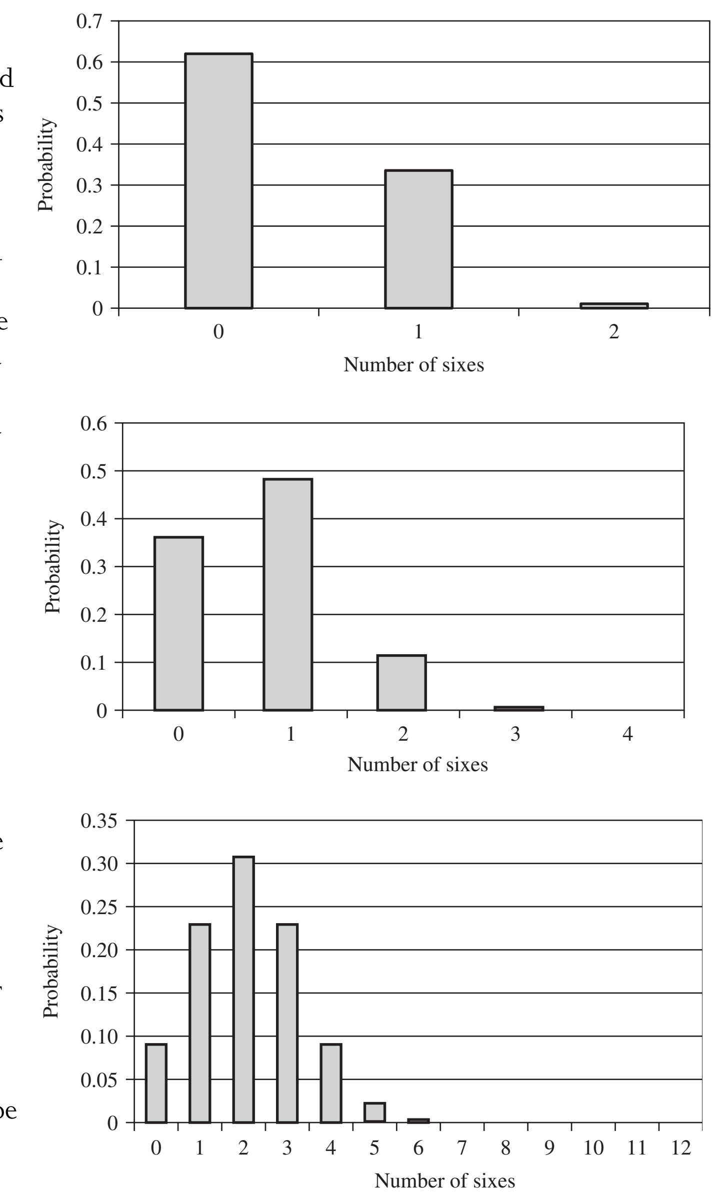
22 Continuous Probability Distributions

In Chapter 21, we saw that the binomial distribution could be used to solve problems such as "If an unbiased cubical die is thrown 50 times, what is the probability of throwing a six more than 25 times?" To solve this problem, we compute the probability of throwing a six 25 times then the probability of throwing a six 26 times, 27 times, etc., which before the introduction of calculators would have taken a very long time to compute. Abraham de Moivre, who we met in Chapter 17, noted that when the number of events (throwing a die in this case) increased to a large enough number, then the shape of the binomial distribution approached a very smooth curve.

Binomial distributions for 2, 4 and 12 throws



De Moivre realised that if he was able to find a mathematical expression for this curve, he would be able to find probabilities much more easily. This curve is what we now call a normal curve and the distribution associated with it is introduced in this chapter. It is shown here approximating the binomial distribution for 12 coin flips.

The normal distribution is of great importance because many natural phenomena are at least approximately normally distributed. One of the earliest applications of the normal distribution was connected to error analysis in astronomical observations. Galileo in the 17th century hypothesized several distributions for these errors, but it was not until two centuries later that it was discovered that they followed a normal distribution. The normal distribution had also been discovered by Laplace in 1778 when he derived the extremely important central limit theorem. Laplace showed that for any distribution, provided that the sample size is large, the distribution of the means of repeated samples from the distribution would be approximately normal, and that the larger the sample size, the closer the distribution would be to a normal distribution.

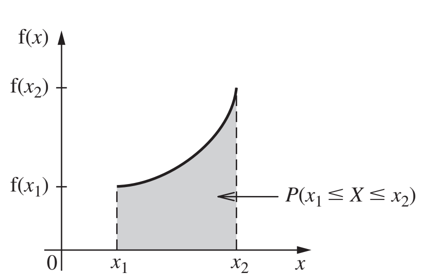
22.1 Continuous random variables

In Chapter 19 we discussed the difference between discrete and continuous data and in Chapter 21 we met discrete data where $\sum_{\text{all } x} \mathrm{P}(X = x) = 1$. In this chapter we consider continuous data. To find the probability that the height of a man is 1.85 metres, correct to 3 significant figures, we need to find $\mathrm{P}(1.845 \leq H < 1.855)$. Hence for continuous data we construct ranges of values for the variable and find the probabilities for these different ranges.

For a discrete random variable, a table of probabilities is normally given. For a continuous random variable, a probability density function is normally used instead. In Chapter 21, we met probability density functions where the variable was discrete. When the variable is continuous the function f(x) can be integrated over a particular range of values to give the probability that the random variable X lies in that particular range. Hence for a continuous random variable valid over the range $a \leq x \leq b$ we can say that $\int_a^b \mathrm{f}(x)\, dx = 1$. This is analogous to $\sum_{x=a}^{x=b} \mathrm{P}(X = x) = 1$ for discrete data and also the idea of replacing sigma notation with integral notation when finding the area under the curve, as seen in Chapter 14.

The area under the curve represents the probability.

Thus, if $a \leq x_1 \leq x_2 \leq b$ then $\mathrm{P}(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \mathrm{f}(x)\, \mathrm{d}x$ as shown in the diagram.



Since many of the calculations involve definite integration, if the questions were to appear on a calculator paper, the calculations could be performed on a calculator.

Example

Consider the function $\mathrm{f}(x) = \frac{3}{4}, \frac{1}{3} \leq x \leq \frac{5}{3}$, which is being used as a probability density function for a continuous random variable X.

a Show that f(x) is a valid probability density function.

b Find the probability that X lies in the range $\frac{3}{4}$ to $\frac{5}{4}$.

c Show this result graphically.

a f(x) is a valid probability density function if $\int_{\frac{1}{3}}^{\frac{5}{3}} \frac{3}{4}\mathrm{d}x = 1$.

$$\int_{\frac{1}{3}}^{\frac{5}{3}} \frac{3}{4}\mathrm{d}x = \left[\frac{3x}{4}\right]_{\frac{1}{3}}^{\frac{5}{3}}$$

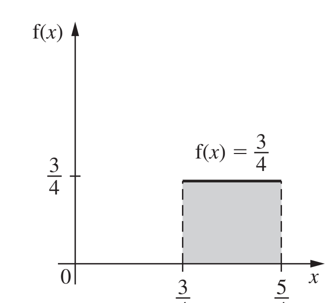
$$= \frac{5}{4} - \frac{1}{4} = 1$$

Hence f(x) can be used as a probability density function for a continuous random variable.

b $\mathrm{P}\left(\frac{3}{4} \leq X \leq \frac{5}{4}\right) = \int_{\frac{3}{4}}^{\frac{5}{4}} \frac{3}{4}\mathrm{d}x$

$$= \left[\frac{3x}{4}\right]_{\frac{3}{4}}^{\frac{5}{4}} = \frac{15}{16} - \frac{9}{16} = \frac{6}{16} = \frac{3}{8}$$

c



In this case we did not have to use integration as the area under the curve is given by the area of a rectangle.

Example

The continuous random variable X has probability density function f(x) where $\mathrm{f}(x) = \frac{3}{26}(x - 1)^2, 2 \leq x \leq k$.

a Find the value of the constant k.

b Sketch $y = \mathrm{f}(x)$.

c Find $\mathrm{P}(2.5 \leq X \leq 3.5)$ and show this on a diagram.

d Find $\mathrm{P}(X \geq 2.5)$.

a $\displaystyle\int_2^k \frac{3}{26}(x-1)^2\,dx = 1$

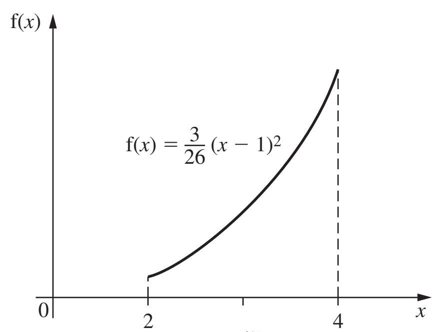
$\Rightarrow \left[\frac{(x-1)^3}{26}\right]_2^k = 1$

$\Rightarrow \frac{(k-1)^3}{26} - \frac{1}{26} = 1$

$\Rightarrow \frac{(k-1)^3}{26} = \frac{27}{26}$

$\Rightarrow k - 1 = 3 \Rightarrow k = 4$

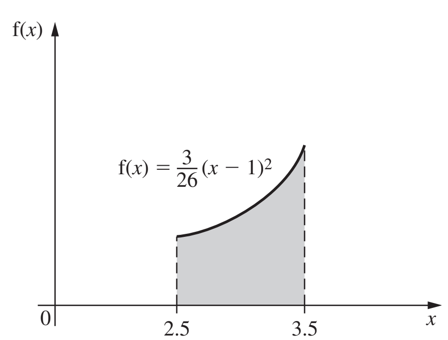
b



The graph has a domain of $2 \le x \le 4$.

c $P(2.5 \le X \le 3.5) = \displaystyle\int_{2.5}^{3.5} \frac{3}{26}(x-1)^2\,dx$

$= \left[\frac{(x-1)^3}{26}\right]_{2.5}^{3.5} = \frac{125}{208} - \frac{27}{208} = \frac{98}{208} = \frac{49}{104}$



d $P(X \ge 2.5) = \displaystyle\int_{2.5}^{4} \frac{3}{26}(x-1)^2\,dx$

$= \left[\frac{(x-1)^3}{26}\right]_{2.5}^{4} = \frac{27}{26} - \frac{27}{208} = \frac{189}{208}$

Sometimes the probability density function for a continuous random variable can use two or more different functions.

This is similar to the piecewise functions met in Chapter 3.

However, it works in exactly the same way.

