

1 CONTINUOUS PROBABILITY DISTRIBUTIONS

Objectives

After studying this chapter you should

- know the definition of $E(X)$ and $E(g(X))$ for both discrete and continuous distributions;
- know the definition of mean and variance in terms of expectations;
- be able to do calculations involving linear combinations of independent normal random variables;
- be able to calculate probabilities using the exponential distribution.

1.0 Introduction

In the text *Statistics*, continuous probability distributions were introduced in Chapter 7. Probability was represented by the area under a curve, known as the probability density function. A probability density function of a random variable, X , must be non-negative for all values of X and the total area under the curve must be 1.

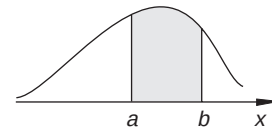
If the **probability density function** is denoted $f(x)$ then the probability that an observed value of X lies between a and b is given by

$$P(a < x < b) = \int_a^b f(x) dx$$

Also note that

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

The most important continuous distribution and indeed the most important distribution of any kind in statistics is the normal distribution. This was introduced in *Statistics*, Chapter 8. A familiarity with the normal distribution is essential for an understanding of what follows.



This chapter starts by drawing together definitions and laws of expectations and variances for discrete and continuous distributions. Many of the ideas have already been met in *Statistics*. Some well known results are proved. Students may prefer to omit the proofs, particularly at first reading. It is possible to leave sections of this chapter, which deal with some of the more mathematical ideas in *Statistics*, until later in the course if preferred.

1.1 Expectation

For a discrete probability distribution

$$E(X) = \sum_x xP(X = x)$$

If $g(x)$ is a function of x then

$$E(g(x)) = \sum_x g(x)P(X = x)$$

If, for example, the random variable X can take the values 1, 2 or 3 with probabilities 0.2, 0.3 and 0.5 respectively, then

$$E(X) = 0.2 \times 1 + 0.3 \times 2 + 0.5 \times 3 = 2.3$$

$$E(\sqrt{X}) = 0.2 \times \sqrt{1} + 0.3 \times \sqrt{2} + 0.5 \times \sqrt{3} = 1.49.$$

For a continuous distribution

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

and

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

Example

The random variable X has probability density function

$$\frac{3}{4}x(2-x) \quad \text{for } 0 < x < 2 \quad \text{Find } E(X) \text{ and } E\left(\frac{1}{X}\right).$$

Solution

$$\begin{aligned} E(X) &= \int_0^2 \frac{3}{4}x^2(2-x)dx \\ &= \frac{3}{4} \int_0^2 (2x^2 - x^3)dx \end{aligned}$$

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

$$= \frac{3}{4} \left(\frac{16}{3} - 4 \right)$$

$$= 1$$

and $E\left(\frac{1}{X}\right) = \int_0^2 \frac{1}{x} \cdot \frac{3}{4} x(2-x) dx$

$$= \int_0^2 \frac{3}{4} (2-x) dx$$

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

$$= 1.5$$

1.2 Mean and variance

For any random variable whether continuous or discrete, the mean is defined to be $E(X)$ and this is usually denoted by μ .

The variance is defined to be

$$E(X - E(X))^2$$

and this is usually denoted σ^2

$$\sigma^2 = E([X - \mu]^2)$$

An equivalent formula is

$$\sigma^2 = E(X^2) - \mu^2 \quad (\text{see section 1.4})$$

Why is the second formula for σ^2 preferred for calculations?

The next example is to revise some of the ideas introduced in the text *Statistics*, Chapter 7.

Example

A charity group raises funds by collecting waste paper. A skip-full will contain an amount, X , of other materials such as plastic bags and rubber bands. X may be regarded as a random variable

with probability density function

$$f(x) = k(x-1)(4-x) \quad 1 < x < 4$$

$$= 0 \quad \text{otherwise.}$$

(All numerical values in this question are in units of 100 kg.)

- (a) Show that $k = \frac{2}{9}$.
- (b) Find the mean and standard deviation of X .
- (c) Find the probability that X exceeds 3.5.

A skip-full may normally be sold for £250 but if X exceeds 3.5 only £125 will be paid. Find the expected value of a skip-full.

