right)\,dx = 2\left[\frac{1}{2}x - \frac{1}{8}x^2\right]_0^1 = 2\left(\frac{1}{2} - \frac{1}{8}\right) = \frac{3}{4}$$

The cumulative distribution function

Analogous to the formula for the cumulative distribution function:

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

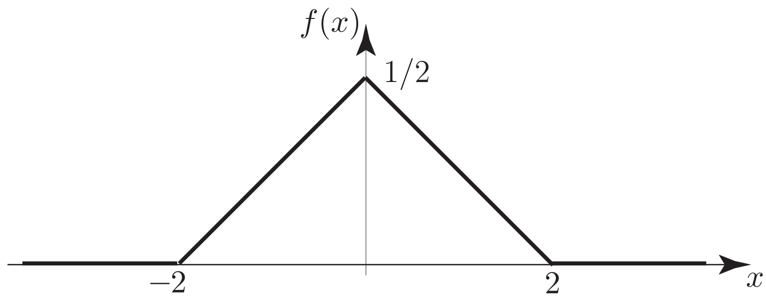
used in the case of a discrete random variable X with associated probabilities $\mathsf{P}(X = x_i)$, we define a **cumulative probability distribution function** $F(x)$ by means of the integral (being a form of a sum):

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

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

Task

For the pdf in the diagram below



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

Your solution

Answer

$$F(x)=\begin{cases} 0, & x\le -2 \\ \frac{1}{2}+\frac{1}{2}x+\frac{1}{8}x^2 & -2<x<0 \\ \frac{1}{2}+\frac{1}{2}x-\frac{1}{8}x^2 & 0<x\le 2 \\ 1 & x\ge 2 \end{cases}$$

$$\begin{aligned} \mathsf{P}(-1\le x\le 1)=F(1)-F(-1) &= \left(\frac{1}{2}+\frac{1}{2}=\frac{1}{8}\right)-\left(\frac{1}{2}-\frac{1}{4}+\frac{1}{8}\right) \\ &= \frac{1}{2}+\frac{1}{2}-\frac{1}{8}-\frac{1}{8}=\frac{3}{4}. \end{aligned}$$

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

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

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

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

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.

Solution

Let T be the random variable 'time to failure'.

The proportion of bulbs expected to fail before 1500 hours is given as

$$\mathrm{P}(T < 1500) = 1 - \mathrm{e}^{-1500/2000} = 1 - \mathrm{e}^{-3/4} = 1 - 0.4724 = 0.5276$$

The proportion of bulbs expected to last longer that 2500 hours is given as

$$\mathrm{P}(T > 2500) = 1 - \mathrm{P}(T \leq 2500) = \mathrm{e}^{-2500/2000} = \mathrm{e}^{-5/4} = 0.2865.$$

Using $r(t) = 1 - p(t)$ we have $r(2500) = 0.2865$.

Hence we expect just under 53% of light bulbs to fail before 1500 hours service and just under 29% of light bulbs to give over 2500 hours service.

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	$\mathrm{P}(X = x)$	$f(x)dx$
Allowed Values	$\mathrm{P}(X = x) \geq 0$	$f(x) \geq 0$
Summation	$\sum \mathrm{P}(X = x)$	$\int f(x)dx$
Expectation	$\mathrm{E}(X) = \sum x\mathrm{P}(X = x)$	?
Variance	$\mathrm{V}(X) = \sum (x - \mu)^2 \mathrm{P}(X = x)$	?

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 $\mathsf{E}(X)$ (or μ) and $\mathsf{V}(X)$ (or σ^2) respectively are given by:

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

and

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

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

$$\mathsf{V}(X) = \mathsf{E}(X^2) - [\mathsf{E}(X)]^2$$

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

Task

For the variable X with pdf

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

find $\mathsf{E}(X)$ and then $\mathsf{V}(X)$.