Example

The continuous random variable X has probability density function

$$f(x) = \begin{cases} k(4-x)^2 & 0 \le x \le 2 \\ 4k & 2 < x \le \frac{8}{3} \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Find the value of the constant k.

b Sketch $y = f(x)$.

c Find $P(1 \le X \le 2.5)$.

d Find $P(X \ge 1)$.

a $\displaystyle\int_0^2 k(4-x)^2\,dx + \int_2^{\frac{8}{3}} 4k\,dx = 1$

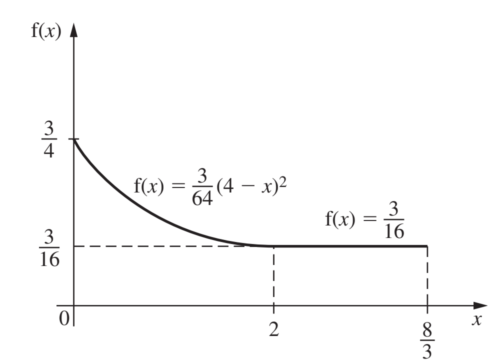
$\Rightarrow \left[\frac{k(4-x)^3}{-3}\right]_0^2 + [4kx]_2^{\frac{8}{3}} = 1$

$\Rightarrow -\frac{8k}{3} + \frac{64k}{3} + \frac{32k}{3} - 8k = 0$

$\Rightarrow \frac{64k}{3} = 1$

$\Rightarrow k = \frac{3}{64}$

b



c As the area spans the two distributions, we integrate over the relevant domains.

$P(1 \le X \le 2.5) = \displaystyle\int_1^2 \frac{3}{64}(4-x)^2\,dx + \int_2^{2.5} \frac{3}{16}dx$

$= \left[-\frac{1}{64}(4-x)^3\right]_1^2 + \left[\frac{3x}{16}\right]_2^{2.5}$

$= -\frac{1}{8} + \frac{27}{64} + \frac{15}{32} - \frac{3}{8} = \frac{25}{64}$

d $P(X \ge 1) = \displaystyle\int_1^2 \frac{3}{64}(4-x)^2\,dx + \int_2^{\frac{8}{3}} \frac{3}{16}dx$

$$= \left[-\frac{1}{64}(4-x)^3\right]_1^2 + \left[\frac{3x}{16}\right]_2^{\frac{8}{3}}$$

$$= -\frac{1}{8} + \frac{27}{64} + \frac{1}{2} - \frac{3}{8} = \frac{27}{64}$$

These integrals could be done directly on a calculator.

Exercise 1

1 A continuous probability density function is defined as

$$f(x) = \begin{cases} k - \dfrac{x}{4} & 1 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Find the value of k. **b** Sketch $y = f(x)$.

c Find $P(1.5 \le X \le 2.5)$ and show this on a sketch. **d** Find $P(X \le 2.5)$.

2 Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} \dfrac{x}{2} - 1 & 2 \le x \le c \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

a Find the value of c.

b Sketch $y = f(x)$.

c Find $P(2.5 \le X \le 3)$ and show this on a sketch.

d Find $P(4.5 \le X \le 5.2)$.

3 A continuous random variable X has probability density function

$$f(x) = \begin{cases} k\cos x & 0 \le x \le \dfrac{\pi}{4} \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Find the value of k. **b** Sketch $y = f(x)$.

c Find $P\left(0 \le X \le \frac{\pi}{6}\right)$ and show this on a sketch. **d** Find $P\left(X \ge \frac{\pi}{12}\right)$.

4 The probability density function $f(x)$ of a continuous random variable X is defined by

$$f(x) = \begin{cases} \dfrac{1}{2}x(4 - x^2) & k \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Find the value of k. **b** Sketch $y = f(x)$.

c Find $P(1.1 \le X \le 1.3)$ and show this on a sketch. **d** Find $P(X \le 1.5)$.

5 The time taken for a worker to perform a particular task, t minutes, has probability density function

$$f(t) = \begin{cases} kt^2 & 0 \le t \le 5 \\ 0.4k(2 + t) & 5 < t \le 15 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Find the value of k. **b** Sketch $y = f(t)$.

c Find $P(4 \le X \le 11)$ and show this on a sketch. **d** Find $P(X \ge 9)$.

22.2 Using continuous probability density functions

Expectation

For a discrete random variable $E(X) = \sum_{\text{all } x} x \cdot P(X = x)$.

Hence $E(X) = \int_a^b x\,f(x)\,dx$ for a continuous random variable valid over the range $a \le x \le b$.

If the probability density function is symmetrical then $E(X)$ is the value of the line of symmetry.

This is similar to the result for discrete data.

For continuous data we are often dealing with a population, so $E(X)$ is denoted as μ. For discrete data we are often dealing with a sample, so $E(X)$ is denoted as $\bar{x}$. In both cases $E(X)$ is referred to as the mean of X.

Example

If X is a continuous random variable with probability density function $f(x) = \frac{1}{9}x^2$, $0 \le x \le 3$, find $E(X)$.

$$E(X) = \int_0^3 x\left(\frac{1}{9}x^2\right)dx$$

$$= \frac{1}{9}\int_0^3 x^3\,dx$$

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

$$= \frac{81}{36} = \frac{9}{4}$$

Example

The continuous random variable X has probability density function
$f(x) = k(1 - x)(x - 5),\ 1 \le x \le 5$.

a Find the value of the constant k.
b Sketch $y = f(x)$.
c Find $E(X)$.
d Find $P(1.5 \le X \le 3.5)$.

a

$$\int_1^5 k(1-x)(x-5)\,dx = 1$$

$$\Rightarrow k\int_1^5 (-x^2 + 6x - 5)\,dx = 1$$

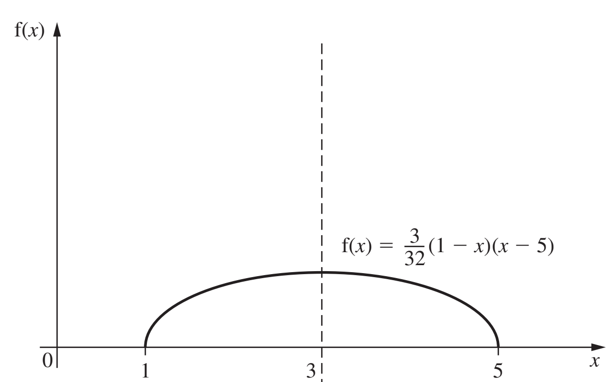
$$\Rightarrow k\left[-\frac{x^3}{3} + 3x^2 - 5x\right]_1^5 = 1$$

$$\Rightarrow k\left[\left(-\frac{125}{3} + 75 - 25\right) - \left(-\frac{1}{3} + 3 - 5\right)\right] = 1$$

$$\Rightarrow \frac{32}{3}k = 1$$

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

b



c Since the distribution is symmetrical, $E(X) = 3$ from the above diagram.

d $$P(1.5 \le X \le 3.5) = \int_{1.5}^{3.5} \frac{3}{32}(1-x)(x-5)\,dx$$

$$= \frac{3}{32}\int_{1.5}^{3.5} (-x^2 + 6x - 5)\,dx$$

$$= \frac{3}{32}\left[-\frac{x^3}{3} + 3x^2 - 5x\right]_{1.5}^{3.5}$$

$$= \frac{3}{32}\left[\left(-\frac{343}{24} + \frac{147}{4} - \frac{35}{2}\right) - \left(-\frac{9}{8} + \frac{27}{4} - \frac{15}{2}\right)\right]$$

$$= \frac{41}{64}$$

Example

The time taken in hours for a particular insect to digest food is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} k(x-1)^2 & 1 \le x \le 2 \\ k(8-x) & 2 < x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Find

a the value of the constant k
b the mean time taken
c the probability that it takes an insect between 1.5 and 3 hours to digest its food
d the probability that two randomly chosen insects each take between 1.5 and 3 hours to digest their food.

a $$\int_1^2 k(x-1)^2\,dx + \int_2^4 k(8-x)\,dx = 1$$

$$\Rightarrow k\left[\frac{(x-1)^3}{3}\right]_1^2 + k\left[8x - \frac{x^2}{2}\right]_2^4 = 1$$

$$\Rightarrow k\left[\left(\frac{1}{3} - 0\right) + (24 - 14)\right] = 1$$

$$\Rightarrow k = \frac{3}{31}$$

b $$E(X) = \int_1^2 kx(x-1)^2\,dx + \int_2^4 kx(8-x)\,dx$$

$$= \frac{3}{31}\left[\int_1^2 (x^3 - 2x^2 + x)\,dx + \int_2^4 (8x - x^2)\,dx\right]$$

$$= \frac{3}{31}\left\{\left[\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}\right]_1^2 + \left[4x^2 - \frac{x^3}{3}\right]_2^4\right\}$$

$$= \frac{3}{31}\left[\left(4 - \frac{16}{3} + 2\right) - \left(\frac{1}{4} - \frac{2}{3} + \frac{1}{2}\right) + \left(64 - \frac{64}{3}\right) - \left(16 - \frac{8}{3}\right)\right]$$

$= 2.90$ hours

c $$P(1.5 \le X \le 3) = \int_{1.5}^2 k(x-1)^2\,dx + \int_2^3 k(8-x)\,dx$$

$$= \frac{3}{31}\left\{\left[\frac{(x-1)^3}{3}\right]_{1.5}^2 + \left[8x - \frac{x^2}{2}\right]_2^3\right\}$$

$$= \frac{3}{31}\left[\left(\frac{1}{3} - \frac{1}{24}\right) + (22 - 14)\right]$$

$= 0.802$

d P(two randomly chosen insects each take between 1.5 and 3 hours to digest their food) $= 0.802^2 = 0.644$

For a continuous random variable valid over the range $a \le x \le b$

$$E[g(X)] = \int_a^b g(x)\, f(x)\, dx$$

This is similar to the result for discrete data.

where $g(x)$ is any function of the continuous random variable X and $f(x)$ is the probability density function.

Hence we have the result

$$E(X^2) = \int_a^b x^2 f(x)\, dx$$

Example

The continuous random variable X has probability density function $f(x)$ where $f(x) = \frac{1}{18}(6 - x),\ 0 \le x \le 6$. Find:

a $E(X)$
b $E(2X - 1)$
c $E(X^2)$

a $$E(X) = \int_0^6 \frac{1}{18} x(6 - x)\, dx$$
$$= \frac{1}{18}\int_0^6 (6x - x^2)\, dx$$
$$= \frac{1}{18}\left[3x^2 - \frac{x^3}{3}\right]_0^6$$
$$= \frac{1}{18}\left(108 - \frac{216}{3}\right) = 2$$

b $$E(2X - 1) = \int_0^6 \frac{1}{18}(2x - 1)(6 - x)\, dx$$
$$= \frac{1}{18}\int_0^6 (-2x^2 + 13x - 6)\, dx$$
$$= \frac{1}{18}\left[\frac{-2x^3}{3} + \frac{13x^2}{2} - 6x\right]_0^6$$
$$= \frac{1}{18}(-144 + 234 - 36) = 3$$

c $$E(X^2) = \int_0^6 \frac{1}{18} x^2(6 - x)\, dx$$

$$= \frac{1}{18}\int_0^6 (6x^2 - x^3)\, dx$$
$$= \frac{1}{18}\left[2x^3 - \frac{x^4}{4}\right]_0^6$$
$$= \frac{1}{18}(512 - 324) = 10.4$$

Variance

We are now in a position to calculate the variance. As with discrete data

$$\text{Var}(X) = E(X - \mu)^2$$
$$= E(X^2) - E^2(X)$$

Therefore for a continuous random variable with probability density function valid over the domain $a \le x \le b$

$$\text{Var}(X) = \int_a^b x^2 f(x)\, dx - \left(\int_a^b x f(x)\, dx\right)^2$$

The standard deviation of X is $\sigma = \sqrt{\text{Var}(X)}$.