Alternatively the paper may be sorted before being placed in the skip. This will ensure a very low value of X and a skip-full may then be sold for £310. However the effort put into sorting means that 25 per cent fewer skips-full will be sold. Advise the charity whether or not to sort the paper. (AEB)

Solution

(a) $\int_1^4 k(x-1)(4-x)dx = 1$

$$k \int_1^4 (-x^2 + 5x - 4)dx = 1$$

$$k \left[\frac{-x^3}{3} + \frac{5x^2}{2} - 4x \right]_1^4 = 1$$

$$k \left[\frac{8}{3} - \left(-\frac{11}{6} \right) \right] = 1$$

$$4.5k = 1$$

$$\frac{9}{2}k = 1 \Rightarrow k = \frac{2}{9}$$

(b) Mean, $E(X) = \int_1^4 \frac{2}{9}x(x-1)(4-x)dx$

$$= \frac{2}{9} \left(\int_1^4 -x^3 + 5x^2 - 4x \right) dx$$

$$= \frac{2}{9} \left[\frac{-x^4}{4} + \frac{5x^3}{3} - \frac{4x^2}{2} \right]_1^4$$

$$= \frac{2}{9} \left(\frac{32}{3} - \left(-\frac{7}{12} \right) \right)$$

$$= 2.5$$

$$\begin{aligned}
E(X^2) &= \int_1^4 \frac{2}{9}x^2(x-1)(4-x)dx \\
&= \frac{2}{9} \int_1^4 (-x^4 + 5x^3 - 4x^2)dx \\
&= \frac{2}{9} \left[-\frac{x^5}{5} + \frac{5x^4}{4} - \frac{4x^3}{3} \right]_1^4 \\
&= \frac{2}{9} \left(\frac{448}{15} - \left(-\frac{17}{60} \right) \right) \\
&= 6.7
\end{aligned}$$

Hence

$$\text{variance} = E(X^2) - \mu^2 = 6.7 - 2.5^2 = 0.45$$

and

$$\text{standard deviation} = \sqrt{\text{variance}} = 0.671.$$

(c) Probability that X exceeds 3.5 is given by

$$\begin{aligned}
P(X > 3.5) &= \int_{3.5}^4 \frac{2}{9}(x-1)(4-x)dx \\
&= \frac{2}{9} \left[-\frac{x^3}{3} + \frac{5x^2}{2} - 4x \right]_{3.5}^4 \\
&= \frac{2}{9} \left(\frac{8}{3} - \frac{7}{3} \right) \\
&= 0.0741
\end{aligned}$$

(d) Expected value of a skip-full

$$\begin{aligned}
&= £125 \times 0.0741 + £250 \times (1 - 0.0741) \\
&= £241
\end{aligned}$$

For each skip-full of unsorted material only 0.75 skips-full of paper would be obtained after sorting. This would be worth

$$0.75 \times £310 = £232.5$$

Thus the charity will make more money in the long run if the paper is not sorted.

Exercise 1A

- A temporary roundabout is installed at a crossroads. The time, X minutes, which vehicles have to wait before entering the crossroads has probability density function

$$f(x) = \begin{cases} 0.8 - 0.32x & 0 < x < 2.5 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and standard deviation of X .
(AEB)
- The random variable, Y , has probability density function

$$f(y) = \begin{cases} k(8 - 2y) & 0 < y < 4 \\ 0 & \text{otherwise} \end{cases}$$
 - Verify that $k = 0.0625$ and that the median is 1.172.
 - Find the mean and standard deviation of Y .
 - What is the probability that the mean of a random sample of size 50 from this distribution will lie between the mean and the median?
(AEB)
- A technique for measuring the density of a silicon compound is a random variable, X , with probability density function

$$f(x) = \begin{cases} k & -0.04 < x < 0.04 \\ 0 & \text{otherwise} \end{cases}$$
 - Find the value of k .
 - Find the mean and standard deviation of X .
 - Find the probability that the error is between -0.03 and 0.01 .
 - Find the probability that the magnitude of the error is greater than 0.035 .

1.3 Laws of expectation

The following laws apply to expectations of both discrete and continuous random variables.

(i) $E(a + bX) = a + bE(X)$

where a and b are constants and X is a random variable.

(ii) $E(X + Y) = E(X) + E(Y)$

where X and Y are any random variables.

Laws (i) and (ii) may be extended to a more general form

(iii)
$$\begin{aligned} E(a_1X_1 + a_2X_2 + \dots + a_nX_n) \\ = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n) \end{aligned}$$

when $X_1, X_2, \dots, X_n$ are random variables and $a_1, a_2, \dots, a_n$ are constants.

(iv) $E(XY) = E(X)E(Y)$

when X and Y are independent random variables.

Note that (iv) is only true if X and Y are independent.

1.4 Some results for variances

The laws of expectation together with the definition of variance can be used to derive some interesting and useful results. In the following $V(X)$ is used to denote the variance of X and $E(X) = \mu$ and $V(X) = \sigma^2$.