First find $\mathsf{E}(X)$:

Your solution

Answer

$$\mathsf{E}(X) = \int_0^2 \frac{1}{2}x.x\,dx = \left[\frac{1}{6}x^3\right]_0^2 = \frac{8}{6} = \frac{4}{3}.$$

Now find $\mathsf{E}(X^2)$:

Your solution

Answer

$\mathsf{E}(X^2) = \int_0^2 \frac{1}{2}x.x^2\,dx = \left[\frac{1}{8}x^4\right]_0^2 = 2.$

Now find $\mathsf{V}(X)$:

Your solution

Answer

$$\begin{aligned} \mathsf{V}(X) &= \mathsf{E}(X^2) - \{\mathsf{E}(X)\}^2 \\ &= 2 - \frac{16}{9} = \frac{2}{9}. \end{aligned}$$

Task

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

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

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

Answer

(a) $\mathsf{P}(a < X < b) = \int_a^b f(x)dx$

$$\begin{aligned}\mathsf{P}(X < 10) &= \int_{-\infty}^{10} f(x)\,dx \\ &= \int_0^{10} \frac{1}{20}e^{-x/20}dx = \Big[-e^{-x/20}\Big]_0^{10} = 0.393\end{aligned}$$

(b) $\mathsf{P}(16 < X < 24) = \int_{16}^{24} \frac{1}{20}e^{-x/20}dx = \Big[-e^{-x/20}\Big]_{16}^{24} = -e^{-1.2} + e^{-0.8} = 0.148$

(c) $\mathsf{P}(X > 30) = \int_{30}^{\infty} \frac{1}{20}e^{-x/20}dx = \Big[-e^{-x/20}\Big]_{30}^{\infty} = e^{-1.5} = 0.223$

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:

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

(HELM 38.2)

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

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

(HELM 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:

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

(HELM 39)

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

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

(HELM 46.1)

Exercises

1. A target is made of three concentric circles of radii $1/\sqrt{3}$, 1 and $\sqrt{3}$ metres. Shots within the inner circle count 4 points, within the middle band 3 points and within the outer band 2 points. (Shots outside the target count zero.) The distance of a shot from the centre of the target is a random variable R with density function $f(r) = \dfrac{2}{\pi(1+r^2)}$, $r > 0$. Calculate the expected value of the score after five shots.

2. A continuous random variable T has the following probability density function.

$$f_T(u) = \begin{cases} 0 & (u < 0) \\ 3(1 - u/k) & (0 \leq u \leq k) \\ 0 & (u > k) \end{cases} .$$

   Find

   (a) k.
   (b) E(T).
   (c) E(T^2).
   (d) V(T).

3. A continuous random variable X has the following probability density function

$$f_X(u) = \begin{cases} 0 & (u < 0) \\ ku & (0 \leq u \leq 1) \\ 0 & (u > 1) \end{cases}$$

   (a) Find k.
   (b) Find the distribution function $F_X(u)$.
   (c) Find E(X).
   (d) Find V(X).
   (e) Find E(e^X).
   (f) Find V(e^X).
   (g) Find the distribution function of e^X. (Hint: For what values of X is $e^X < u$?)
   (h) Find the probability density function of e^X.
   (i) Sketch $f_X(u)$.
   (j) Sketch $F_X(u)$.

Answers

1.

$$\begin{aligned}\text{P (inner circle)} &= \text{P}\left(0 < r < \frac{1}{\sqrt{3}}\right) = \int_0^{\frac{1}{\sqrt{3}}} \frac{2}{\pi(1+r^2)}dr = \frac{2}{\pi}\left[\ \tan^{-1} r\right]_0^{\frac{1}{\sqrt{3}}} \\ &= \frac{2}{\pi}\tan^{-1}\frac{1}{\sqrt{3}} = \frac{2}{\pi}\left(\frac{\pi}{6}\right) = \frac{1}{3}\end{aligned}$$

$$\begin{aligned}\text{P (middle band)} &= \text{P}\left(\frac{1}{\sqrt{3}} < r < 1\right) \\ &= \int_{\frac{1}{\sqrt{3}}} \frac{2}{\pi(1+r^2)}dr = \frac{2}{\pi}\left[\ \tan^{-1} r\right]_{\frac{1}{\sqrt{3}}}^{1} = \frac{2}{\pi}\tan^{-1} 1 - \frac{1}{3} = \frac{1}{6}.\end{aligned}$$

$$\begin{aligned}\text{P(outer band)} &= \text{P}(1 < r < \sqrt{3}) = \frac{2}{\pi}\left[\ \tan^{-1} r\right]_1^{\sqrt{3}} = \frac{2}{\pi}\tan^{-1}\sqrt{3} - \frac{1}{2} = \frac{1}{6} \\ \text{P(miss target)} &= 1 - \frac{1}{6} - \frac{1}{6} - \frac{1}{3} = \frac{1}{3}\end{aligned}$$

Let S be the random variable equal to 'score'.

s	0	2	3	4
$\text{P}(S = s)$	$^1/_3$	$^1/_6$	$^1/_6$	$^1/_3$

$$\text{E}(S) = 0 + \frac{2}{6} + \frac{3}{6} + \frac{4}{3} = \frac{13}{6}$$

The expected score after 5 shots is this value times 5 which is: $= 5\left(\frac{13}{6}\right) = 10.83.$

2.

