1 Binomial

1. (a) If $X \sim B(13, 0.12)$ find
(i) $P(X > 2)$ (ii) $P(X = 10)$

(b) If $X \sim B(10, 0.2)$ find
(i) $P(X \leqslant 3)$ (ii) $P(X < 7)$

2. If $X \sim B(9, 0.18)$ find $P(X < \mu - \sigma)$
where $\mu = E(X)$, $\sigma = \sqrt{Var(X)}$

3. X is a Binomial random variable based on 20 trials. Given that $E(X) = 15$

(a) find the standard deviation of X

(b) find $P(X > \mu)$, where $\mu = E(X)$.

4. A quality control agent tests sets of 20 components on a production line from which it is known that 95% of components are free from defects.

(i) Find the probability that a set chosen at random is free from defects.
(ii) If she tests five sets before lunch, find the probability that three of these sets were free from defects.
(iii) She tests sixty sets in a week's work. On average, how many sets does she find which are not perfect?

5. (a) A student taking a multiple choice test has done enough work that he can normally discount two of the five possible answers, but has to guess between the remaining options. The test has 20 questions. Find the probability that he gets less than 4 correct.

(b) Another student taking the same test has done more work, and finds that he knows the correct answers to half of the questions. In each of the other questions he can rule out 3 of the 5 given answers, but then has to guess.

(i) Find the probability that he scores more than 17 marks.
(ii) Find the probability that he gets no more than 1 wrong.

6. Suppose that 20% of fruit pastilles produced are blackcurrant flavoured. The number of blackcurrant pastilles in a packet may be modelled by a binomial distribution with parameter p equal to 0.2.

(a) A standard packet contains 8 fruit pastilles. Find the probability that a standard packet does not have any blackcurrant pastilles.

(b) A large packet contains 25 pastilles.
Determine the probability that a large packet contains:

(i) at most 5 blackcurrant pastilles
(ii) exactly 5 blackcurrant pastilles
(iii) more than 5 but fewer than 10 blackcurrant pastilles.

© Oxford University Press 2009

1 Binomial

Question Number	Solution	Marks
1 a	i $P(X > 2) = 1 - P(X \leqslant 2)$	M1
	or $1 - P(0) - P(1) - P(2)$	
	or similar	
	$= 1 - 0.88^{13} - 13 \times 0.88^{12} \times 0.12$	
	$- \frac{13 \times 12}{2 \times 1} \times 0.88^{11} \times 0.12^2$	M3
	$= 1 - 0.1898 - 0.3364 - 0.2753$	
	$= 0.20$ (2 dp)	A1
	ii $P(X = 10) = \binom{13}{10} \times 0.88^3 \times 0.12^{10}$	M1
	$= \frac{13 \times 12 \times 11}{3 \times 2 \times 1} \times 0.88^3 \times 0.12^{10}$	
	$= 0.00000012$ (2 sf)	A1
b	i $P(X \leqslant 3) = 0.879$ (3 dp) using tables	B1
	ii $P(X < 7) = P(X \leqslant 6)$	M1
	$= 0.999$ (3 dp) using tables	A1
		10
2	$\mu = np = 1.62$	M1
	$\sigma^2 = np(1 - p) = 1.3284$	
	So $\sigma \simeq 1.1526$.	
	Therefore	
	$P(X < \mu - \sigma)$	M1
	$= P(X < 0.4674)$	A1
	$= P(X = 0)$	
	$= (0.82)^9$	
	$\simeq 0.168$ (3 dp)	
		3
3 a	$X \sim B(20, p)$	
	$E(X) = 15$ so $np = 15$	M1
	$20p = 15$	
	$p = 0.75$	A1
	$Var(X) = np(1 - p)$	
	$= 20 \times 0.75 \times 0.25$	M1
	$= 3.75$	
	$sd(X) = \sqrt{3.75}$	
	$= 1.94$ (2 dp)	A1
b	$P(X > 15) = P(\text{not } X \leqslant 4)$	M1
	Where not $X \sim B(20, 0.25)$	
	(since if $X = 16$, not $X = 4$ etc)	
	hence using tables	
	$P(X > 15) = 0.415$ (3 dp)	A1
		6