Note: Brief proofs are included but will not be tested in the AEB examination and may be omitted.

$$\begin{aligned}
 \text{(i)} \quad & \boxed{V(X) = E([X - \mu]^2)} \\
 & = E(X^2) - 2\mu E(X) + \mu^2 \\
 & = E(X^2) - 2\mu^2 + \mu^2 \\
 & = E(X^2) - \mu^2
 \end{aligned}$$

This result has been used in *Statistics* and in Section 1.2 above.

$$\text{(ii)} \quad \boxed{V(a + bX) = b^2 V(X)}$$

where a and b are constants.

Since

$$\begin{aligned}
 E(a + bX) &= a + bE(X) \\
 &= a + b\mu,
 \end{aligned}$$

then

$$\begin{aligned}
 V(a + bX) &= E([a + bX - a - b\mu]^2) \\
 &= E(b^2[X - \mu]^2) \\
 &= b^2 E([X - \mu]^2) \\
 &= b^2 V(X)
 \end{aligned}$$

(iii) If X and Y are **independent** random variables

$$\boxed{V(X \pm Y) = V(X) + V(Y)}$$

Note: (a) this is only true if X and Y are independent;

(b) the left hand side has a sign but the variances are always added. Clearly it would not make sense to take one variance from another as this could lead to a negative result. A negative variance is impossible.

The proof follows from

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ &= \mu_x + \mu_y, \text{ say} \end{aligned}$$

$$\begin{aligned} V(X \pm Y) &= E\left(\left[X \pm Y - \{\mu_x \pm \mu_y\}\right]^2\right) \\ &= E\left(\left[\{X - \mu_x\} \pm \{Y - \mu_y\}\right]^2\right) \\ &= E\left(\left[X - \mu_x\right]^2\right) + E\left(\left[Y - \mu_y\right]^2\right) \pm 2E\left(\left[X - \mu_x\right]\left[Y - \mu_y\right]\right) \\ &= V(X) + V(Y) \pm 2E(X - \mu_x)E(Y - \mu_y) \end{aligned}$$

since X and Y are independent.

Now

$$\begin{aligned} E(X - \mu_x) &= E(X) - \mu_x \\ &= \mu_x - \mu_x \\ &= 0 \end{aligned}$$

Hence

$$V(X \pm Y) = V(X) + V(Y).$$

(iv) The results above can be extended to

$$\begin{aligned} &V(a_1X_1 + a_2X_2 + \dots + a_nX_n) \\ &= a_1^2V(X_1) + a_2^2V(X_2) + \dots + a_n^2V(X_n) \end{aligned}$$

where $a_1, a_2, \dots, a_n$ are constants and $X_1, X_2, \dots, X_n$ are **independent** random variables.

1.5 Distribution of the sample mean

The results above can be used to derive the mean and variance of a sample mean.

(a) If $\bar{X}$ is the mean of a random sample of n observations of the random variable X , then

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)$$

where X_i denotes the i th observation in the sample, giving

$$\begin{aligned}
 E(\bar{X}) &= \frac{1}{n} (E[X_1] + E[X_2] + \dots + E[X_n]) \\
 &= \frac{1}{n} (\mu + \mu + \dots + \mu) \\
 &= \frac{1}{n} (n\mu) \\
 &= \mu
 \end{aligned}$$

(b) $V(\bar{X}) = E(\bar{X} - \mu)^2$ since the mean of $\bar{X}$ has just been shown to be μ .

$$\begin{aligned}
 V(\bar{X}) &= V\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) \\
 &= \frac{1}{n^2} (V(X_1) + V(X_2) + \dots + V(X_n)) \\
 &= \frac{1}{n^2} (\sigma^2 + \sigma^2 + \dots + \sigma^2) \\
 &= \frac{1}{n^2} (n\sigma^2) \\
 &= \frac{\sigma^2}{n}
 \end{aligned}$$

You have now shown that the sample mean $\bar{X}$ is distributed with mean μ and variance $\frac{\sigma^2}{n}$, a result which has been used frequently in *Statistics*.

1.6 Unbiased estimate of variance

The results above can be used to find an unbiased estimate of the variance.