(a) $1 = \int_0^k 3(1 - u/k)\, du = \left[3\left(u - \frac{u^2}{2k}\right)\right]_0^k = 3(k - k/2)$ so $k = 2/3.$

(b) $\text{E}(T) = \int_0^{2/3} 3u(1 - 3u/2)\, du = 3\int_0^{2/3} u - 3u^2/2\, du$

$$3\left[\frac{u^2}{2} - \frac{u^3}{2}\right]_0^{2/3} = 3\left(\frac{2}{9} - \frac{4}{27}\right) = 3\left(\frac{6-4}{27}\right) = \frac{2}{9}.$$

(c) $\text{E}(T^2) = \int_0^{2/3} 3u^2(1 - 3u/2)\, du = 3\int_0^{2/3} u^2 - 3u^3/2\, du$

$$= 3\left[\frac{u^3}{3} - \frac{3u^4}{8}\right]_0^{2/3} = 3\left(\frac{8}{81} - \frac{6}{81}\right) = 3\left(\frac{8-6}{81}\right) = \frac{2}{27}$$

(d) $\text{V}(T) = \text{E}(T^2) - \{\text{E}(T)\}^2 = \frac{2}{27} - \frac{4}{81} = \frac{2}{81}.$

Answers

3.

(a) $1 = \int_0^1 ku\,du = \left[\frac{ku^2}{2}\right]_0^1 = \frac{k}{2}$, so $k = 2$.

$$F_X(u) = \begin{cases} 0 & (u < 0) \\ u^2 & (0 \le u \le 1) \\ 1 & (1 < u) \end{cases}$$

(b) $\mathsf{E}(X) = \int_0^1 2u^2\,du = \left[\frac{2u^3}{3}\right]_0^1 = \frac{2}{3}$.

(c) $\mathsf{E}(X^2) = \int_0^1 2u^3\,du = \left[\frac{2u^4}{4}\right]_0^1 = \frac{1}{2}$. so $\mathsf{V}(X) = \mathsf{E}(X^2) - \{\mathsf{E}(X)\}^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$.

(e) $\mathsf{E}(e^X) = \int_0^1 2ue^u\,du = \Big[2ue^u\Big]_0^1 - 2\int_0^1 e^u\,du = \Big[2ue^u - 2e^u\Big]_0^1 = 2e - 2e + 2 = 2$

(f) $\mathsf{E}(e^{2X}) = \int_0^1 2ue^{2u}\,du = \Big[ue^{2u}\Big]_0^1 - \int_0^1 e^{2u}\,du = \Big[ue^{2u} - e^{2u}/2\Big]_0^1 = e^2$

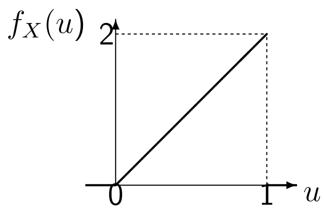
$= e^2/2 + 1/2 = (e^2+1)/2$ so $\mathsf{V}(e^X) = \mathsf{E}(e^{2X}) - \{\mathsf{E}(e^X)\}^2 = (e^2+1)/2 - 4$.

(g) $\mathsf{P}(e^X < u) = \mathsf{P}(X < \ln u) = (\ln u)^2$ for $0 < \ln u < 1$, i.e. $1 < u < e$.

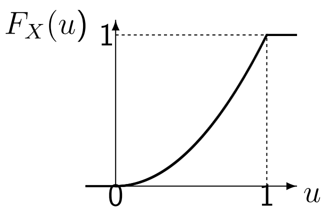
Hence the distribution function of e^X is $F_{e^X}(u) = \begin{cases} 0 & (u < 1) \\ (\ln u)^2 & (\le u \le e) \\ 1 & (e < u) \end{cases}$

(h) The pdf of e^X is $f_{e^X}(u) = \begin{cases} 0 & (u < 1) \\ \dfrac{2\ln u}{u} & (\le u \le e) \\ 0 & (e < u) \end{cases}$

(i) Sketch of pdf:



(j) Sketch of distribution function:



The Uniform Distribution

38.2

Introduction

This Section introduces the simplest type of continuous probability distribution which features a continuous random variable X with probability density function $f(x)$ which assumes a constant value over a finite interval.