© Oxford University Press 2009

4	**i** $n = 20$, P(defect) $= p = 0.05$ X = number of defects in batch of 20 $X \sim B(20, 0.05)$ $P(X = 0) = 0.359$ (3 dp) using tables **ii** Test 5 sets P(defect-free set) $= 1 - 0.3585 = 0.6415$ Y = number of defect-free sets $Y \sim B(5, 0.642)$ $P(Y = 3) = \binom{5}{3} \times 0.6415^3 \times 0.3585^2$ $= \frac{5 \times 4 \times 3}{3 \times 2 \times 1} \times 0.6415^3 \times 0.3585^2$ $= 0.339$ (3 dp) **iii** P(set is defective) $= 0.3585$ Test 60 sets Average number of sets with defects $= 60 \times 0.3585 = 21.5$ (1 dp)	M1 A1 M1 M1 A1 M1 A1 **7**
5 a	(Assuming there are 5 possible answers to each question!) X = correct answers out of 20 $X \sim B\left(20, \frac{1}{3}\right)$ $P(X < 4) = P(0) + P(1) + P(2) + P(3)$ $= 0.667^{20}$ $+ \binom{20}{1} \times 0.667^{19} \times 0.333^1$ $+ \binom{20}{2} \times 0.667^{18} \times 0.333^2$ $+ \binom{20}{3} \times 0.667^{17} \times 0.333^3$ $= 0.0003 + 0.0030 + 0.0143 + 0.0429$ $= 0.061$(3 dp)	 M1 M2 A1
b	**i** Since he knows 10 answers, and P(correct) = 0.5 for the other 10, use Y = correct answers out of 10 $Y \sim B(10, 0.5)$ P(scores more than 17) $= P(Y > 7)$ $= 1 - P(Y \leqslant 7)$ $= 1 - 0.945$ $= 0.055$ (3 dp) using tables **ii** P(no more than 1 wrong) = P(scores 19 or 20) $= P(Y > 8)$ $= 1 - P(Y \leqslant 8)$ $= 1 - 0.989$ $= 0.011$ (3 dp) using tables	 M1 M1 A1 M1 A1 **9**

6 a	X = number of blackcurrant $X \sim B(8, 0.2)$ $P(X = 0) = 0.8^8$ $= 0.168$ (3 dp)	 M1 A1
b	**i** Y = number of blackcurrant $Y \sim B(25, 0.2)$ $P(Y \leqslant 5)$ $= P(0) + P(1) + P(2) + P(3) + P(4) + P(5)$ $= 0.8^{25}$ $+ \binom{25}{1} \times 0.8^{24} \times 0.2^1$ $+ \binom{25}{2} \times 0.8^{23} \times 0.2^2$ $+ \binom{25}{3} \times 0.8^{22} \times 0.2^3$ $+ \binom{25}{4} \times 0.8^{21} \times 0.2^4$ $+ \binom{25}{5} \times 0.8^{20} \times 0.2^5$ $= 0.0038$ $+0.0236$ $+0.0708$ $+0.1358$ $+0.1867$ $+0.1960$ $= 0.617$ (3 dp)	 M1 M2
	ii $P(Y = 5) = \binom{25}{5} \times 0.8^{20} \times 0.2^5$ $= 0.196$ (3 dp) from above	 B1
	iii $P(6 \leqslant Y \leqslant 9) = P(6) + P(7) + P(8) + P(9)$ $= \binom{25}{6} \times 0.8^{19} \times 0.2^6$ $+ \binom{25}{7} \times 0.8^{18} \times 0.2^7$ $+ \binom{25}{8} \times 0.8^{17} \times 0.2^8$ $+ \binom{25}{9} \times 0.8^{16} \times 0.2^9$ $= 0.1633 + 0.1108 + 0.0623 + 0.0294$ $= 0.366$ (3 dp)	M1 M2 A2
		13

© Oxford University Press 2009

2 Poisson

1. (a) If $X \sim \text{Po}(2.4)$ find
(i) $P(X = 1)$ (ii) $P(X \leqslant 2)$

(b) If $X \sim \text{Po}(8.5)$ find
(i) $P(X < 6)$ (ii) $P(X = 5)$

2. The number of accidents which occur on a particular stretch of road in a day may be modelled by a Poisson distribution with parameter 2.5.

(a) Find the probability that on that stretch of road during a day
(i) exactly 2 accidents occur
(ii) less than 5 accidents occur.

(b) Assuming that the Poisson distribution continues to hold, find the probability that more than 8 accidents occur over the 3 days of a bank holiday weekend.

(c) Comment on whether you think the assumption in part **b** is reasonable.

3. Given that $X \sim \text{Po}(8)$

(a) write down an expression for $P(X = 8)$ in terms of $P(X = 7)$

(b) explain why X has two modes at 7 and 8

(c) find $P(X < \mu - \sigma)$ where $\mu = E(X)$, $\sigma = \sqrt{Var(X)}$

4. Customers enter a supermarket independently of one another, and at random intervals of time. The average rate is 1.8 people per minute during the time that the shop is open.

(a) If X is the number of customers entering the shop in a period of five minutes, write down the distribution that you would use to model X.

(b) Use this distribution to calculate the probability that
(i) there are 5 customers in one particular five minute period
(ii) there are more than 6 customers in a given five minute period.

(c) Find the probability that
(i) there are less than 9 customers in a ten minute period
(ii) exactly one customer enters in each of 5 one minute periods.

(d) Compare the probabilities obtained in **b i** and **c ii**.

© Oxford University Press 2009

5. A dual carriageway has one lane blocked off because of roadworks. The number of cars passing a point in a road in a number of one minute intervals is summarised in the table:

Number of cars	0	1	2	3	4	5	6
Frequency	2	3	6	32	45	4	2

(a) Calculate the mean and variance of the number of cars passing in any one minute interval.

(b) The Poisson distribution is not likely to provide an adequate model for the distribution of the number of cars passing in any one minute interval.

(i) Give a reason why, using your answer to part **a**.