First consider

$$\begin{aligned}
 \Sigma (X_i - \mu)^2 &= \Sigma ([X_i - \bar{X}] + [\bar{X} - \mu])^2 \\
 &= \Sigma (X_i - \bar{X})^2 + \Sigma (\bar{X} - \mu)^2 + 2\Sigma (X_i - \bar{X})(\bar{X} - \mu)
 \end{aligned}$$

$$\begin{aligned}
 \text{Now } \sum (X_i - \bar{X})(\bar{X} - \mu) &= (\bar{X} - \mu) \sum (X_i - \bar{X}) \\
 &= (\bar{X} - \mu) \left(\sum X_i - n\bar{X} \right) \\
 &= (\bar{X} - \mu) (\sum X_i - \sum X_i) \\
 &= 0
 \end{aligned}$$

Hence

$$\begin{aligned}
 \sum (X_i - \mu)^2 &= \sum (X_i - \bar{X})^2 + \sum (\bar{X} - \mu)^2 \\
 &= \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.
 \end{aligned}$$

Taking expectations of both sides gives

$$\begin{aligned}
 \sum E[(X_i - \mu)^2] &= E\left(\sum [X_i - \bar{X}]^2\right) + nE[(\bar{X} - \mu)^2] \\
 \Rightarrow n\sigma^2 &= E\left(\sum [X_i - \bar{X}]^2\right) + \frac{n\sigma^2}{n}.
 \end{aligned}$$

Hence

$$\begin{aligned}
 E\left(\sum [X_i - \bar{X}]^2\right) &= (n-1)\sigma^2 \\
 \text{or } \frac{E\left(\sum [X_i - \bar{X}]^2\right)}{(n-1)} &= \sigma^2.
 \end{aligned}$$

The expression on the left hand side was met in *Statistics* and is denoted $\hat{\sigma}^2$. You have just shown that $E(\hat{\sigma}^2) = \sigma^2$. $\hat{\sigma}^2$ is said to be an unbiased estimator of σ^2 and this is why it is generally used to estimate σ^2 . Note however that $\hat{\sigma}$ is not an unbiased estimator of σ .

Why is $\hat{\sigma}$ not an unbiased estimator of σ ?

1.7 Distribution of a linear combination of independent normal random variables

The results in Sections 1.3 and 1.4 cover the mean and variance of linear combinations of independent random variables. If the variables are all normally distributed then a linear combination of them will also be normally distributed. The proof of this is beyond the scope of this book.

To take a simple example, in the mass production of jars of jam the weight of jam put in each jar is a normally distributed random variable, X , with mean 456 g and standard deviation 4 g. The weight of the jar (including the lid) is an independent normally distributed random variable, Y , with mean 35 g and standard deviation 3 g.

The total weight of the jam plus the jar is $X + Y$.

The mean of $X + Y = E(X + Y) = E(X) + E(Y) = 456 + 35 = 491$ g

The variance of $X + Y = V(X) + V(Y) = 4^2 + 3^2 = 25$.

The standard deviation will be $\sqrt{25} = 5$ g.

Since the two variables are independently normally distributed the distribution of the total weight will also be normal.

A child opens a new jar of jam and takes a spoonful out. The weight of jam in the spoon is a random variable, Z , with mean 22 g standard deviation 2 g. Z is independent of X .

The weight of jam remaining in the jar is $X - Z$.

The mean is $E(X) - E(Z) = 456 - 22 = 434$ g.

The variance of $X - Z$ is $V(X) + V(Z) = 4^2 + 2^2 = 20$.

The standard deviation is $\sqrt{20} = 4.47$ g.

Since X and Z are independently normally distributed the distribution of $X - Z$ is also normal.

Example

A machine produces rubber balls whose diameters are normally distributed with mean 5.50 cm and standard deviation 0.08 cm.

- (a) What proportion of balls will have diameters
- (i) less than 5.60 cm,
 - (ii) between 5.34 and 5.44 cm?

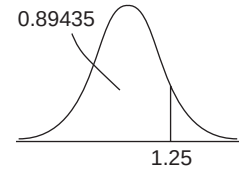
- (b) The balls are packed in cylindrical tubes whose internal diameters are normally distributed with mean 5.70 cm and standard deviation 0.12 cm. If a ball, selected at random, is placed in a tube, selected at random, what is the distribution of the clearance? (The clearance is the internal diameter of the tube minus the diameter of the ball.) What is the probability that the clearance is between 0.05 cm and 0.25 cm?

(AEB)

Solution

$$(a) \quad (i) \quad z = \frac{(5.60 - 5.50)}{0.08} = 1.25.$$

Using Normal tables, the probability of being less than 5.60 cm is 0.894.