Example

The continuous random variable X has probability density function $f(x)$ where $f(x) = \frac{2}{63}(1 - 2x)^2,\ 2 \le x \le \frac{7}{2}$. Find:

a $E(X)$
b $E(X^2)$
c $\text{Var}(X)$
d σ

a $$E(X) = \int_2^{\frac{7}{2}} \frac{2}{63} x(1 - 2x)^2\, dx$$
$$= \frac{2}{63}\int_2^{\frac{7}{2}} (x - 4x^2 + 4x^3)\, dx$$
$$= \frac{2}{63}\left[\frac{x^2}{2} - \frac{4x^3}{3} + x^4\right]_2^{\frac{7}{2}}$$
$$= \frac{2}{63}\left[\left(\frac{49}{8} - \frac{343}{6} + \frac{2401}{16}\right) - \left(2 - \frac{32}{3} + 16\right)\right]$$
$$= \frac{163}{56}$$

b $\mathrm{E}(X^2) = \displaystyle\int_2^{\frac{7}{2}} \frac{2}{63}x^2(1-2x)^2\,\mathrm{d}x$

$= \dfrac{2}{63}\displaystyle\int_2^{\frac{7}{2}} (x^2 - 4x^3 + 4x^4)\,\mathrm{d}x$

$= \dfrac{2}{63}\left[\dfrac{x^3}{3} - x^4 + \dfrac{4x^5}{5}\right]_2^{\frac{7}{2}}$

$= \dfrac{2}{63}\left[\left(\dfrac{343}{24} - \dfrac{2401}{16} + \dfrac{16\,807}{40}\right) - \left(\dfrac{8}{3} - 16 + \dfrac{128}{5}\right)\right]$

$= \dfrac{2433}{280}$

c $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$

$= \dfrac{2433}{280} - \left(\dfrac{163}{56}\right)^2 = 0.217$

d $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{0.217} = 0.466$

Example

A particular road has been altered so that the traffic has to keep to a lower speed and at one point in the road traffic can only go through one way at a time. At this point traffic in one direction will have to wait. The time in minutes that vehicles have to wait has probability density function

$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{2}\left(1 - \dfrac{x}{4}\right) & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

a Find the mean waiting time.
b Find the standard deviation of the waiting time.
c Find the probability that three cars out of the first six to arrive after 8.00 am in the morning have to wait more than 2 minutes.

a The mean waiting time is given by $\mathrm{E}(X)$.

$\mathrm{E}(X) = \displaystyle\int_0^4 \frac{1}{2}x\left(1 - \frac{x}{4}\right)\mathrm{d}x$

$= \displaystyle\int_0^4 \frac{1}{2}\left(x - \frac{x^2}{4}\right)\mathrm{d}x$

$= \dfrac{1}{2}\left[\dfrac{x^2}{2} - \dfrac{x^3}{12}\right]_0^4$

$= \dfrac{1}{2}\left(8 - \dfrac{16}{3}\right) = \dfrac{4}{3}$

b $\mathrm{E}(X^2) = \displaystyle\int_0^4 \frac{1}{2}x^2\left(1 - \frac{x}{4}\right)\mathrm{d}x$

$= \displaystyle\int_0^4 \frac{1}{2}\left(x^2 - \frac{x^3}{4}\right)\mathrm{d}x$

$= \dfrac{1}{2}\left[\dfrac{x^3}{3} - \dfrac{x^4}{16}\right]_0^4$

$= \dfrac{1}{2}\left(\dfrac{64}{3} - 16\right) = \dfrac{8}{3}$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X)$

$= \dfrac{8}{3} - \left(\dfrac{4}{3}\right)^2 = \dfrac{8}{9}$

Hence $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\dfrac{8}{9}} = 0.943$

c First we need to calculate the probability that a car has to wait more than 2 minutes.

$\mathrm{P}(X > 2) = \displaystyle\int_2^4 \frac{1}{2}\left(1 - \frac{x}{4}\right)\mathrm{d}x$

$= \dfrac{1}{2}\left[x - \dfrac{x^2}{8}\right]_2^4$

$= \dfrac{1}{2}\left[(4-2) - \left(2 - \dfrac{1}{2}\right)\right]$

$= \dfrac{1}{4}$

Since we are now considering six cars, this follows the binomial distribution $Y \sim \mathrm{Bin}\left(6, \dfrac{1}{4}\right)$.

We want $\mathrm{P}(Y = 3) = {}^6\mathrm{C}_3\left(\dfrac{1}{4}\right)^3\left(\dfrac{3}{4}\right)^3 = 0.132$.

Example

A continuous random variable has probability density function $\mathrm{f}(x)$ where

$$\mathrm{f}(x) = \begin{cases} \dfrac{3}{128}x^2 & 0 \le x \le 4 \\ \dfrac{1}{4} & 4 < x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

Calculate:
a $\mathrm{E}(X)$
b $\mathrm{Var}(X)$
c σ
d $\mathrm{P}(|X - \mu| < \sigma)$

a $E(X) = \int_0^4 \frac{3}{128}x^3\,dx + \int_4^6 \frac{1}{4}x\,dx$

$= \left[\frac{3x^4}{512}\right]_0^4 + \left[\frac{x^2}{8}\right]_4^6$

$= \frac{3}{2} + \frac{9}{2} - 2 = 4$

b $E(X^2) = \int_0^4 \frac{3}{128}x^4\,dx + \int_4^6 \frac{1}{4}x^2\,dx$

$= \left[\frac{3x^5}{640}\right]_0^4 + \left[\frac{x^3}{12}\right]_4^6$

$= \frac{24}{5} + 18 - \frac{16}{3} = \frac{262}{15}$

$\mathrm{Var}(X) = E(X^2) - E^2(X)$

$= \frac{262}{15} - (4)^2 = \frac{22}{15}$

c $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{22}{15}} = 1.21$

d $P(|X - \mu| < \sigma) = P(|X - 4| < 1.21)$

$= P(-1.21 < X - 4 < 1.21)$

$= P(2.79 < X < 5.21)$

$= \int_{2.79}^{4} \frac{3}{128}x^2\,dx + \int_4^{5.21} \frac{1}{4}dx$

$= \left[\frac{x^3}{128}\right]_{2.79}^4 + \left[\frac{x}{4}\right]_4^{5.31}$

$= 0.5 - 0.169 + 1.33 - 1 = 0.658$

The mode

Since the mode is the most likely value for X, it is found at the value of X for which $f(x)$ is greatest, in the given range of X. Provided the probability density function has a maximum point, it is possible to determine the mode by finding this point.

To find the mode we differentiate, but when we find the mean and the median we integrate.

Example

The continuous random variable X has probability density function $f(x)$ where $f(x) = \frac{3}{38}(3 + 2x)(3 - x)$, $1 \le x \le 3$. Find the mode.

To find the mode we differentiate the function and justify that it has a maximum value.

$f(x) = \frac{3}{38}(3 + 2x)(3 - x)$

$= \frac{3}{38}(-2x^2 + 3x + 9)$

$f'(x) = \frac{3}{38}(-4x + 3) = 0$ for a maximum or minimum point

$\Rightarrow x = \frac{3}{4}$

This is a maximum because $f''\left(\frac{3}{4}\right) = -\frac{6}{19}$.

Therefore the mode is $x = \frac{3}{4}$.

The median

Since the probability is given by the area under the curve, the median splits the area under the curve $y = f(x)$, $a \le x \le b$, into two halves. So if the median is m, then

$$\int_a^m f(x)\,dx = 0.5$$

Example

The continuous random variable X has probability density function $f(x)$ where $f(x) = \frac{3}{10}(x - 4)(1 - x)$, $1 \le x \le 3$. Find the median.

Let the median be m.

$$\int_1^m \frac{3}{10}(x - 4)(1 - x)\,dx = 0.5$$

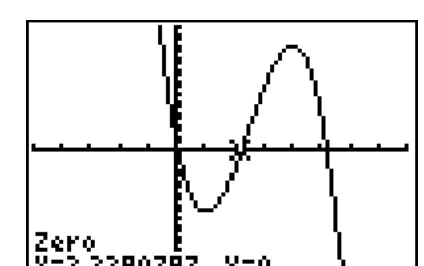
$$\Rightarrow \frac{3}{10}\int_1^m (-x^2 + 5x - 4)\,dx = 0.5$$

$$\Rightarrow \frac{3}{10}\left[-\frac{x^3}{3} + \frac{5x^2}{2} - 4x\right]_1^m = 0.5$$

$$\Rightarrow \frac{3}{10}\left[\left(-\frac{m^3}{3} + \frac{5m^2}{2} - 4m\right) - \left(-\frac{1}{3} + \frac{5}{2} - 4\right)\right] = 0.5$$

$$\Rightarrow -2m^3 + 15m^2 - 24m + 1 = 0$$

We now solve this on a calculator.



There are three solutions to this equation, but only one lies in the domain and hence $m = 2.24$.

This becomes a little more complicated if the probability density function is made up of more than one function, as we have to calculate in which domain the median lies.