(ii) Give a reason why you would not have expected it to provide a good model without reference to the observed data.

6. For the following situations state what assumptions are needed if a Poisson distribution is to be used to model them, and give the value of λ that would be used. You are **not** expected to do any calculations!

(a) On average cars pass a point in the road at a rate of 5 per minute. X = the number of cars passing that point in the road in a quarter of an hour.

(b) On average there are 3.2 dandelions per m^2 of a lawn. X = number of dandelions in a rectangular section of the lawn which is 3 m by 2 m.

(c) On average there are 17.5 particle emissions per second from a piece of radioactive material. X = the number of particle emissions in a minute.

7. A supermarket sells bags of apples, with each bag containing 10 apples. It is found that, on average, 3% of the apples are bad.

(a) Find the probability that a bag contains

(i) one bad apple

(ii) more than one bad apple.

(b) A box containing 15 of these bags is opened and inspected. Identify and use a suitable approximation to find the probability that there are no more than 3 bad apples in this box.

8. A firm is proud of their production statistics which show that only 0.1% of their components are faulty. The components are packed in boxes of 400.

(a) Write down the appropriate exact distribution to model the number of faulty components in a box chosen at random.

(i) Give two conditions which need to be satisfied for this distribution to be a suitable model.

(ii) Is it reasonable to assume that these conditions would be satisfied?

(b) (i) Write down the distribution you would use to calculate the probability that a box contains more than one faulty component.

(ii) Calculate this probability.

© Oxford University Press 2009

2 Poisson

Question Number	Solution	Marks
1 a	i If $X \sim \text{Po}(\lambda)$ then $\text{P}(X = r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$	
	$X \sim \text{Po}(2.4)$	
	$\text{P}(X = 1) = \dfrac{e^{-2.4}(2.4)^1}{1!}$	M1
	$= 0.218$ (3 dp)	A1
	ii $\text{P}(X \leqslant 2) = \text{P}(0) + \text{P}(1) + \text{P}(2)$	
	$= \dfrac{e^{-2.4}(2.4)^0}{0!} + \dfrac{e^{-2.4}(2.4)^1}{1!} + \dfrac{e^{-2.4}(2.4)^2}{2!}$	M1
	$= 0.0907 + 0.2177 + 0.2613$	
	$= 0.570$ (3 dp)	A2
b	i $X \sim \text{Po}(8.5)$	
	Using tables	
	$\text{P}(X < 6) = \text{P}(X \leqslant 5)$	M1
	$= 0.150$ (3 dp)	A1
	ii $\text{P}(X = 5) = \text{P}(X \leqslant 5) - \text{P}(X \leqslant 4)$	M1
	$= 0.1496 - 0.0744$	A1
	$= 0.075$ (3 dp)	
		7
2 a	i $X \sim \text{Po}(2.5)$	
	Using tables	
	$\text{P}(X = 2) = \text{P}(X \leqslant 2) - \text{P}(X \leqslant 1)$	M1
	$= 0.5438 - 0.2873$	
	$= 0.257$ (3 dp)	A1
	ii $\text{P}(X < 5) = \text{P}(X \leqslant 4)$	M1
	$= 0.891$ (3 dp)	A1
b	In 3 days, the average number of accidents should be $3 \times 2.5 = 7.5$	M1
	Use $Y \sim \text{Po}(7.5)$	
	$\text{P}(Y > 8) = 1 - \text{P}(Y \leqslant 8)$	M1
	$= 1 - 0.6620$	
	$= 0.338$ (3 dp)	A1
c	The assumption is risky since there is generally more traffic during a bank holiday weekend, and more accidents.	B1
		8

© Oxford University Press 2009

3 a	If $X \sim \text{Po}(8)$ then $\text{P}(X = r) = \dfrac{e^{-8}8^r}{r!}$ $\text{P}(X = 8) = \dfrac{e^{-8}8^8}{8!}$ $\text{P}(X = 7) = \dfrac{e^{-8}8^7}{7!}$ $\text{P}(X = 8) = \dfrac{8}{8} \times \text{P}(X = 7) = \text{P}(X = 7)$ or use recurrence relation $p_{x+1} = \dfrac{\lambda}{x+1} \times p_x$ $p_8 = \dfrac{8}{7+1} \times p_7$	M1 A1
b	Since the probabilities of 8 and 7 are equal, the distribution has 2 modes. These are modes because for $x < 8$, recurrence relation is increasing and for $x > 8$, it is decreasing.	B1
c	For a Poisson distribution $\mu = \sigma^2 = \lambda$. $\text{P}(X < \mu - \sigma) = \text{P}(X < 8 - \sqrt{8})$ $= \text{P}(X < 5.2)$ $= \text{P}(X \leqslant 5)$ since Poisson is discrete $= 0.191$ (3 dp) using tables	M1 M1 A1 **6**
4 a	$X \sim \text{Po}(9)$	B1
b	**i** Using tables $\text{P}(X = 5) = \text{P}(X \leqslant 5) - \text{P}(X \leqslant 4)$ $= 0.116 - 0.055$ $= 0.061$ (3 dp) **ii** $\text{P}(X > 6) = 1 - \text{P}(X \leqslant 6)$ $= 1 - 0.207$ $= 0.793$ (3 dp)	 M1 A1 M1 A1
c	**i** Y = number of customers in a random 10 minute period $Y \sim \text{Po}(18)$ $\text{P}(Y < 9) = \text{P}(Y \leqslant 8)$ $= 0.007$ (3 dp) using tables **ii** X = number of customers in a random 1 minute period $X \sim \text{Po}(1.8)$ $\text{P}(X = 1) = \dfrac{e^{-1.8}(1.8)^1}{1!}$ $= 0.2975$ (4 dp) Probability this happens 5 times $= (0.2975)^5$ $= 0.002$ (3 dp)	 M1 A1 M1 A1 M1 A1
d	This is less than the answer b(i) because there is only 1 pattern to achieve it, namely 1 – 1 – 1– 1 – 1 whereas the other answer allowed 1 – 1 – 1 – 1 – 1 1 – 2 – 0 – 1 – 1 0 – 5 – 0 – 0 – 0 etc	 B1 **12**