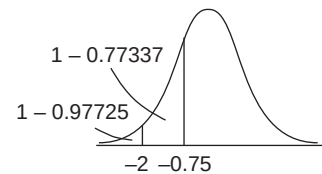


$$(ii) \quad z_1 = \frac{(5.34 - 5.50)}{0.08} = -2.0$$

$$z_2 = \frac{(5.44 - 5.50)}{0.08} = -0.75.$$

Hence the probability of being between 5.34 cm and 5.44 cm is given by

$$0.97725 - 0.77337 = 0.204.$$



- (b) If X is the diameter of the ball and Y is the diameter of the tube the clearance is $Y - X$. This will be normally distributed with

$$\text{mean} = 5.70 - 5.50 = 0.20$$

$$\text{variance} = 0.08^2 + 0.12^2 = 0.0208$$

$$\text{standard deviation} = 0.1442.$$

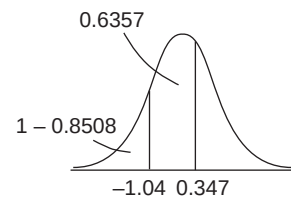
To calculate the probability that the clearance is between 0.05 cm and 0.25 cm you first find

$$z_1 = \frac{(0.05 - 0.20)}{0.1442} = -1.040$$

$$z_2 = \frac{(0.25 - 0.20)}{0.1442} = 0.347$$

so that probability of the clearance between 0.05 cm and 0.25 cm is given by

$$0.6357 - (1 - 0.8508) = 0.4865.$$



Note: Interpolation has been used in reading the normal tables, but the effect on the final answer is small.

Example

A baker makes digestive biscuits whose masses are normally distributed with mean 24.0 g and standard deviation 1.9 g. The biscuits are packed by hand into packets of 25.

- (i) Assuming the biscuits included in each packet are a random sample from the population, what is the distribution of the total mass of biscuit in a packet and what is the probability that it lies between 598 g and 606 g?
- (ii) Ten packets of biscuits are placed in a box. What is the probability that the total mass of biscuit in the box lies between 6010 g and 6060 g?
- (iii) A new packer was including 26 biscuits in each packet. What is the probability that a packet selected at random from those containing 25 biscuits would contain a greater mass of biscuits than a packet selected at random from those containing 26 biscuits? (AEB)

Solution

- (i) Since the sample is random the masses of the 25 biscuits included in the packet will be independent of each other. The distribution of the total mass will therefore be normal.

The mean mass of the biscuits in the packet will be
 $24 + 24 + \dots + 24 = 25 \times 24 = 600$ g.

The variance of the total mass will be
 $1.9^2 + 1.9^2 + \dots + 1.9^2 = 25 \times 1.9^2 = 90.25$,
 giving a standard deviation = 9.5 g.

The distribution is $N(600, 9.5^2)$.

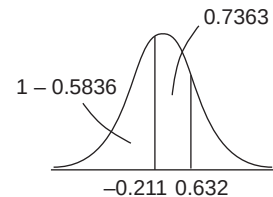
To find the probability the total mass lies between 598 g and 606 g, note that

$$z_1 = \frac{(598 - 600)}{9.5} = -0.211$$

$$z_2 = \frac{(606 - 600)}{9.5} = 0.632.$$

Probability between 598 g and 606 g

$$= 0.7363 - (1 - 0.5836) = 0.320.$$



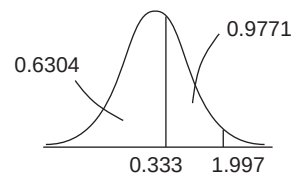
- (ii) Ten packets of biscuits are placed in a box. Provided their masses are independent the total mass will be normally distributed with mean $10 \times 600 = 6000$ g.

The variance will be $10 \times 90.25 = 902.5$ and the standard deviation $\sqrt{902.5} = 30.04$ g. To calculate the probability between 6010 g and 6060 g,

$$z_1 = \frac{(6010 - 6000)}{30.04} = 0.333$$

$$z_2 = \frac{(6060 - 6000)}{30.04} = 1.997.$$

Probability of being between 6010 g and 6060 g
 $= 0.9771 - 0.6304 = 0.347.$



- (iii) If X is the mass of biscuit in a packet containing 25 biscuits then we already know that X is $N(600, 90.25)$. If Y is the mass of biscuit in a packet containing 26 biscuits then Y is normally distributed with mean $26 \times 24 = 624$ g and variance $26 \times 1.9^2 = 93.86$.