Prerequisites

Before starting this Section you should ...

- understand the concepts of probability
- be familiar with the concepts of expectation and variance
- be familiar with the concept of continuous probability distribution

Learning Outcomes

On completion you should be able to ...

- explain what is meant by the term uniform distribution
- calculate the mean and variance of a uniform distribution

1. The uniform distribution

The Uniform or Rectangular distribution has random variable X restricted to a finite interval $[a, b]$ and has $f(x)$ a constant over the interval. An illustration is shown in Figure 3:

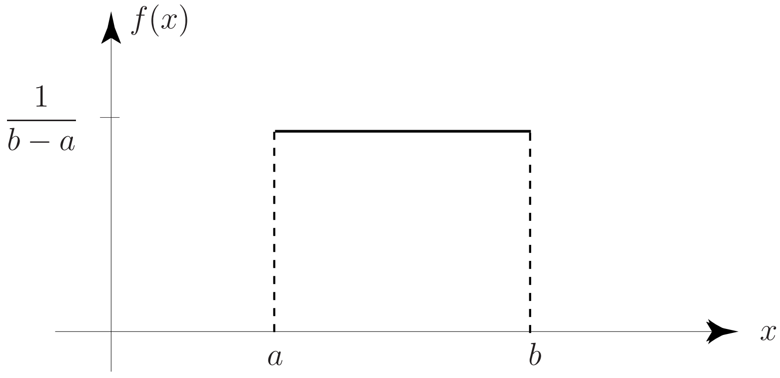


Figure 3

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

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

Mean and variance of a uniform distribution

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:

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

Using the formula for the variance, we may write:

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

Key Point 3

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

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

Example 2

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

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

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

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

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

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

Task

The thickness x of a protective coating applied to a conductor designed to work in corrosive conditions follows a uniform distribution over the interval $[20, 40]$ microns. Find the mean, standard deviation and cumulative distribution function of the thickness of the protective coating. Find also the probability that the coating is less than 35 microns thick.

Your solution

Answer

Over the interval $[20, 40]$ the probability density function $f(x)$ is given by the formula

$$f(x) = \begin{cases} 0.05, & 20 \leq x \leq 40 \\ 0 & \text{otherwise} \end{cases}$$

Using the formulae developed for the mean and variance gives

$$\mathsf{E}(X) = 10\ \mu\text{m} \qquad \text{and} \qquad \sigma = \sqrt{\mathsf{V}(X)} = \frac{20}{\sqrt{12}} = 5.77\ \mu\text{m}$$

The cumulative distribution function is given by

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

Hence, choosing appropriate ranges for x, the cumulative distribution function is obtained as:

$$F(x) = \begin{cases} 0, & x < 20 \\[2ex] \dfrac{x-20}{20} & 20 \leq x \leq 40 \\[2ex] 1 & x \geq 40 \end{cases}$$

Hence the probability that the coating is less than 35 microns thick is

$$F(x < 35) = \frac{35 - 20}{20} = 0.75$$

Exercises

1. In the manufacture of petroleum the distilling temperature (T°C) is crucial in determining the quality of the final product. T can be considered as a random variable uniformly distributed over 150°C to 300°C. It costs $£C_1$ to produce 1 gallon of petroleum. If the oil distills at temperatures less than 200°C the product sells for $£C_2$ per gallon. If it distills at a temperature greater than 200°C it sells for $£C_3$ per gallon. Find the expected net profit per gallon.

2. Packages have a nominal net weight of 1 kg. However their actual net weights have a uniform distribution over the interval 980 g to 1030 g.

   (a) Find the probability that the net weight of a package is less than 1 kg.

   (b) Find the probability that the net weight of a package is less than w g, where $980 < w < 1030$.

   (c) If the net weights of packages are independent, find the probability that, in a sample of five packages, all five net weights are less than wg and hence find the probability density function of the weight of the heaviest of the packages. (Hint: all five packages weigh less than w g if and only if the heaviest weighs less that w g).