5 a

x	f	fx	fx^2
0	2	0	0
1	3	3	3
2	6	12	24
3	32	96	288
4	45	180	720
5	4	20	100
6	2	12	72
	94	323	1207

M3(table)

$\text{mean} = \frac{323}{94}$ — M1

$= 3.436$ (3 dp) — A1

$\text{variance} = \frac{1207}{94} - 3.436^2$ — M1

$= 1.033$ (3 dp) — A1

b

i In a Poisson distribution, the mean is equal to the variance. They are a poor match here. — B1

ii The number of cars passing is not truly random. The closure of one lane could mean the traffic has to move more slowly from to time, so you count fewer cars. — B1

9

6 a

You can use $X \sim \text{Po}(75)$ — B1

provided that each of the 15 minutes is independently P(5). This might not be the case if you are approaching rush hour, for instance, or a nearby school closing time. — B1

b

You can use $X \sim \text{Po}(19.2)$

provided that the whole area has an even spread of dandelions. — B1

This might not be the case if there is a variation in soil condition, for instance. — B1

c

You can use $X \sim \text{Po}(1050)$

provided that the radiation of particles doesn't vary much during the hour. — B1

This would probably be okay for a substance with a very long half-life, but not for a shorter one. — B1

6

7 a

X = number of bad apples in a bag — M1

$X \sim \text{B}(10, 0.03)$

i $\text{P}(X = 1) = 10 \times 0.03 \times 0.97^9$ — M1

$= 0.228$ (3 dp) — A1

ii $\text{P}(X > 1) = 1 - \text{P}(0) - \text{P}(1)$

$= 1 - 0.97^{10} - 0.228$ — M1

$= 1 - 0.737 - 0.228$ — M1

$= 0.035$ (3 dp) — A1

b

You could say that the average number of bad apples per bag is $10 \times 0.03 = 0.3$

Then if the 15 bags in the box are independent, there would be an average of $15 \times 0.3 = 4.5$ bad apples per box. — M1

Then use

$X \sim \text{Po}(4.5)$ for the distribution of number of bad apples in a box.

$\text{P}(X \leqslant 3) = 0.342$ (3 dp) from tables — A1

8

© Oxford University Press 2009

8 a	X = number of faulty components in a box $X \sim B(400, 0.001)$	B1
	i The probability of each component being faulty is independent, and equal to 0.001.	B1 B1
	ii It is reasonable to assume that the conditions are satisfied. (unless their is a suspicion of a bad batch caused by substandard materials)	B1
b	**i** For calculations use the approximation $X \sim Po(0.4)$	B1
	ii $P(X > 1) = 1 - P(X \leqslant 1)$ $= 1 - P(0) - P(1)$	M1
	$= 1 - \frac{e^{-0.4}(0.4)^0}{0!} - \frac{e^{-0.4}(0.4)^1}{1!}$	M1
	$= 1 - 0.6703 - 0.2681$ $= 0.062$ (3 dp)	A1
		8

3 Continuous random variables

1. A continuous random variable X has a probability density function of the form

$$f(x) = \begin{cases} A(x^2 + 1) & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Calculate the value of A.

(b) Find $\Pr\left(\frac{1}{2} < x \leqslant 1\right)$ correct to 2 decimal places.

(c) Calculate $E(X)$ and $Var(X)$.

2. A continuous random variable X has a probability density function of the form

$$f(x) = \begin{cases} \frac{x}{5} + C & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Calculate the value of C.

(b) Find $\Pr(1 \leqslant x \leqslant 2)$ correct to 2 decimal places.

(c) Calculate $E(X)$ and $Var(X)$.

3. A continuous random variable X has a probability density function of the form

$$f(x) = \begin{cases} k\left(1 - \frac{x}{4}\right) & 0 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show that $k = 0.5$

(b) Calculate the cumulative distribution function $F(x)$.

(c) Find $\Pr(1 \leqslant x \leqslant 2)$.

(d) Calculate $E(X)$ and $Var(X)$.

(e) Calculate $\Pr(X \leqslant \mu)$ where $\mu = E(X)$.