Example

A continuous random variable X has a probability density function

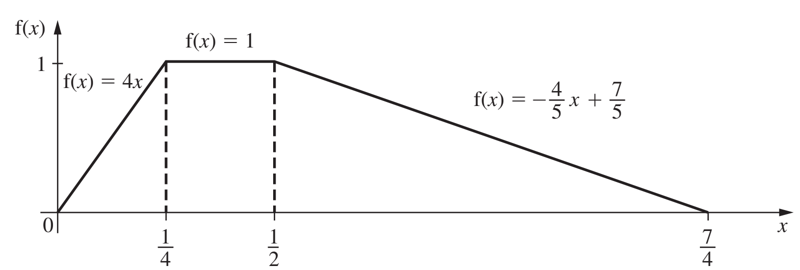
$$f(x) = \begin{cases} 4x & 0 \le x \le \frac{1}{4} \\ 1 & \frac{1}{4} \le x \le \frac{1}{2} \\ -\frac{4}{5}x + \frac{7}{5} & \frac{1}{2} \le x \le \frac{7}{4} \\ 0 & \text{otherwise} \end{cases}$$

a Sketch $y = f(x)$.

b Find the median m.

c Find $P\left(|X - m| > \frac{1}{2}\right)$.

a



b We now have to determine in which section of the function the median occurs. We will do this by integration, but it can be done using the areas of triangles and rectangles.

$$P\left(X \le \frac{1}{4}\right) = \int_0^{\frac{1}{4}} 4x \, dx$$

$$= [2x^2]_0^{\frac{1}{4}} = \frac{1}{8}$$

Since $\frac{1}{8} < 0.5$ the median does not lie in this region.

$$P\left(X \le \frac{1}{2}\right) = \int_0^{\frac{1}{4}} 4x \, dx + \int_{\frac{1}{4}}^{\frac{1}{2}} 1 \, dx$$

$$= [2x^2]_0^{\frac{1}{4}} + [x]_{\frac{1}{4}}^{\frac{1}{2}} = \frac{1}{8} + \frac{1}{2} - \frac{1}{4} = \frac{3}{8}$$

Since $\frac{3}{8} < 0.5$ the median does not lie in this region so it must be in the third region.

Hence $\int_0^{\frac{1}{4}} 4x \, dx + \int_{\frac{1}{4}}^{\frac{1}{2}} 1 \, dx + \int_{\frac{1}{2}}^{m} \left(-\frac{4}{5}x + \frac{7}{5}\right) dx = 0.5$

$$\Rightarrow \frac{3}{8} + \frac{1}{5}[-2x^2 + 7x]_{\frac{1}{2}}^{m} = 0.5$$

$$\Rightarrow \frac{3}{8} + \frac{1}{5}\left(-2m^2 + 7m + \frac{1}{2} - \frac{7}{2}\right) = 0.5$$

$$\Rightarrow 15 - 16m^2 + 56m - 24 = 20$$

$$\Rightarrow 16m^2 - 56m + 29 = 0$$

$$\Rightarrow m = 0.632 \text{ or } 2.86$$

Since $m = 2.86$ is not defined for the probability density function, $m = 0.632$.

c $P\left(|X - m| > \frac{1}{2}\right) = P\left(|X - 0.632| > \frac{1}{2}\right)$

$$= P\left(-\frac{1}{2} < X - 0.632 < \frac{1}{2}\right)$$

$$= P(0.132 < X < 1.13)$$

$$= \int_{0.132}^{\frac{1}{4}} 4x \, dx + \int_{\frac{1}{4}}^{\frac{1}{2}} 1 \, dx + \int_{\frac{1}{2}}^{1.13} \left(-\frac{4}{5}x + \frac{7}{5}\right) dx$$

$$= [2x^2]_{0.132}^{\frac{1}{4}} + [x]_{\frac{1}{4}}^{\frac{1}{2}} + \frac{1}{5}[-2x^2 + 7x]_{\frac{1}{2}}^{1.13}$$

$$= 0.125 - 0.0348 + 0.5 - 0.25 + 1.07 - 0.6 = 0.812$$

Exercise 2

1 A continuous random variable has probability density function

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

where k is a constant.

a Find the value of k. **b** Find $E(X)$. **c** Find $\text{Var}(X)$.

2 A continuous random variable has probability density function

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

where k is a constant.

Without using a calculator, find:

a k **b** $E(X)$ **c** $\text{Var}(X)$

3 The probability density function of a continuous random variable Y is given by

$$f(x) = \begin{cases} y(2 + 3y) & 0 < y < c \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

a Find the value of c. **b** Find the mean of Y.

4 A continuous random variable X has probability density function

$$p(x) = \begin{cases} kx & 0 \le x \le 4 \\ 4k & 4 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

Find:

a k **b** $E(X)$ **c** $Var(X)$ **d** the median of x **e** $P(3 \le X \le 5)$

5 A continuous random variable X has a probability density function

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

where k is a constant.

Find:

a k **b** $E(X)$ **c** $Var(X)$ **d** the median of X.

6 A continuous probability density function is described as

$$f(x) = \begin{cases} ce^x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

a Find the value of c. **b** Find the mean of the distribution.

7 A continuous random variable X has a probability density function

$f(x) = k \cos x,\ 0 \le x \le \frac{\pi}{2}$.

Find:

a k **b** $E(X)$ **c** $Var(X)$ **d** the median of X **e** $P\left(|X - m| > \frac{1}{2}\right)$

8 A continuous probability distribution is defined as

$$p(x) = \begin{cases} \dfrac{1}{1 + x^2} & 0 \le x \le k \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

Find:

a k
b the mean, μ
c the standard deviation, σ
d the median, m
e $P\left(|X - m| > \frac{1}{4}\right)$

9 **a** If $x \ge 0$, what is the largest domain of the function $f(x) = \dfrac{1}{\sqrt{1 - 4x^2}}$?

The function $f(x) = \dfrac{1}{\sqrt{1 - 4x^2}}$ is now to be used as a probability density function for a continuous random variable X.

b For it to be a probability density function for a continuous random variable X, what is the domain, given that the lower bound of the domain is 0?

c Find the mean of X.

d Find the standard deviation of X.

10 A continuous random variable has a probability density function given by

$$f(x) = \begin{cases} \dfrac{2}{\pi(1 + x^2)} & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Without using a calculator, find $P\left(|X| \le \tan \frac{\pi}{4}\right)$.

b Find the mean of X.

c Find the standard deviation of X.

11 A continuous random variable X has probability density function

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

where k and c are positive constants. Show that $k = \dfrac{-c^3e^{2c}}{2(2c^2 + c + 1)}$.

12 The time taken in minutes for a carpenter in a factory to make a wooden shelf follows the probability density function

$$f(t) = \begin{cases} \dfrac{6}{56}(15t - t^2 - 50) & 6 \le t \le 10 \\ 0 & \text{otherwise} \end{cases}$$

a Find:
i μ
ii σ^2

b A carpenter is chosen at random. Find the probability that the time taken for him to complete the shelf lies in the interval $[\mu - \sigma, \mu]$.

13 The lifetime of Superlife batteries is X years where X is a continuous random variable with probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ ke^{\frac{-x}{3}} & 0 \le x \le 6 \end{cases}$$

where k is a constant.

a Find the exact value of k.

b Find the probability that a battery fails after 4 months.

c A computer keyboard takes six batteries, but needs a minimum of four batteries to operate. Find the probability that the keyboard will continue to work after 9 months.

14 The probability that an express train is delayed by more than X minutes is modelled by the probability density function $f(x) = \dfrac{1}{72\,000}(x - 60)^2,\ 0 \le x \le 60$. It is assumed that no train is delayed by more than 60 minutes.

a Sketch the curve.

b Find the standard deviation of X.

c Find the median, m, of X.

15 A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx^2 + c & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

where k and c are constants.

The mean of X is $\frac{3}{2}$.

a Find the values of k and c.

b Find the variance of X.

c Find the median, m, of X.

d Find $P(|X - 1| < \sigma)$ where σ is the standard deviation of X.

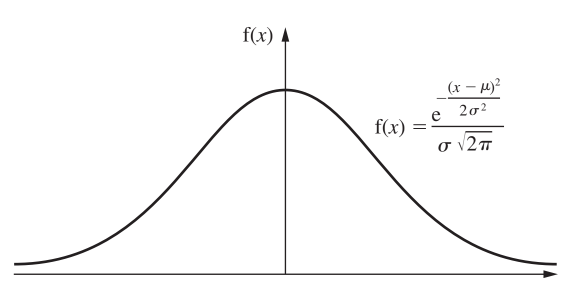
22.3 Normal distributions

We found in Chapter 21 that there were special discrete distributions, which modelled certain types of data. The same is true for continuous distributions and the normal distribution is probably the most important continuous distribution in statistics since it models data from natural situations quite effectively. This includes heights and weights of human beings. The probability density function for this curve is quite complex and contains two parameters, μ the mean and σ^2 the variance.

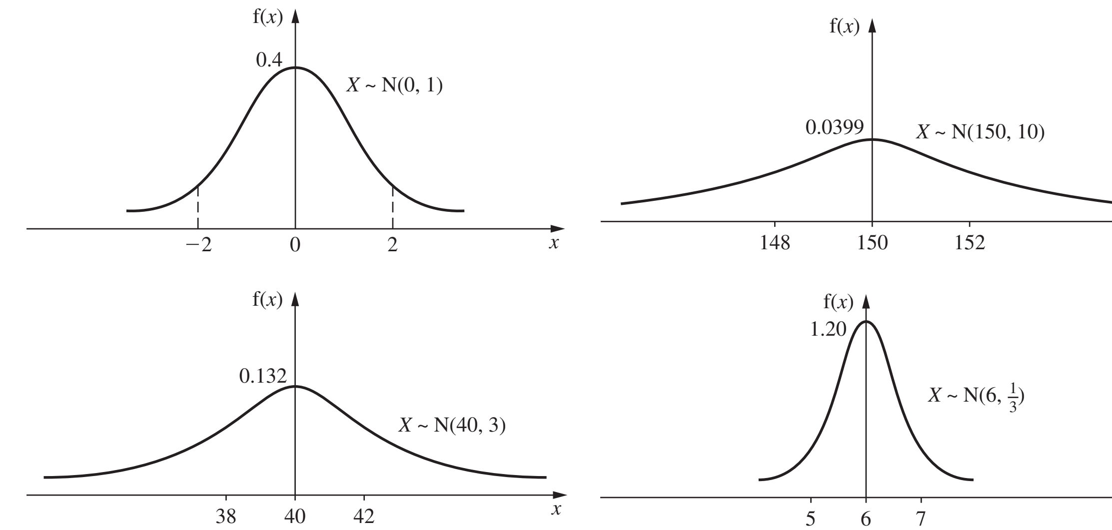
The probability density function for a normal distribution is $f(x) = \dfrac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$.

If a random variable X follows a **normal distribution**, we say $X \sim N(\mu, \sigma^2)$.

When we draw the curve it is a bell-shaped distribution as shown below.



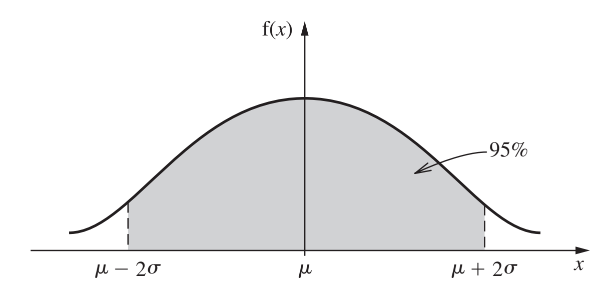
The exact shape of the curve is dependent on the values of μ and σ and four examples are shown below.