The difference $Y - X$ is normally distributed with

$$\text{mean} = 624 - 600 = 24 \text{ g,}$$

$$\text{variance} = 90.25 + 93.86 = 184.11,$$

$$\text{standard deviation} = \sqrt{184.11} = 13.57.$$

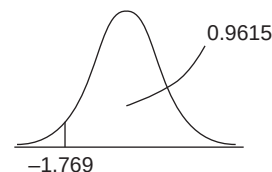
A packet containing 25 biscuits will weigh more than a packet containing 26 biscuits if $Y - X$ is negative, i.e. $Y - X < 0$.

Thus

$$z = \frac{(0 - 24)}{13.57} = -1.769.$$

This gives the probability of this occurring as

$$1 - 0.9615 = 0.0385.$$



Exercise 1B

- A dispenser discharges an amount of soft drink which is normally distributed with mean 475 ml and standard deviation 20 ml.
 - What is the distribution of the total amount in two independent drinks?
 - If the capacity of the cups into which the drink is dispensed is normally distributed with mean 500 ml and standard deviation 30 ml, what is the distribution of the difference between the capacity of a cup and the amount dispensed? Assume the two are independent.
- The weights of pieces of home made fudge are normally distributed with mean 34 g and standard deviation 5 g.
 - A bag contains 15 pieces of fudge chosen at random. What is the distribution of the total weight of fudge in the bag? What is the probability that the total weight is between 490 g and 540 g?
 - What is the probability that the total weight of fudge in a bag containing 15 pieces exceeds that in another bag containing 16 pieces? (AEB)

3. Audrey is a regular customer of Toto's taxis. When she rings from home the time, X , a taxi takes to arrive is normally distributed with mean 19 minutes and standard deviation 3 minutes.

- (a) (i) What is the probability of her having to wait less than 15 minutes for a taxi?
(ii) What waiting time will be exceeded with probability 0.1?

Audrey decides to try Blue Star taxis. The standard deviation of her waiting time, Y , is 7 minutes and the probability of Y exceeding 8 minutes is 0.97725.

- (b) Find the mean of Y , assuming a normal distribution.
(c) What is the distribution of T where $T = X - Y$? (X and Y may be assumed independent.) If both firms were rung at the same time, what is the probability that Toto would arrive first?
(d) In order to catch a train Audrey needs a taxi within 10 minutes. Which firm would you advise her to ring? Explain your answer.

(AEB)

1.8 The exponential distribution

If cars passing a point on a motorway follow a Poisson distribution, the interval of time between successive cars passing follows an exponential distribution.

The exponential distribution has probability density function

$$f(x) = \begin{cases} me^{-mx} & x > 0, \text{ where } m \text{ is constant} \\ 0 & \text{otherwise} \end{cases}$$

The mean of this exponential distribution is

$$\int_0^{\infty} mxe^{-mx} dx = \left[-xe^{-mx} \right]_0^{\infty} - \int_0^{\infty} -e^{-mx} dx = \left[-\frac{e^{-mx}}{m} \right]_0^{\infty} = \frac{1}{m}$$

$$E(X^2) = \int_0^{\infty} mx^2e^{-mx} dx = \left[-x^2e^{-mx} \right]_0^{\infty} - \int_0^{\infty} -2xe^{-mx} dx = \frac{2}{m^2}$$

since we have already shown that

$$\int_0^{\infty} mxe^{-mx} dx = \frac{1}{m}.$$

Thus

$$V(X) = \frac{2}{m^2} - \frac{1}{m^2} = \frac{1}{m^2}.$$

Hence the standard deviation is $\frac{1}{m}$.

Note: the AEB syllabus requires polynomial integration only and these proofs will not be tested.

1.9 Cumulative distribution function

The probability that an observed value from the negative exponential distribution is less than x is given by

$$P(X < x) = \int_0^x me^{-mx} dx = \left[-e^{-mx} \right]_0^x = 1 - e^{-mx}$$

This expression is known as the **cumulative distribution function**.

It is denoted $F(x)$.

For this distribution

$$F(x) = 1 - e^{-mx}$$

The cumulative distribution function may be used as an alternative to integration for evaluating probabilities. If a and b are two constants and $a > b$, the probability that X takes a value between a and b is

$$P(X < a) - P(X < b) = F(a) - F(b)$$

For example, the interval, X seconds, between cars passing a point on a motorway follows an exponential distribution with probability density function

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

(This distribution may alternatively be described as an exponential distribution with parameter 2.)