4. A random variable has a cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x < 0 \\ 4x^3 - 3x^4 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

Find

(a) the density function $f(x)$

(b) the mean and variance of X

(c) the mode of X.

5. A random variable X has probability density function given by

$$f(x) = \begin{cases} kx^3(4 - x) & 0 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Show that $k = \frac{5}{256}$, and find the mean and standard deviation of X.

© Oxford University Press 2009

6. At a supermarket, the time, T minutes, that customers have to wait in order to get to the checkout has the following cumulative distribution function.

$$F(t) = \begin{cases} 0 & t < 0 \\ \dfrac{t^2}{18} & 0 \leqslant t < 3 \\ \dfrac{1}{2} + \dfrac{1}{4}(t-3) & 3 \leqslant t \leqslant 5 \\ 1 & t > 5 \end{cases}$$

(a) Write down the values of $P(t \leqslant 2)$.

(b) Show that the median waiting time is 3 minutes.

(c) Find the probability that customers have to wait for less than 4 minutes in order to get to the checkout.

(d) Calculate the mean time that customers have to wait in order to get to the checkout.

7. (a) For the random variable with the probability density function given by

$$f(x) = \begin{cases} \dfrac{2}{9}x & 4 \leqslant x \leqslant 5 \\ 0 & \text{otherwise} \end{cases}$$

find (i) the median
(ii) the upper quartile
(iii) the mode.

(b) For the random variable with the probability density function given by

$$f(x) = \frac{4}{27}(3x^2 - x^3) \quad 0 \leqslant x \leqslant 3,$$

find (i) the mean
(ii) the mode.

© Oxford University Press 2009

3 Continuous random variables

Question Number	Solution	Marks
1 a	$f(x)=\begin{cases}A(x^2+1) & 0\leqslant x\leqslant 1\\ 0 & \text{elsewhere}\end{cases}$	
	$A\int_0^1 (x^2+1)\,dx = 1$	M1
	$A\left[\frac{x^3}{3}+x\right]_0^1 = 1$	M1
	$\frac{4A}{3}=1 \qquad A=\frac{3}{4}$	A1
b	$P\left(\frac{1}{2}<x\leqslant 1\right)=\int_{\frac{1}{2}}^1 \frac{3}{4}(x^2+1)\,dx$	M1
	$=\frac{3}{4}\left[\frac{x^3}{3}+x\right]_{\frac{1}{2}}^1$	M1
	$=\frac{3}{4}\left[\frac{4}{3}-\frac{13}{24}\right]$	
	$=\frac{19}{32}=0.59\ (2\text{ dp})$	A1
c	$E(X)=\frac{3}{4}\int_0^1 (x^3+x)\,dx$	M1
	$=\frac{3}{4}\left[\frac{x^4}{4}+\frac{x^2}{2}\right]_0^1$	M1
	$=\frac{3}{4}\left(\frac{1}{4}+\frac{1}{2}\right)=\frac{9}{16}$	A1
	$\text{Var}(X)=E(X^2)-E(X)^2$	M1
	$=\frac{3}{4}\int_0^1 (x^4+x^2)dx-\left(\frac{9}{16}\right)^2$	M1
	$=\frac{3}{4}\left[\frac{x^5}{5}+\frac{x^3}{3}\right]_0^1-\frac{81}{256}$	M1
	$=\frac{3}{4}\times\frac{8}{15}-\frac{81}{256}$	
	$=0.084\ (3\text{ dp})$	A1
		13

© Oxford University Press 2009

2 a	$f(x)=\begin{cases}\frac{x}{5}+C & 0\leqslant x\leqslant 2\\ 0 & \text{elsewhere}\end{cases}$	
	$\int_0^2\left(\frac{x}{5}+C\right)dx=1$	M1
	$\left[\frac{x^2}{10}+Cx\right]_0^2=1$	M1
	$0.4+2C=1$	
	$C=0.3$	A1
b	$P(1\leqslant X\leqslant 2)=\int_1^2\left(\frac{x}{5}+\frac{3}{10}\right)dx$	M1
	$=\left[\frac{x^2}{10}+\frac{3x}{10}\right]_1^2$	M1
	$=\frac{4}{10}+\frac{6}{10}-\left(\frac{1}{10}+\frac{3}{10}\right)$	
	$=0.6$	A1
c	$E(X)=\int_0^2\left(\frac{x^2}{5}+\frac{3x}{10}\right)dx$	M1
	$=\left[\frac{x^3}{15}+\frac{3x^2}{20}\right]_0^2$	M1
	$=\frac{8}{15}+\frac{12}{20}$	
	$=\frac{17}{15}$ OR 1.13 (2 dp)	A1
	$\text{Var}(X)=E(X^2)-E(X)^2$	M1
	$=\int_0^2\left(\frac{x^3}{5}+\frac{3x^2}{10}\right)dx-\left(\frac{17}{15}\right)^2$	M1
	$=\left[\frac{x^4}{20}+\frac{x^3}{10}\right]_0^2-\left(\frac{17}{15}\right)^2$	M1
	$=\frac{16}{20}+\frac{8}{10}-\left(\frac{17}{15}\right)^2$	
	$=0.316$ (3 dp)	A1
		13