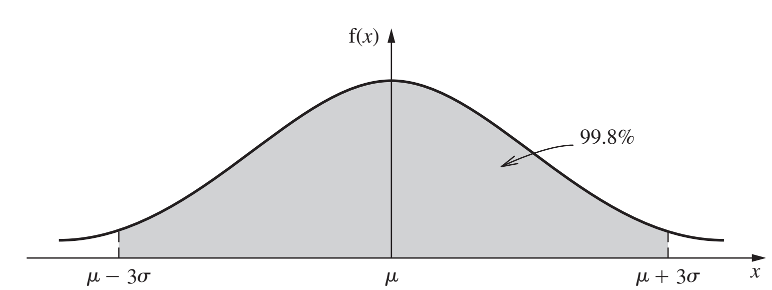
We normally make μ the axis of symmetry, but we could draw them as translations of the normal curve centred on $\mu = 0$.

Important results

1. The area under the curve is 1, meaning that $f(x)$ is a probability density function.
2. The curve is symmetrical about μ, that is the part of curve to the left of $x = \mu$ is the mirror image of the part to the right. Hence $P(-a \le X \le a) = 2P(0 \le X \le a)$ and $P(X \ge \mu) = P(X \le \mu) = 0.5$.
3. We can find the probability for any value of x since the probability density function is valid for all values between $\pm\infty$. The further away the value of x is from the mean, the smaller the probability becomes.
4. Approximately 95% of the distribution lies within two standard deviations of the mean.



5. Approximately 99.8% of the distribution lies within three standard deviations of the mean.



6. The maximum value of f(x) occurs when $x = \mu$ and is given by $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}$. Hence in the case of a normal distribution, the mean and the mode are the same.
7. $E(X) = \mu$. The proof of this involves mathematics beyond the scope of this syllabus.
8. $\text{Var}(X) = \sigma^2$. Again the proof of this involves mathematics beyond the scope of this syllabus.
9. The curve has points of inflexion at $x = \mu - \sigma$ and $x = \mu + \sigma$.

Finding probabilities from the normal distribution

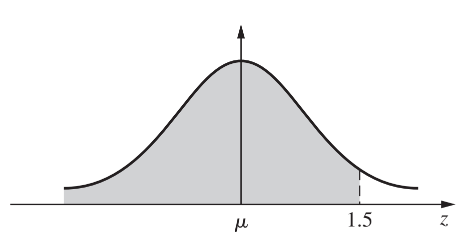
Theoretically, this works in exactly the same way as for any continuous random variable and hence if we have a normal distribution with $\mu = 0$ and $\sigma^2 = 1$, that is $X \sim N(0, 1)$, and we want to find $P(-0.5 \leq X \leq 0.5)$ the calculation we do is $\displaystyle\int_{-0.5}^{0.5} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$. This could be done on a calculator, but would be very difficult to do manually. In fact there is no direct way of integrating this function manually and approximate methods would need to be used. In the past this problem was resolved by looking up values for the different probabilities in tables of values, but now graphing calculators will do the calculation directly. Within this syllabus you will not be required to use tables of values and it is unlikely that a question on normal distributions would appear on a non-calculator paper.

Since there are infinite values of μ and σ there are an infinite number of possible distributions. Hence we designate what we call a standard normal variable Z and these are the values that appear in tables and are the default values on a calculator. The standard normal distribution is one that has mean 0 and variance 1, that is $Z \sim N(0, 1)$.

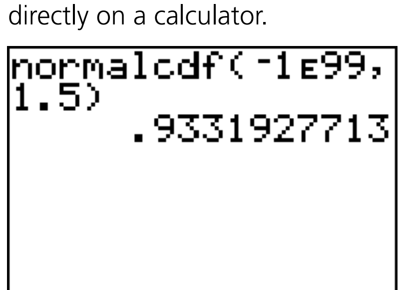
Example

Find $P(Z \leq 1.5)$.

The diagram for this is shown below.



We do the calculation directly on a calculator.



$P(Z \leq 1.5) = 0.933$

It is often a good idea to draw a sketch showing what you need.

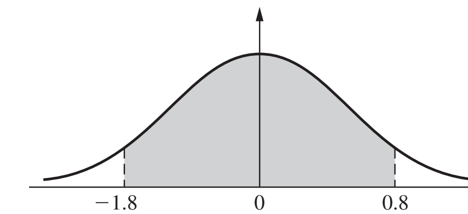
The value -1×10^{99} was chosen as the lower bound because the number is so small that the area under the curve to the left of that bound is negligible.

If we need to find a probability where Z is greater than a certain value or between two values, this works in the same way.

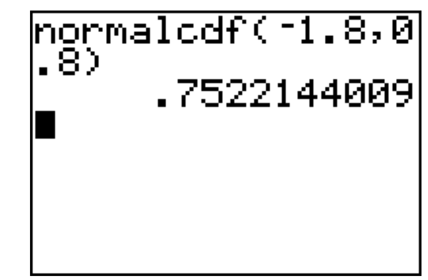
Example

Find $P(-1.8 \leq Z \leq 0.8)$.

The diagram for this is shown below.



Again, we do the calculation directly on a calculator.



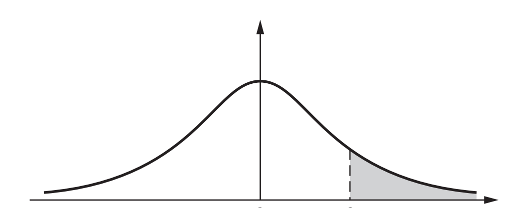
$P(-1.8 \leq Z \leq 0.8) = 0.752$

More often than not we will be using normal distributions other than the standard normal distribution. This works in the same way, except we need to tell the calculator the distribution from which we are working.

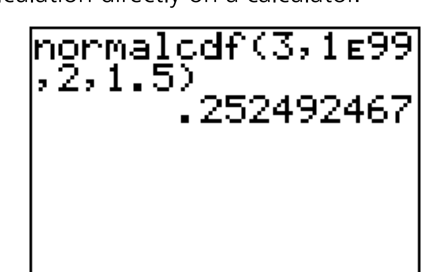
Example

If $X \sim N(2, 1.5^2)$, find $P(X \geq 3)$.

The diagram for this is shown below.



Again, we do the calculation directly on a calculator.



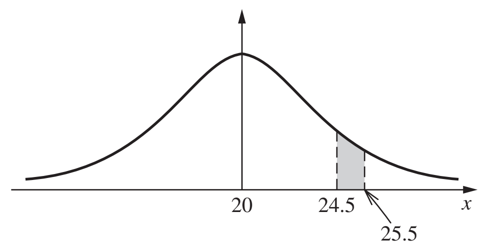
$P(X \geq 3) = 0.252$

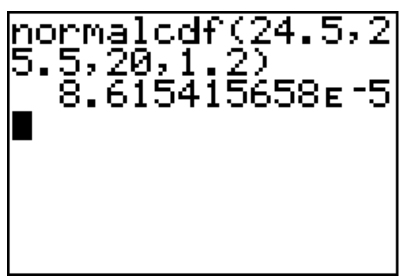
Since we are dealing with continuous distributions, if we are asked to find the probability of X being a specific value, then we need to turn this into a range.

Example

If $X \sim N(20, 1.2^2)$, find $P(X = 25)$, given that 25 is correct to 2 significant figures.

In terms of continuous data $P(X = 25) = P(24.5 \leq X < 25.5)$ and hence this is what we calculate.
This is shown in the diagram.





$P(X = 25) = 8.62 \times 10^{-5}$

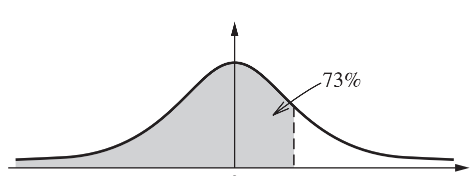
For the normal distribution, calculating the probability of X "less than" or the probability of X "less than or equal to" amounts to exactly the same calculation.

Up until now we have been calculating probabilities. Now we also need to be able to find the values that give a defined probability using a calculator.

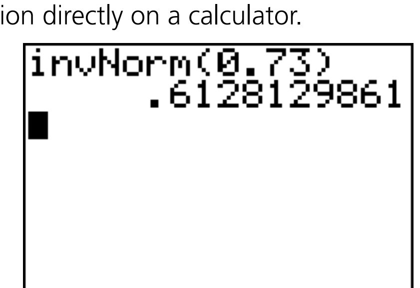
Example

Find a if $P(Z \leq a) = 0.73$.

In this case we are using the standard normal distribution and we are told the area is 0.73, that is the probability is 0.73, and we want the value. This is shown in the diagram.



We do the calculation directly on a calculator.



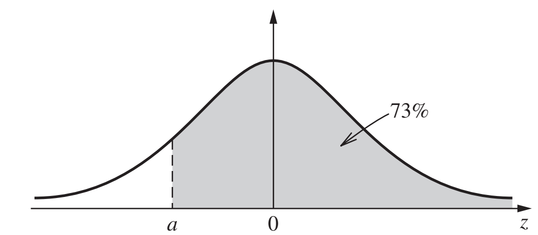
$a = 0.613$

The calculator will only calculate the value that provides what we call the **lower tail** of the graph and if we wanted $P(Z \geq a) = 0.73$, which we call the **upper tail**, we would need to undertake a different calculation. An upper tail is an area greater than a certain value and a lower tail is an area less than a certain value. This is why it can be very useful to draw a sketch first to see what is required.

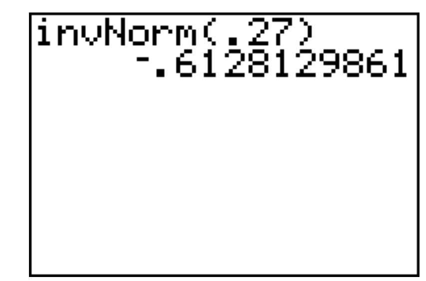
Example

Find a if $P(Z \geq a) = 0.73$.

This is shown in the diagram.



In this case $P(Z \leq a) = 1 - 0.73 = 0.27$.
We do this calculation directly on a calculator.



$a = -0.613$

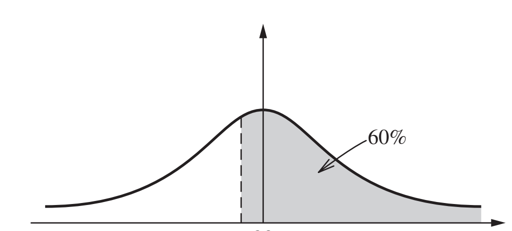
Because of the symmetry of the curve, the answer is the negative of the answer in the previous example. You can use this property, but it is probably easier to always use the lower tail of the distribution. This negative property only appears on certain distributions, including the standard normal distribution, since the values to the left of the mean depend on the value of the mean.