The probability that the next interval is between 1 and 2 seconds is

$$F(2) - F(1) = 1 - e^{-4} - (1 - e^{-2}) = 0.9817 - 0.8647 = 0.117.$$

You may check this result by evaluating

$$\int_1^2 2e^{-2x} dx.$$

The probability that the next interval is longer than 3 seconds is

$$F(\infty) - F(3) = 1 - (1 - e^{-6}) = 0.0025.$$

The probability that the next interval is less than 1.5 seconds is

$$F(1.5) - F(0) = (1 - e^{-3}) - (1 - e^0) = 0.9502.$$

This exponential distribution will have a mean of $\frac{1}{2} = 0.5$ seconds.

Therefore the average number of cars passing the point per second is $\frac{1}{0.5} = 2$.

In general the intervals between successive events from a Poisson distribution with mean m are distributed according to the exponential distribution with parameter m .

1.10 Conditional probability and the exponential distribution

Perhaps the most interesting feature of the exponential distribution is that it is 'memory less'. Using the cumulative distribution for the exponential distribution with parameter m , the probability that the next interval will last longer than a seconds is

$$F(\infty) - F(a) = 1 - (1 - e^{-ma}) = e^{-ma}.$$

The probability that the interval lasts longer than $a + b$ seconds is similarly $e^{-m(a+b)}$.

The probability that an interval lasts longer than $a + b$, given that it has already lasted a seconds, may be written

$$P(X > a + b \mid X > a).$$

Using the laws of probability

$$P[(X > a + b) \cap (X > a)] = P(X > a) \times P(X > a + b \mid X > a).$$

If an interval lasts longer than $a + b$ seconds it must have lasted longer than a seconds. Therefore

$$P[(X > a + b) \cap (X > a)] = P(X > a + b) = e^{-m(a+b)}$$

Therefore $e^{-m(a+b)} = e^{-ma} P(X > a + b \mid X > a)$.

$$\text{i.e.} \quad P(X > a + b \mid X > a) = e^{-mb} = P(X > b).$$

That is, the probability of the next event occurring within b seconds is the same whether an event has just occurred or whether the last event occurred a seconds ago.

Why is a distribution with this probability described as 'memory less'?

In the earlier example the probability that the interval between cars passing a point on a motorway would be longer than 3 seconds was calculated. The calculation would have been identical if the probability required had been that no car will pass in the next 3 seconds.

Exercise 1C

1. The interval, T minutes, between successive telephone calls to a school office follows an exponential distribution with parameter 0.2. Find the probability that the interval between the next two telephone calls will be

- (a) between 3 and 6 minutes,
- (b) between 2 and 7 minutes,
- (c) longer than 8 minutes,
- (d) less than 10 minutes.

2. A factory worker is employed to watch a monitor and give a warning when the monitor signals that action is needed to adjust the process. The interval, X hours, between successive signals follows an exponential distribution with parameter 0.08.

What is the probability that the interval between the next two signals is

- (a) between 10 and 20 hours,
- (b) between 6 and 16 hours
- (c) longer than 50 hours,
- (d) less than 2 hours?

The factory worker decides to read a newspaper for a few minutes instead of watching the monitor. How long can she read for if the probability of missing a signal is to be less than 0.01?

3. The lives of electric light bulbs, T hours, follows an exponential distribution with probability distribution function

$$f(t) = \begin{cases} 0.004e^{-0.004t} & 0 < t \\ 0 & \text{otherwise} \end{cases}$$

What is the probability that a bulb will last

- (a) between 200 and 300 hours,
- (b) between 250 and 350 hours,
- (c) more than 400 hours
- (d) less than 100 hours?

What is the mean and standard deviation of T ?

What is the mean and standard deviation of the mean life of 225 bulbs?

Find, approximately, the probability that the mean life of 225 bulbs will be less than 240 hours.

1.11 Miscellaneous Exercises

1. A clothing factory uses rolls of cloth for making suits. The length, in cm, of cloth wasted (because it is too short to use) at the end of each roll may be regarded as a random variable, X , with probability density function

$$f(x) = \begin{cases} \frac{1}{a} & 0 < x < a \\ 0 & \text{otherwise} \end{cases}$$

- (a) Derive the mean, μ , and variance, σ^2 , in terms of a .

- (b) Write down the mean and variance of $\bar{X}$, the mean of a sample of size n from the distribution.

- (c) The median, Y , of a random sample of size 3 from the distribution has a probability density function

$$g(y) = \begin{cases} \frac{6y}{a^2} - \frac{6y^2}{a^3} & 0 < y < a \\ 0 & \text{otherwise} \end{cases}$$

Find the mean and variance of Y . (AEB)

2. The Sunset Times has an average of 1 typing error per 500 words. The distribution of X , the number of words between successive typing errors, has probability density function

$$f(x) = \begin{cases} 0.002e^{-0.002x} & 0 < x \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that the number of words between successive errors is

- (a) between 400 and 600,
- (b) between 200 and 800,
- (c) greater than 1000,
- (d) less than 100.