Answers

1.

$$\text{P}(X < 200) = 50 \times \frac{1}{150} = \frac{1}{3} \qquad \text{P}(X > 200) = \frac{2}{3}$$

Let F be a random variable defining profit.
F can take two values $£(C_2 - C_1)$ or $£(C_3 - C_1)$

x	$C_2 - C_1$	$C_3 - C_1$
$\text{P}(F = x)$	$^1/_3$	$^2/_3$

$$\text{E}(F) = \left[\frac{C_2 - C_1}{3}\right] + \frac{2}{3}[C_3 - C_1] = \frac{C_2 - 3C_1 + 2C_3}{3}$$

2.

(a) The required probability is $\text{P}(W < 1000) = \dfrac{1000 - 98}{1030 - 980} = \dfrac{20}{50} = 0.4$

(b) The required probability is $\text{P}(W < w) = \dfrac{w - 980}{1030 - 980} = \dfrac{w - 980}{50}$

(c) The probability that all five weigh less than w g is $\left(\dfrac{w - 980}{50}\right)^5$ so the pdf of the heaviest is

$$\frac{d}{dw}\left(\frac{w - 980}{50}\right)^5 = \frac{5}{50}\left(\frac{w - 980}{50}\right)^4 = 0.1\left(\frac{w - 980}{50}\right)^4 \quad \text{for } 980 < w < 1030.$$

The Exponential Distribution

38.3

Introduction

If an engineer is responsible for the quality of, say, copper wire for use in domestic wiring systems, he or she might be interested in knowing both the number of faults in a given length of wire and also the distances between such faults. While the number of faults may be analysed by using the Poisson distribution, the distances between faults along the wire may be shown to give rise to the exponential distribution defined and used in this Section.

Prerequisites

Before starting this Section you should . . .

- understand the concepts of probability
- be familiar with the concepts of expectation and variance
- be familiar with the concepts of continuous distributions, in particular the Poisson distribution.

Learning Outcomes

On completion you should be able to . . .

- understand what is meant by the term exponential distribution
- calculate the mean and variance of an exponential distribution
- use the exponential distribution to solve simple practical problems

1. The exponential distribution

The exponential distribution is defined by

$$f(t) = \lambda \mathrm{e}^{-\lambda t} \qquad t \geq 0 \qquad \lambda \text{ a constant}$$

or sometimes (see the Section on Reliability in HELM 46) by

$$f(t) = \frac{1}{\mu}\mathrm{e}^{-t/\mu} \qquad t \geq 0 \qquad \mu \text{ a constant}$$

The advantage of this latter representation is that it may be shown that the mean of the distribution is μ.

Example 3

The lifetime T (years) of an electronic component is a continuous random variable with a probability density function given by

$$f(t) = \mathrm{e}^{-t} \qquad t \geq 0 \qquad (\text{i.e. } \lambda = 1 \text{ or } \mu = 1)$$

Find the lifetime L which a typical component is 60% certain to exceed. If five components are sold to a manufacturer, find the probability that at least one of them will have a lifetime less than L years.

Solution

We require $\mathrm{P}(T > L) = 0.6$. We know that this probability is given by the relationship

$$\mathrm{P}(T > L) = \int_L^\infty \mathrm{e}^{-t}\,dt = \Big[-\mathrm{e}^{-t} \Big]_L^\infty = \mathrm{e}^{-L}$$

Solving $\mathrm{e}^{-L} = 0.6$ for the least value of L we obtain $L = 0.51$ years.
Assuming that the lifetime of each component is independent we have

P(at least one component has a lifetime less than 0.51 years)

$= 1 -$ P(no component has a lifetime less than 0.51 years)

$= 1 - 0.6^5$

$= 0.92$

Task

Commonly, car cooling systems are controlled by electrically driven fans. Assuming that the lifetime T in hours of a particular make of fan can be modelled by an exponential distribution with $\lambda = 0.0003$ find the proportion of fans which will give at least 10000 hours service. If the fan is redesigned so that its lifetime may be modelled by an exponential distribution with $\lambda = 0.00035$, would you expect more fans or fewer to give at least 10000 hours service?