If we do not have a standard normal distribution, we can still do these questions on a calculator, but this time we need to specify μ and σ.

Example

If $X \sim N(20, 3.2^2)$, find a where $P(X \geq a) = 0.6$.

This is shown in the diagram.



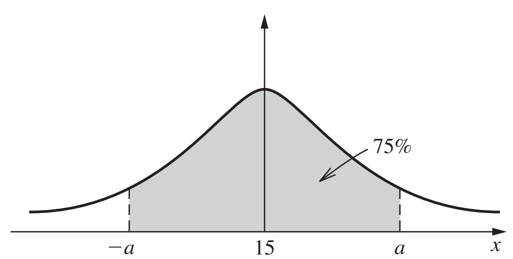
In this case $P(X \le a) = 1 - 0.6 = 0.4$.
We do this calculation directly on a calculator.

```
invNorm(.4,20,3.
2)
          19.18928928
```

$a = 19.2$

Example

If $X \sim N(15, 0.8^2)$, find a where $P(|X| \le a) = 0.75$.
This is the same as finding $P(-a \le X \le a) = 0.75$.
This is shown in the diagram.



In this case $P(X \le a) = \dfrac{1 - 0.75}{2} = 0.125$.
We do this calculation directly on a calculator.

```
invNorm(0.125,15
,0.8)
          14.0797205
```

$a = 14.1$

Exercise 3

1 If $Z \sim N(0, 1)$, find:

a $P(Z \le 0.756)$ **b** $P(Z \le 0.224)$
c $P(Z \ge -0.341)$ **d** $P(Z \le -1.76)$
e $P(Z \le 1.43)$ **f** $P(0.831 < Z < 1.25)$
g $P(-0.561 < Z < -0.0232)$ **h** $P(-1.28 < Z < 0.419)$
i $P(|Z| < 1.41)$ **j** $P(|Z| > 0.614)$

2 If $Z \sim N(0, 1)$, find a where

a $P(Z < a) = 0.548$ **b** $P(Z < a) = 0.937$
c $P(Z < a) = 0.346$ **d** $P(Z < a) = 0.249$
e $P(Z > a) = 0.0456$ **f** $P(Z > a) = 0.686$
g $P(Z > a) = 0.159$ **h** $P(Z > a) = 0.0598$
i $P(|Z| > a) = 0.611$ **j** $P(|Z| < a) = 0.416$

3 If $X \sim N(250, 49)$, find:

a $P(X > 269)$ **b** $P(X > 241)$
c $P(X < 231)$ **d** $P(X < 263)$

4 If $X \sim N(63, 9)$, find:

a $P(X > 67)$ **b** $P(X > 54.5)$ **c** $P(X < 68)$
d $P(X < 59.5)$ **e** $P(X = 62)$

5 If $X \sim N(-15, 16)$, find:

a $P(X > -10)$ **b** $P(X > -18.5)$ **c** $P(X < -3.55)$
d $P(X < -20.1)$ **e** $P(X = -14)$

6 If $X \sim N(125, 70)$, find:

a $P(85 < X < 120)$ **b** $P(90 < X < 100)$
c $P(|X - 125| < \sqrt{70})$ **d** $P(|X - 100| < 9)$

7 If $X \sim N(80, 22)$, find:

a $P(75 < X < 90)$ **b** $P(60 < X < 73)$
c $P(|X - 80| < \sqrt{22})$ **d** $P(|X - 80| < 3\sqrt{22})$

8 If $X \sim N(40, 4)$, find a where

a $P(X < a) = 0.617$ **b** $P(X < a) = 0.293$
c $P(X > a) = 0.173$ **d** $P(X > a) = 0.651$

9 If $X \sim N(85, 15)$, find a where

a $P(X < a) = 0.989$ **b** $P(X < a) = 0.459$
c $P(X > a) = 0.336$ **d** $P(X > a) = 0.764$

10 If $X \sim N(300, 49)$, find a where

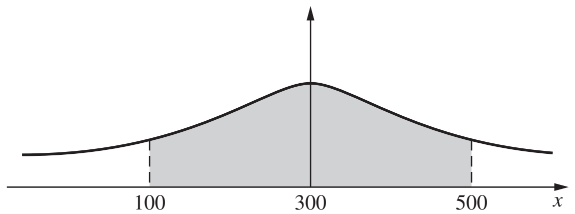
a $P(|X - 300| < a) = 0.6$ **b** $P(|X - 300| < a) = 0.95$
c $P(|X - 300| < a) = 0.99$ **d** $P(|X - 300| < a) = 0.45$

11 Z is a standardized normal random variable with mean 0 and variance 1. Find the upper quartile and the lower quartile of the distribution.

12 Z is a standardized normal random variable with mean 0 and variance 1. Find the value of a such that $P(|Z| \le a) = 0.65$.

13 A random variable X is normally distributed with mean -2 and standard deviation 1.5. Find the probability that an item chosen from this distribution will have a positive value.

14 The diagram below shows the probability density function for a random variable X which follows a normal distribution with mean 300 and standard deviation 60.



Find the probability represented by the shaded region.

15 The random variable Y is distributed normally with mean 26 and standard deviation 1.8. Find $P(23 \leq Y \leq 30)$.

16 A random variable X is normally distributed with mean zero and standard deviation 8. Find the probability that $|X| > 12$.

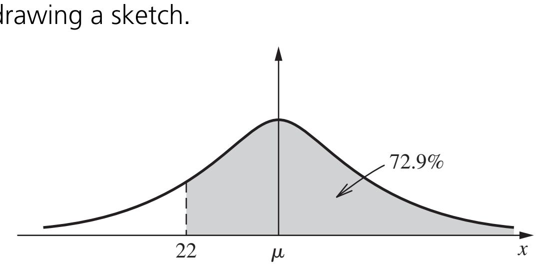
22.4 Problems involving finding μ and σ

To do this, we need to know how to convert any normal distribution to the standard normal distribution. Since the standard normal distribution is N(0,1) we standardize X which is $N(\mu, \sigma^2)$ using $Z = \frac{X - \mu}{\sigma}$. This enables us to find the area under the curve by finding the equivalent area on a standard curve. Hence if $X \sim N(2, 0.5^2)$ and we want $P(X \leq 2.5)$, this is the same as finding $P\left(Z \leq \frac{2.5 - 2}{0.5}\right) = P(Z \leq 1)$ on the standard curve. The way to find μ and/or σ if they are unknown is best demonstrated by example.

Example

If $X \sim N(\mu, 7)$ and $P(X \geq 22) = 0.729$, find the value of μ.

We begin by drawing a sketch.



It is clear from the sketch that $\mu > 22$.

Since the question gives an upper tail, we want the value of Z associated with a probability of $1 - 0.729 = 0.271$ which can be found on a calculator to be -0.610.

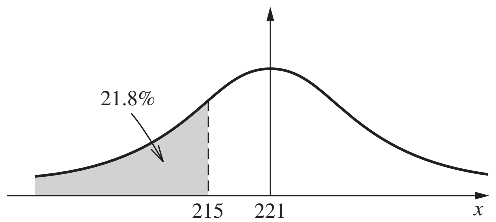
Since $Z = \frac{X - \mu}{\sigma}$ we have $-0.610 = \frac{22 - \mu}{7}$

$\Rightarrow \mu = 26.3$

Example

If $X \sim N(221, \sigma^2)$ and $P(X \leq 215) = 0.218$, find the value of σ.

Again, we begin by drawing a sketch.



The question gives a lower tail and hence we want the value of Z associated with a probability of 0.218, which can be found on a calculator to be -0.779.

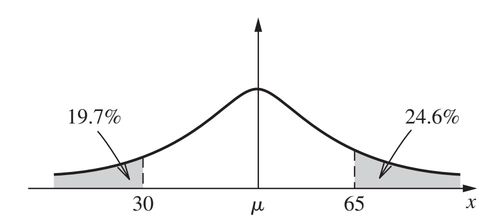
Since $Z = \frac{X - \mu}{\sigma}$ we have $-0.779 = \frac{215 - 221}{\sigma}$

$\Rightarrow \sigma = 7.70$

Example

If $X \sim N(\mu, \sigma^2)$, $P(X \leq 30) = 0.197$ and $P(X \geq 65) = 0.246$, find the values of μ and σ.

In this case we will have two equations and we will need to solve them simultaneously. Again we begin by drawing a sketch.



We first want the value of Z associated with a probability of 0.197, which can be found on a calculator to be -0.852.

$$-0.852 = \frac{30 - \mu}{\sigma}$$

$\Rightarrow -0.852\sigma = 30 - \mu$ equation (i)

To find the value of Z associated with 0.246 we need to use $1 - 0.246 = 0.754$ as it is an upper tail that is given. From the calculator we find the required value is 0.687.

$$0.687 = \frac{65 - \mu}{\sigma}$$

$\Rightarrow 0.687\sigma = 65 - \mu$ equation (ii)

We now subtract equation (i) from equation (ii) to find $\sigma = 19.5$.

Substituting back in equation (i) allows us to find $\mu = 51.6$.

Exercise 4

1 If $X \sim N(\mu, 1.5)$ and $P(X > 15.5) = 0.372$, find the value of μ.

2 If $X \sim N(\mu, 18)$ and $P(X > 72.5) = 0.769$, find the value of μ.

3 If $X \sim N(\mu, 7)$ and $P(X < 28.5) = 0.225$, find the value of μ.

4 If $X \sim N(\mu, 3.5)$ and $P(X < 41) = 0.852$, find the value of μ.

5 If $X \sim N(56, \sigma^2)$ and $P(X \leq 49) = 0.152$, find the value of σ.

6 If $X \sim N(15, \sigma^2)$ and $P(X \leq 18.5) = 0.673$, find the value of σ.

7 If $X \sim N(535, \sigma^2)$ and $P(X \geq 520) = 0.856$, find the value of σ.

8 If $X \sim N(125, \sigma^2)$ and $P(X \geq 135) = 0.185$, find the value of σ.

9 If $X \sim N(\mu, \sigma^2)$, $P(X \leq 8.5) = 0.247$ and $P(X \geq 14.5) = 0.261$, find the values of μ and σ.

10 If $X \sim N(\mu, \sigma^2)$, $P(X \leq 45) = 0.384$ and $P(X \geq 42.5) = 0.811$, find the values of μ and σ.

11 If $X \sim N(\mu, \sigma^2)$, $P(X \leq 268) = 0.0237$ and $P(X \geq 300) = 0.187$, find the values of μ and σ.

12 A random variable X is normally distributed with mean μ and standard deviation σ such that $P(X > 30.1) = 0.145$ and $P(X < 18.7) = 0.211$.

a Find the values of μ and σ. **b** Hence find $P(|X - \mu| < 3.5)$.