What is the probability that a reader starting at a random point in the paper will read more than 750 words before reaching an error?

The editor, Mr B. Pad, always reads the first paragraph before the paper is published. How many words can the first paragraph contain if the probability of it containing an error is to be not more than 0.05?

3. The thickness of a certain grade of hardboard stocked by a DIY shop is normally distributed with mean 7.3 mm and standard deviation 0.5 mm.
- (a) What proportion of sheets will be between 7 mm and 8 mm thick?
 - (b) Sheets of the same grade bought from a second shop contain 9.1% over 8 mm thick and 2.3% less than 7 mm thick. Assuming that the thickness is normally distributed, find its mean and standard deviation correct to the nearest tenth of a mm.
 - (c) What is the distribution of $Y - X$ where X and Y are the thickness of pieces of hardboard selected at random from the first and second shops respectively? Find the probability that X exceeds Y .
 - (d) It is possible to buy batches of hardboard from the first shop with any required mean and with the standard deviation remaining 0.5 mm. What value of the mean should be chosen
 - (i) to minimise the proportion of sheets outside the range 7 mm to 8 mm (no proof required),
 - (ii) so that 0.1% of sheets are less than 7 mm thick?
4. A certain brand of beans is sold in tins, the tins being filled and sealed by a machine. The mass of beans in each tin is normally distributed with mean 425 g and standard deviation 25 g and the mass of the tin is normally distributed with mean 90 g and standard deviation 10 g.

- (a) Find the probability that the total mass of the sealed tin and its beans
 - (i) exceeds 550 g,
 - (ii) lies between 466 g and 575 g.
- (b) Calculate an interval within which approximately 90% of the masses of the filled tins will lie.

The tins are packed in boxes of 24, the mass of the box being normally distributed with mean 500 g and standard deviation 30 g.
- (c) Find the probability that a full box weighs less than 12.75 kg.

5. Lin Ying belongs to an athletic club. In 800 m races her times are normally distributed with mean 128 seconds and standard deviation 4 seconds.
- (a) What is the probability of her time in an 800 being between 120 and 130 seconds?
 - (b) What time will she beat in 70% of her races?
- Julie belongs to the same club. In 800 m races 85% of her times are less than 140 seconds and 70% are less than 135 seconds. Her times are normally distributed.
- (c) Find the mean and standard deviation of Julie's times, each correct to two significant figures.
 - (d) What is the probability that in an 800 m race Lin Ying will beat Julie?
 - (e) The club has to choose one of these two athletes to enter a major competition. In order to qualify for the final rounds it is necessary to achieve a time of 114 seconds or less in the heats. Which athlete should be chosen and why?
6. The contents of bags of oats are normally distributed with mean 3.05 kg, standard deviation 0.08 kg.
- (a) What proportion of bags contain less than 3.11 kg?
 - (b) What proportion of bags contain between 3.00 and 3.15 kg?
 - (c) What weight is exceeded by the contents of 99.9% of the bags?
 - (d) If 6 bags are selected at random, what is the probability that the mean weight of the contents will be between 3.00 and 3.15 kg?

The weight of the bags when empty is normally distributed with mean 0.12 kg, standard deviation 0.02 kg. Full bags are packed into boxes each of which holds 6 bags.

- (e) What is the distribution of the weight in a box, i.e. 6 bags together with their contents?
Assume that the weight of all bags and contents in a box are independent of each other.
 - (f) Within what limits will the weight in a box lie with probability 0.9?
7. Fertilizer is packed, by a machine, into bags of nominal mass 12 kg. The random mass of each bag may be regarded as a normally distributed random variable with mean 12.05 kg, standard deviation 0.20 kg.
- (a) What is the probability that the mass of a bag exceeds 12 kg?
 - (b) What mass is exceeded by exactly 95% of the bags?
- A farmer buys 20 bags at a time.
- (c) What is the probability that their mean mass will exceed 12 kg?
 - (d) What is the distribution of the total mass of the 20 bags and what is the probability that it lies between 239.5 kg and 240.5 kg?

The mass of the bags packed by a second machine may be regarded as a normally distributed random variable with mean 12.05 kg, standard deviation 0.05 kg. If the farmer's 20 bags are made up of n from the first machine and the rest from the second machine, what is the largest possible value of n which gives a probability of at least 0.95 of the total mass of the bags exceeding 240 kg?

(AEB)