© Oxford University Press 2009

3 a	$f(x)=\begin{cases} k\left(1-\frac{x}{4}\right) & 0\leqslant x\leqslant 4 \\ 0 & \text{elsewhere}\end{cases}$	
	$k\int_0^4\left(1-\frac{x}{4}\right)dx=1$	M1
	$k\left[x-\frac{x^2}{8}\right]_0^4=1$	M1
	$k[4-2]=1$	
	$k=0.5$	A1
b	Between 0 and 4	
	$F(x)=\frac{1}{2}\int_0^x\left(1-\frac{t}{4}\right)dt$	M1
	$=\frac{1}{2}\left[t-\frac{t^2}{8}\right]_0^x$	M1
	$=\frac{1}{2}\left(x-\frac{x^2}{8}\right)$ OR $\frac{x}{16}(8-x)$	M1
	$F(x)=\begin{cases} 0 & x<0 \\ \frac{x}{16}(8-x) & 0\leqslant x\leqslant 4 \\ 1 & x>4\end{cases}$	A1
c	$P(1\leqslant X\leqslant 2)=F(2)-F(1)$	M1
	$=\frac{2}{16}\times 6-\frac{1}{16}\times 7=\frac{5}{16}$	A1
d	$E(X)=\frac{1}{2}\int_0^4\left(x-\frac{x^2}{4}\right)dx$	M1
	$=\frac{1}{2}\left[\frac{x^2}{2}-\frac{x^3}{12}\right]_0^4$	M1
	$=\frac{1}{2}\left(8-\frac{64}{12}\right)=\frac{4}{3}$	A1
	$\text{Var}(X)=E(X^2)-E(X)^2$	M1
	$=\frac{1}{2}\int_0^4\left(x^2-\frac{x^3}{4}\right)dx-\left(\frac{4}{3}\right)^2$	M1
	$=\frac{1}{2}\left[\frac{x^3}{3}-\frac{x^4}{16}\right]_0^4-\left(\frac{4}{3}\right)^2$	M1
	$=\frac{1}{2}\left(\frac{64}{3}-16\right)-\frac{16}{9}=\frac{8}{9}$	A1
e	$P\left(X\leqslant\frac{4}{3}\right)=F\left(\frac{4}{3}\right)$	M1
	$=\frac{4}{3\times 16}\left(8-\frac{4}{3}\right)=\frac{5}{9}$	A1
		18

© Oxford University Press 2009

4	$F(x) = 4x^3 - 3x^4 \quad 0 \leqslant x \leqslant 1$	M1
a	$f(x) = F'(x)$ $= 12x^2 - 12x^3$	M1
	$f(x) = \begin{cases} 12x^2(1-x) & 0 \leqslant x \leqslant 1 \\ 0 & \text{elsewhere} \end{cases}$	A1
b	$E(X) = \int_0^1 12x^3(1-x)\,dx$	M1
	$= 12\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1$	M1
	$= 12\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{12}{20} = 0.6$	A1
	$\text{Var}(X) = E(X^2) - E(X)^2$	M1
	$= \int_0^1 12x^4(1-x)\,dx - 0.6^2$	M1
	$= 12\left(\frac{x^5}{5} - \frac{x^6}{6}\right)_0^1 - 0.36$	M1
	$= \frac{12}{30} - 0.36$ $= 0.04$	A1
c	mode is where $f'(x) = 0$	M1
	$f'(x) = 24x - 36x^2$	M1
	$24x - 36x^2 = 12x(2 - 3x) = 0$ $\Rightarrow x = 0$ or $\frac{2}{3}$	M1
	mode is at $x = \frac{2}{3}$	A1
		14

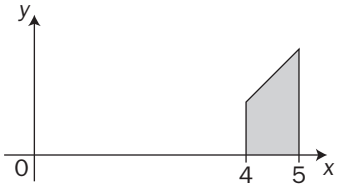
© Oxford University Press 2009

5	$f(x) = \begin{cases} kx^3(4-x) & 0 \leqslant x \leqslant 4 \\ 0 & \text{elsewhere} \end{cases}$	
	$\int_0^4 kx^3(4-x)\,dx = 1$	M1
	$k\left[x^4 - \frac{x^5}{5}\right]_0^4 = 1$	M1
	$k\left(256 - \frac{1024}{5}\right) = 1$	
	$k\left(\frac{1280-1024}{5}\right) = 1$	
	$k = \frac{5}{256}$	A1
	$E(X) = \frac{5}{256}\int_0^4 x^4(4-x)\,dx$	M1
	$= \frac{5}{256}\left[4\frac{x^5}{5} - \frac{x^6}{6}\right]_0^4$	M1
	$= \frac{5}{256}\left(\frac{4^6}{5} - \frac{4^6}{6}\right) = \frac{5\times 4^6}{4^4}\left(\frac{1}{5} - \frac{1}{6}\right)$	
	$= 5 \times 16 \times \frac{1}{30}$	
	$= \frac{8}{3}$	A1
	$\text{Var}(X) = E(x^2) - E(X)^2$	M1
	$= \frac{5}{256}\int_0^4 x^5(4-x)\,dx - \left(\frac{8}{3}\right)^2$	M1
	$= \frac{5}{256}\left(4\frac{x^6}{6} - \frac{x^7}{7}\right)_0^4 - \frac{64}{9}$	M1
	$= \frac{5\times 4^7}{4^4}\left(\frac{1}{6} - \frac{1}{7}\right) - \frac{64}{9}$	
	$= 5 \times 64 \times \frac{1}{42} - \frac{64}{9}$	
	$= \frac{32}{63}$	
	$= 0.5079$ (4 dp)	M1
	$\text{sd}(X) = \sqrt{\text{Var}(X)}$	M1
	$= 0.713$ (3 dp)	A1
		12