13 The random variable X is normally distributed and $P(X \leq 14.1) = 0.715$, $P(X \leq 18.7) = 0.953$. Find $E(X)$.

22.5 Applications of normal distributions

Normal distributions have many applications and are used as mathematical models within science, commerce etc. Hence problems with normal distributions are often given in context, but the mathematical manipulation is the same.

Example

The life of a certain make of battery is known to be normally distributed with a mean life of 150 hours and a standard deviation of 15 hours. Estimate the probability that the life of such a battery will be

a greater than 170 hours

b less than 120 hours

c within the range 135 hours to 155 hours.

Six batteries are chosen at random. What is the probability that

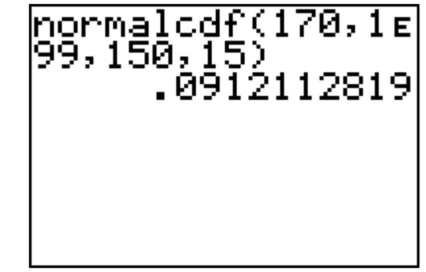
d exactly three of them have a life of between 135 hours and 155 hours

e at least one of them has a life of between 135 hours and 155 hours?

$X \sim N(150, 15^2)$

a We require $P(X > 170)$.

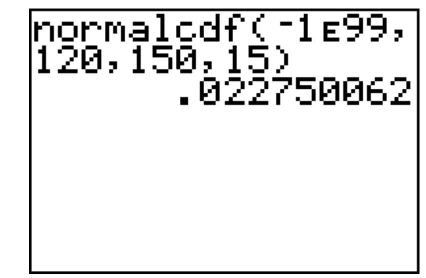
This is shownbelow.



$P(X > 170) = 0.0912$

b We require $P(X < 120)$.

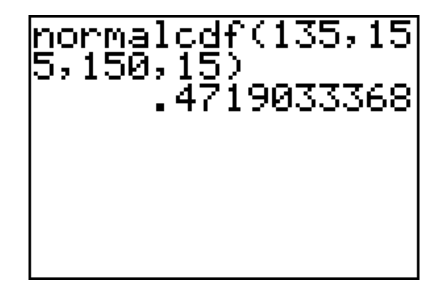
This is shown below.



$P(X < 120) = 0.0228$

c We require $P(135 < X < 155)$.

This is shown below.



$P(135 < X < 155) = 0.472$

d We can now model this using a binomial distribution, $Y \sim Bin(6, 0.472)$.

$$P(Y = 3) = {}^6C_3(0.472)^3(0.528)^3 = 0.310$$

e Again, we use the binomial distribution and in this case we need

$$P(Y \geq 1) = 1 - P(Y = 0) = 0.978$$

It is quite common for questions to involve finding a probability using the normal distribution and then taking a set number of these events, which leads to setting up a binomial distribution.

Example

The weight of chocolate bars produced by a particular machine follows a normal distribution with mean 80 grams and standard deviation 4.5 grams. A chocolate bar is rejected if its weight is less than 75 grams or more than 83 grams.

a Find the percentage of chocolate bars which are accepted.

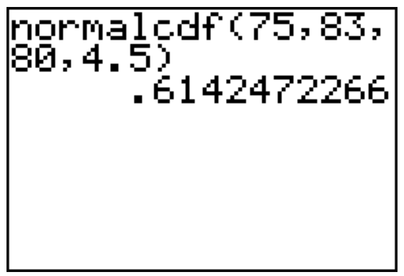
The setting of the machine is altered so that both the mean weight and the standard deviation change. With the new setting, 2% of the chocolate bars are rejected because they are too heavy and 3% are rejected because they are too light.

b Find the new mean and the new standard deviation.

c Find the range of values of the weight such that 95% of chocolate bars are equally distributed about the mean.

$X \sim N(80, 4.5^2)$

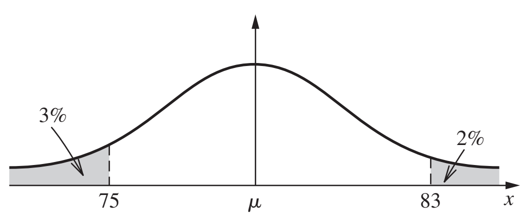
a We require $P(75 < X < 83)$.
This is shown below.



$P(75 < X < 83) = 0.614$

$\Rightarrow 61.4\%$ are accepted.

b In this case we need to solve two equations simultaneously.
We begin by drawing a sketch.



We first want the value of Z associated with a probability of 0.03. Since this is a lower tail, it can be found directly on a calculator to be -1.88.

$P(X < 75) = -1.88 = \dfrac{75 - \mu}{\sigma}$

$\Rightarrow -1.88\sigma = 75 - \mu$ equation (i)

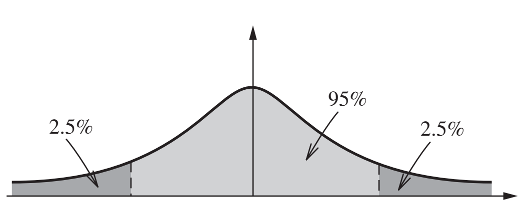
To find the value of Z associated with 0.02 we need to use $1 - 0.02 = 0.98$ as it is an upper tail that is given. From the calculator we find the required value is 2.05.

$2.05 = \dfrac{83 - \mu}{\sigma}$

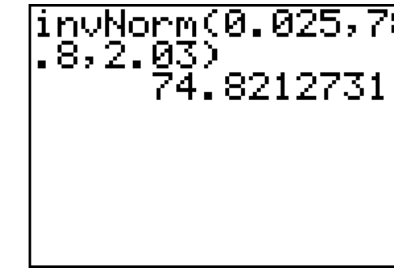
$\Rightarrow 2.05\sigma = 83 - \mu$ equation (ii)

If we now subtract equation (i) from equation (ii) we find $\sigma = 2.03$. Substituting back in equation (i) allows us to find $\mu = 78.8$.

c Again we begin by drawing a sketch.



In this case $X \sim N(78.8, 2.03^2)$ and we require $P(X < a) = 0.025$.
This is shown below.



Hence the lower bound of the range is 74.8.

The upper bound is given by $78.8 + (78.8 - 74.8) = 82.8$.

The range of values required is $74.8 < X < 82.8$.

Exercise 5

1 The weights of a certain breed of otter are normally distributed with mean 2.5 kg and standard deviation 0.55 kg.

a Find the probability that the weight of a randomly chosen otter lies between 2.25 kg and 2.92 kg.

b What is the weight of less than 35% of this breed of otter?

2 Jars of jam are produced by Jim's Jam Company. The weight of a jar of jam is normally distributed with a mean of 595 grams and a standard deviation of 8 grams.

a What percentage of jars has a weight of less than 585 grams?

b Given that 50% of the jars of jam have weights between m grams and n grams, where m and n are symmetrical about 595 grams and $m < n$, find the values of m and n.

3 The temperature T on the first day of July in England is normally distributed with mean 18°C and standard deviation 4°C. Find the probability that the temperature will be

a more than 20°C

b less than 15°C

c between 17°C and 22°C.

4 The heights of boys in grade 11 follow a normal distribution with mean 170 cm and standard deviation 8 cm. Find the probability that a randomly chosen boy from this grade has height

a less than 160 cm

b less than 175 cm

c more than 168 cm

d more than 178 cm

e between 156 cm and 173 cm

f between 167 cm and 173 cm.

5 The mean weight of 600 male students in a college is 85 kg with a standard deviation of 9 kg. The weights are normally distributed.

a Find the number of students whose weight lies in the range 75 kg to 95 kg.

b 62% of students weigh more than a kg. Find the value of a.

6 The standard normal variable has probability density function $f(x) = \dfrac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$.
Find the coordinates of the two points of inflexion.

7 A manufacturer makes ring bearings for cars. Bearings below 7.5 cm in diameter are too small while those above 8.5 cm are too large. If the diameter of bearings produced is normally distributed with mean 7.9 cm and standard deviation 0.3 cm, what is the probability that a bearing chosen at random will fit?

8 The mean score for a mathematics quiz is 70 with a standard deviation of 15. The test scores are normally distributed.

a Find the number of students in a class of 35 who score more than 85 in the quiz.

b What score should more than 80% of students gain?

9 For the delivery of a package to be charged at a standard rate by a courier company, the mean weight of all the packages must be 1.5 kg with a standard deviation of 100 g. The packages are assumed to be normally distributed. A company sends 50 packages, hoping they will all be charged at standard rate. Find the number of packages that should have a weight

a of less than 1.4 kg

b of more than 1.3 kg

c of between 1.2 kg and 1.45 kg.

10 At Sandy Hollow on a highway, the speeds of cars have been found to be normally distributed. 80% of cars have speeds greater than 55 kilometres per hour and 10% have speeds less than 50 kilometres per hour. Calculate the mean speed and its standard deviation.

11 Packets of biscuits are produced such that the weight of the packet is normally distributed with a mean of 500 g and a standard deviation of 50 g.

a If a packet of biscuits is chosen at random, find the probability that the weight lies between 490 g and 520 g.

b Find the weight exceeded by 10% of the packets.

c If a supermarket sells 150 packets in a day, how many will have a weight less than 535 g?

12 Bags of carrots are sold in a supermarket with a mean weight of 0.5 kg and standard deviation 0.05 kg. The weights are normally distributed. If there are 120 bags in the supermarket, how many will have a weight

a less than 0.45 kg

b more than 0.4 kg

c between 0.45 kg and 0.6 kg?

13 The examination scores in an end of year test are normally distributed with a mean of 70 marks and a standard deviation of 15 marks.

a If the pass mark is 50 marks, find the percentage of candidates who pass the examination.

b If 5% of students gain a prize for scoring above y marks, find the value of y.

14 The time taken to get to the desk in order to check in on a flight operated by Surefly Airlines follows a normal distribution with mean 40 minutes and standard deviation 12 minutes. The latest time that David can get to the desk for a flight is 1400. If he arrives at the airport at 1315, what is the probability that he will miss the flight?

15 Loaves of bread made in a particular bakery are found to follow a normal distribution X with mean 250 g and standard deviation 30 g.

a 3% of loaves are rejected for being underweight and 4% of loaves are rejected for being overweight. What is the range of weights of a loaf of bread such that it should be accepted?

b If three loaves of bread are chosen at random, what is the probability that exactly one of them has a weight of more than 270 g?