Your solution

Answer

We know that $f(t) = 0.0003\mathrm{e}^{-0.0003t}$ so that the probability that a fan will give at least 10000 hours service is given by the expression

$$\mathrm{P}(T > 10000) = \int_{10000}^{\infty} f(t)\,dt = \int_{10000}^{\infty} 0.0003\mathrm{e}^{-0.0003t}\,dt = -\left[\mathrm{e}^{-0.0003t}\right]_{10000}^{\infty} = e^{-3} \approx 0.0498$$

Hence about 5% of the fans may be expected to give at least 10000 hours service. After the redesign, the calculation becomes

$$\mathrm{P}(T > 10000) = \int_{10000}^{\infty} f(t)\,dt = \int_{10000}^{\infty} 0.00035\mathrm{e}^{-0.00035t}\,dt = -\left[\mathrm{e}^{-0.00035t}\right]_{10000}^{\infty} = e^{-3.5} \approx 0.0302$$

and so only about 3% of the fans may be expected to give at least 10000 hours service.

Hence, after the redesign we expect *fewer* fans to give 10000 hours service.

Exercises

1. The time intervals between successive barges passing a certain point on a busy waterway have an exponential distribution with mean 8 minutes.

   (a) Find the probability that the time interval between two successive barges is less than 5 minutes.

   (b) Find a time interval t such that we can be 95% sure that the time interval between two successive barges will be greater than t.

2. It is believed that the time X for a worker to complete a certain task has probability density function $f_X(x)$ where

$$f_X(x) = \begin{cases} 0 & (x \leq 0) \\ kx^2e^{-\lambda x} & (x > 0) \end{cases}$$

   where λ is a parameter, the value of which is unknown, and k is a constant which depends on λ.

   (a) Show that if $I_n = \displaystyle\int_0^\infty x^n e^{-\lambda x}\, dx$ then $I_n = \dfrac{n}{\lambda} I_{n-1}$, where $n > 0$ and $\lambda > 0$.

   Evaluate $I_0 = \displaystyle\int_0^\infty e^{-\lambda x}\, dx$ and hence find a general expression for I_n.

   This result can be used in the rest of this question.

   (b) Find, in terms of λ, the value of k.

   (c) Find, in terms of λ, the expected value of X.

   (d) Find, in terms of λ, the variance of X.

   (e) Write down the expected value and variance of the sample mean of a sample of n independent observations on X.

   (f) Find, in terms of λ, the expected value of X^{-1}.

Answers

1. We have $\mu = 8$ so $\lambda = 0.125$.

   (a) The probability is

$$\text{P}(T < 5) = \int_0^5 0.125e^{-0.125t}\ dt = 1 - e^{-0.125\times 5} = 0.4647.$$

   (b) We require

$$\int_t^\infty 0.125e^{-0.125x}\ dx = e^{-0.125t} = 0.95.$$

   So $-0.125t = \log 0.95$ and

$$t = -\frac{\log 0.95}{0.125} = 0.4103.$$

   That is, 24.6 s.

2.

(a) $I_n = \displaystyle\int_0^\infty x^n e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda}x^n e^{-\lambda x}\right]_0^\infty + \frac{n}{\lambda} \qquad \int_0^\infty x^{n-1}e^{-\lambda}\, dx = \frac{n}{\lambda}I_{n-1}$

$I_0 = \displaystyle\int_0^\infty e^{-\lambda x}\, dx = \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}$ hence $I_n = \dfrac{n!}{\lambda^{n+1}}$.

(b) $\displaystyle\int_0^\infty kx^2 e^{-\lambda x}\, dx = 1 \Rightarrow kI_2 = 1 \Rightarrow k = \frac{1}{I_2} = \frac{\lambda^3}{2}$

(c) $\text{E}(X) = \displaystyle\int_0^\infty x f_X(x)\, dx = kI_3 = \frac{\lambda^3}{2}\frac{6}{\lambda^4} = \frac{3}{\lambda}$

(d) $\text{E}(X^2) = \displaystyle\int_0^\infty x^2 f_X(x)\, dx = kI_4 = \frac{\lambda^3}{2}\frac{24}{\lambda^5} = \frac{12}{\lambda^2}$

so $\text{V}(X) = \text{E}(X^2) - \{\text{E}(X)\}^2 = \dfrac{12}{\lambda^2} - \dfrac{9}{\lambda^2} = \dfrac{3}{\lambda^2}$

(e) $\text{E}(\bar{X}) = \dfrac{3}{\lambda} \qquad \text{V}(\bar{X}) = \dfrac{3}{n\lambda^2}$

(f) $\text{E}\left(\dfrac{1}{X}\right) = \displaystyle\int_0^\infty \frac{1}{x} f_X(x)\, dx - kI_1 = \frac{\lambda^3}{2}\frac{1}{\lambda^2} = \frac{\lambda}{2}$