© Oxford University Press 2009

6	$f(t) = \begin{cases} 0 & t < 0 \\ \frac{t^2}{18} & 0 \leqslant t \leqslant 3 \\ \frac{1}{2} + \frac{1}{4}(t-3) & 3 \leqslant t \leqslant 5 \\ 1 & t > 5 \end{cases}$	
a	$P(T \leqslant 2) = F(2)$	M1
	$= \frac{2^2}{18}$	
	$= \frac{2}{9}$	A1
b	$P(T < 3) = \frac{3^2}{18}$	
	$= \frac{1}{2}$	
	hence median = 3 mins	B1
c	$P(T < 4) = \frac{1}{2} + \frac{1}{4}(4-3)$	M1
	$= \frac{3}{4}$	A1
d	$f(t) = \begin{cases} \frac{t}{9} & 0 \leqslant t \leqslant 3 \\ \frac{1}{4} & 3 \leqslant t \leqslant 5 \\ 0 & \text{elsewhere} \end{cases}$	M2
	Mean $= \int_0^3 \frac{t^2}{9}\,dt + \int_3^5 \frac{t}{4}\,dt$	M1
	$= \left[\frac{t^3}{27}\right]_0^3 + \left[\frac{t^2}{8}\right]_3^5$	M1
	$= 1 + \frac{1}{8}(25 - 9)$	
	$= 1 + \frac{1}{8} \times 16$	
	= 3 minutes	A1
		10

© Oxford University Press 2009

7	$f(x)=\begin{cases}\frac{2}{9}x & 4\leqslant x\leqslant 5\\ 0 & \text{elsewhere}\end{cases}$	
a	**i** $\int_4^M \frac{2}{9}x\,dx=\frac{1}{2}$	M1
	$\left[\frac{x^2}{9}\right]_4^M=\frac{1}{2}$	M1
	$\frac{1}{9}(M^2-16)=\frac{1}{2}$	
	$M^2=16+\frac{9}{2}=20.5$	
	$M=\sqrt{20.5}=4.528$ (3 dp)	A1
	(M cannot be –4.528, since f(x) = 0 in this region)	
	ii $\int_4^u \frac{2}{9}x\,dx=\frac{3}{4}$ $\left[\text{OR}\int_u^5 \frac{2}{9}x\,dx=\frac{1}{4}\right]$	M1
	$\left[\frac{x^2}{9}\right]_4^u=\frac{3}{4}$	M1
	$\frac{1}{9}(u^2-16)=\frac{3}{4}$	
	$u^2=16+\frac{27}{4}=22.75$	
	$u=\sqrt{22.75}=4.770$ (3 dp)	A1
	iii graph of f(x)	
		
	From the graph, mode = 5	M1 A1
b	$f(x)=\begin{cases}\frac{4}{27}(3x^2-x^3) & 0\leqslant x\leqslant 3\\ 0 & \text{elsewhere}\end{cases}$	
	i mean $=\frac{4}{27}\int_0^3(3x^3-x^4)\,dx$	M1
	$=\frac{4}{27}\left[\frac{3x^4}{4}-\frac{x^5}{5}\right]_0^3$	M1
	$=\frac{4}{27}\left(\frac{3^5}{4}-\frac{3^5}{5}\right)=\frac{4\times 3^5}{3^3}\left(\frac{1}{4}-\frac{1}{5}\right)$	
	$=4\times 9\times\frac{1}{20}=1.8$	A1

	ii $f'(x) = \frac{4}{27}(6x - 3x^2)$	M1
	$= \frac{4x}{9}(2 - x) = 0$ for mode	M1
	$x = 0$ or 2	M1
	but f(0) = 0 so mode is $x = 2$	A1 **15**

4 Sampling

1. The manager of an office wants some information on the number of staff who were absent each day during 2008. He records the number of staff who were absent on each day in February.

(a) Why is this sampling procedure likely to be biased?

(b) Suggest a better procedure he could use.

2. (a) Explain briefly what you understand by

(i) a population
(ii) a sampling unit.

(b) A market research organisation wants to take a sample of

(i) people living in Newcastle who are members of a fitness club
(ii) people living in London who were the victim of a burglary during July 2008.

Suggest a suitable sampling frame in each case and identify any difficulties with it.

3. The manager of a DVD rental store wishes to establish the number of times DVDs in the store are borrowed during the course of a year.