16 Students' times to run a 200 metre race are measured at a school sports day. There are ten races and five students take part in each race. The results are shown in the table below.

Time to nearest second	26	27	28	29	30	31
Number of students	3	7	15	14	9	2

a Find the mean and the standard deviation of these times.

b Assuming that the distribution is approximately normal, find the percentage of students who would gain a time between 27.5 seconds and 29.5 seconds.

17 Apples are sold on a market stall and have a normal distribution with mean 300 grams and standard deviation 30 grams.

a If there are 500 apples on the stall, what is the expected number with a weight of more than 320 grams?

b Given that 25% of the apples have a weight less than m grams, find the value of m.

18 The lengths of screws produced in a factory are normally distributed with mean μ and standard deviation 0.055 cm. It is found that 8% of screws have a length less than 1.35 cm.

a Find μ.

b Find the probability that a screw chosen at random will be between 1.55 cm and 1.70 cm.

19 In a zoo, it is found that the height of giraffes is normally distributed with mean height H metres and standard deviation 0.35 metres. If 15% of giraffes are taller than 4.5 metres, find the value of H.

20 The weights of cakes sold by a baker are normally distributed with a mean of 280 grams. The weights of 18% of the cakes are more than 310 grams.

a Find the standard deviation.

b If three cakes are chosen at random, what is the probability that exactly two of them have weights of less than 260 grams?

21 A machine in a factory is designed to produce boxes of chocolates which weigh 0.5 kg. It is found that the average weight of a box of chocolates is 0.57 kg. Assuming that the weights of the boxes of chocolate are normally distributed, find the variance if 2.3% of the boxes weigh below 0.5 kg.

22 The marks in an examination are normally distributed with mean μ and standard deviation σ. 5% of candidates scored more than 90 and 15% of candidates scored less than 40. Find the mean μ and the standard deviation σ.

23 The number of hours, T, that a team of secretaries works in a week is normally distributed with a mean of 37 hours. However, 15% of the team work more than 42 hours in a week.

a Find the standard deviation of T.

b Andrew and Balvinder work on the team. Find the probability that both secretaries work more than 40 hours in a week.

Review exercise

A calculator may be used in all questions unless exact answers are required.

1 A man is arranging flowers in a vase. The lengths of the flowers in the vase are normally distributed with a mean of μ cm and a standard deviation of σ cm. When he checks, he finds that 5% of the flowers are longer than 41 cm and 8% of flowers are shorter than 29 cm.

a Find the mean μ and the standard deviation σ of the distribution.

b Find the probability that a flower chosen at random is less than 45 cm long.

2 In a school, the heights of all 14-year-old students are measured. The heights of the girls are normally distributed with mean 155 cm and standard deviation 10 cm. The heights of the boys are normally distributed with mean 160 cm and standard deviation 12 cm.

a Find the probability that a girl is taller than 170 cm.

b Given that 10% of the girls are shorter than x cm, find x.

c Given that 90% of the boys have heights between q cm and r cm where q and r are symmetrical about 160 cm, and $q < r$, find the values of q and r.

In a group of 14-year-old students, 60% are girls and 40% are boys. The probability that a girl is taller than 170 cm was found in part **a**. The probability that a boy is taller than 170 cm is 0.202. A 14-year-old student is selected at random.

d Calculate the probability that the student is taller than 170 cm.

e Given that the student is taller than 170 cm, what is the probability that the student is a girl? [IB May 06 P2 Q4]

3 In a certain college the weight of men is normally distributed with mean 80 kg and standard deviation 6 kg. Find the probability that a man selected at random will have a weight which is

a between 65 kg and 90 kg

b more than 75 kg.

Three men are chosen at random from the college. Find the probability that

c none of them weigh more than 70 kg, giving your answer to 5 decimal places.

d at least one of them will weigh more than 70 kg.

4 A random variable X has probability density function f(x) where

$$f(x) = \begin{cases} \frac{1}{4}x & 0 \le x < 1 \\ \frac{1}{4} & 1 \le x < 3 \\ \frac{1}{12}(6-x) & 3 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

Find the median value of X. [IB Nov 97 P1 Q15]

5 A factory makes hooks which have one hole in them to attach them to a surface. The diameter of the hole produced on the hooks follows a normal distribution with mean diameter 11.5 mm and a standard deviation of 0.15 mm. A hook is rejected if the hole on the hook is less than 10.5 mm or more than 12.2 mm.

a Find the percentage of hooks that are accepted.

The settings on the machine are altered so that the mean diameter changes but the standard deviation remains unchanged. With the new settings 5% of hooks are rejected because the hole is too large.

b Find the new mean diameter of the hole produced on the hooks.

c Find the percentage of hooks rejected because the hole is too small in diameter.

d Six hooks are chosen at random. What is the probability that exactly three of them will have a hole in them that is too small in diameter?

6 a A machine is producing components whose lengths are normally distributed with a mean of 8.00 cm. An upper tolerance limit of 8.05 cm is set and on one particular day it is found that one in sixteen components is rejected. Estimate the standard deviation.

b The next day, due to production difficulties, it is found that one in twelve components is rejected. Assuming that the standard deviation has not changed, estimate the mean of the day's production.

c If 3000 components are produced during each day, how many would be expected to have lengths in the range 7.95 cm to 8.05 cm on each of the two days? [IB May 93 P2 Q8]

7 A continuous random variable X has probability density function defined by

$$f(x) = \begin{cases} \frac{k}{1+x^2} & \text{for } -\frac{1}{\sqrt{3}} \le x \le \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = \frac{2}{\pi}$.

b Sketch the graph of f(x) and state the mode of X.

c Find the median of X.

d Find the expected value of X.

e Find the variance of X. [IB Nov 93 P2 Q8]

8 A continuous random variable X has probability density function defined by

$$f(x) = \begin{cases} k|\sin x| & 0 \le x \le 2\pi \\ 0 & \text{otherwise} \end{cases}$$

a Find the exact value of k.

b Calculate the mean and the variance of X.

c Find $P\left(\frac{\pi}{2} \le X \le \frac{5\pi}{4}\right)$.

9 A company buys 44% of its stock of bolts from manufacturer A and the rest from manufacturer B. The diameters of the bolts produced by each manufacturer follow a normal distribution with a standard deviation of 0.16 mm.

The mean diameter of the bolts produced by manufacturer A is 1.56 mm. 24.2% of the bolts produced by manufacturer B have a diameter less than 1.52 mm.

a Find the mean diameter of the bolts produced by manufacturer B.

A bolt is chosen at random from the company's stock.

b Show that the probability that the diameter is less than 1.52 mm is 0.312 to 3 significant figures.

c The diameter of the bolt is found to be less than 1.52 mm. Find the probability that the bolt was produced by manufacturer B.

d Manufacturer B makes 8000 bolts in one day. It makes a profit of \$1.50 on each bolt sold, on condition that its diameter measures between 1.52 mm and 1.83 mm. Bolts whose diameters measure less than 1.52 mm must be discarded at a loss of \$0.85 per bolt. Bolts whose diameters measure over 1.83 mm are sold at a reduced profit of \$0.50 per bolt. Find the expected profit for manufacturer B. [IB May 05 P2 Q4]

10 The ages of people in a certain country with a large population are presently normally distributed. 40% of the people in this country are less than 25 years old.

a If the mean age is twice the standard deviation, find, in years, correct to 1 decimal place, the mean and the standard deviation.

b What percentage of the people in this country are more than 45 years old?

c According to the normal distribution, 2.28% of the people in this country are less than x years old. Find x and comment on your answer.

d If three people are chosen at random from this population, find the probability that

i all three are less than 25

ii two of the three are less than 25

iii at least one is less than 25.

e 40% of the people on a bus are less than 25 years old. If three people on this bus are chosen at random, what is the probability that all three are less than 25 years old?

f Explain carefully why there is a difference between your answers to **d i** and **e**.

[IB Nov 91 P2 Q8]

11 A business man spends X hours on the telephone during the day. The probability density function of X is given by

$$f(x) = \begin{cases} \frac{1}{12}(8x - x^3) & \text{for } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a **i** Write down an integral whose value is E(X).

ii Hence evaluate E(X).

b **i** Show that the median, m, of X satisfies the equation $m^4 - 16m^2 + 24 = 0$.

ii Hence evaluate m.

c Evaluate the mode of X. [IB May 03 P2 Q4]

12 A machine is set to produce bags of salt, whose weights are distributed normally with a mean of 110 g and standard deviation 1.142 g. If the weight of a bag of salt is less than 108 g, the bag is rejected. With these settings, 4% of the bags are rejected.
The settings of the machine are altered and it is found that 7% of the bags are rejected.

a **i** If the mean has not changed, find the new standard deviation, correct to 3 decimal places.

The machine is adjusted to operate with this new value of the standard deviation.

ii Find the value, correct to 2 decimal places, at which the mean should be set so that only 4% of the bags are rejected.

b With the new settings from part **a**, it is found that 80% of the bags of salt have a weight which lies between A g and B g, where A and B are symmetric about the mean. Find the values of A and B, giving your answers correct to 2 decimal places. [IB May 00 P2 Q4]

13 A farmer's field a yields crop of potatoes. The number of thousands of kilograms of potatoes the farmer collects is a continuous random variable X with probability density function $f(x) = k(3 - x)^3$, $0 \le x \le 3$ and $f(x) = 0$ otherwise, where k is a constant.

a Find the value of k.

b Find the mean of X.

c Find the variance of X.

d Potatoes are sold at 30 cents per kilogram, but cost the farmer 15 cents per kilogram to dig up. What is the expected profit?

14 The difference of two independent normally distributed variables is itself normally distributed. The mean is the difference between the means of the two variables, but the variance is the sum of the two variances.

Two brothers, Oliver and John, cycle home from school every day. The times taken for them to travel home from school are normally distributed and are independent. Oliver's times have a mean of 25 minutes and a standard deviation of 4 minutes. John's times have a mean of 20 minutes and a standard deviation of 5 minutes. What is the probability that on a given day, John arrives home before Oliver?

15 The continuous random variable X has probability density function f(x) where

$$f(x) = \begin{cases} e - ke^{kx}, & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a Show that $k = 1$.

b What is the probability that the random variable X has a value that lies between $\frac{1}{4}$ and $\frac{1}{2}$? Give your answer in terms of e.

c Find the mean and variance of the distribution. Give your answer exactly in terms of e.

The random variable X above represents the lifetime, in years, of a certain type of battery.

d Find the probability that a battery lasts more than six months.

A calculator is fitted with three of these batteries. Each battery fails independently of the other two. Finds the probability that at the end of six months

e none of the batteries has failed

f exactly one of the batteries has failed. [IB Nov 99 P2 Q4]