(a) Suggest a suitable sampling frame.

(b) Identify the sampling units.

(c) Explain *briefly* how the manager could take a sample for this purpose.

A new member of staff examined the DVDs returned by 100 people using the store one Saturday. She calculated that the mean number of times these DVDs had been borrowed over the last 12 months was 6.4. The manager said that the mean value for all the DVDs in the store is unlikely to be close to 6.4.

(d) State, giving a reason, whether you would expect the mean value for all the DVDs to be higher or lower than 6.4.

4. The committee of a golf club are discussing whether or not to incorporate a driving range as part of the club's facilities. They decide to ask a sample of members for their views.

(a) Suggest a suitable sampling frame.

(b) Explain the difference between a sample and a census, using this example to illustrate your answer.

(c) State why you think the committee might have decided to take a sample in this case rather than a census.

© Oxford University Press 2009

5. A computer takes tokens for 10 minutes and for 30 minutes. 30% of the tokens used are for 10 minutes. The tokens cost £1 for 10 minutes and £2 for 30 minutes.

(a) Find the mean μ for the amount of time per token.

(b) A random sample of 3 tokens is taken when the computer is emptied.

(i) List all possible samples.
(ii) Find the sampling distribution for the mean time per token.
(iii) Find the sampling distribution for the mode of the time per token.
(iv) Find the sampling distribution for the median of the time per token.

(c) If you were interested in the value of the tokens rather than the time they represented, which information from parts **a** and **b** could you write down immediately?

© Oxford University Press 2009

4 Sampling

Question Number	Solution	Marks
1 a	Absence varies with the time of year. There is more illness in winter. There are periods of school holidays when families are more likely to be away.	B1
b	Better to survey a week in February, a week in May, a week in August and a week in November, or some similar pattern over a whole year.	
	Best to use whole weeks since there is a well known pattern of illness on Mondays and Fridays.	B1
		2
2 a	i A population is all the people or things you are studying.	
	It could be all the people in London. It could be all the people in England. It could be all the people in the world.	
	You decide this when you set up your study.	B1
	ii A sampling unit is an individual member of the population.	B1
	Ideally you would study the whole population, but this is often impractical due to constraints of money, or time, for instance.	
b	i Select a fitness club at random, study all the members.	B1
	The members may all be from the same social background.	
	The club may cater exclusively for women, or youngsters, or some other particular section of the community.	B1
	Select a few members from each fitness club in Newcastle.	
	Now the problem is to ensure that you have enough variety in age and sex and social background from each club.	
	Also it will be more time consuming, since many clubs have to be visited. Other solutions will be combinations of the above.	
	ii Select an area at random, study all the victims [will need access to police databases].	B1
	The victims may all be from the same social background. The area may not be generally representative of London.	
	Select a few victims from each area of London [again access to police databases required].	B1
	You need to make sure you don't over-represent a particular section of the community, eg. single people or the elderly. It will be more time consuming, since all areas have to be visited.	
		6

© Oxford University Press 2009

3 a	Her sample frame could be (eg.) 50 DVDs, chosen from an alphabetical list of stock.	B1
b	The sample units are each of the selected DVDs.	B1
c	He/She could use: Every 10th DVD in her stocklist Each DVD which is currently at the left edge of a shelf A random selection of 50 from the stocklist A stratified sample by genre	B1
d	I would expect the average to be lower. By using the returned DVDs as the sample frame, she has probably got a biased sample which includes many popular titles which are taken out more often.	B1 **4**
4 a	The sampling frame is the alphabetical list of members In a census, every member of the club would be asked.	B1
b	In a sample, a smaller, but respresentative (hopefully) group of members would be asked	B1
c	A sample would be quicker to do. It may be quite difficult to contact every member of the club and get a reply. Opinions are more easily heard in smaller groups. etc	B1 **4**

© Oxford University Press 2009

5 a	mean time per token, μ $= 0.3 \times 10 + 0.7 \times 30$ $= 24$ mins	 M1 A1
b	**i** possible samples: 10 – 10 – 10 10 – 10 – 30 10 – 30 – 10 30 – 10 – 10 10 – 30 – 30 30 – 10 – 30 30 – 30 – 10 30 – 30 – 30	 B1
	ii sample mean for 10 – 10 – 10 is 10 probability $= 0.3 \times 0.3 \times 0.3 = 0.027$ similarly 10 – 10 – 30 is $16\frac{2}{3}$ probability $= 3 \times 0.3 \times 0.3 \times 0.7 = 0.189$ 10 – 30 – 30 is $23\frac{1}{3}$ probability $= 3 \times 0.3 \times 0.7 \times 0.7 = 0.441$ 30 – 30 – 30 is 30 probability $= 0.7 \times 0.7 \times 0.7 = 0.343$	A1 M2
	sample mean — probability 10 — 0.027 $16\frac{2}{3}$ — 0.189 $23\frac{1}{3}$ — 0.441 30 — 0.343	 A2
	Using above probabilities (adding first two, and second two) sample mode — probability 10 — 0.216 30 — 0.784	
	iii Similarly: sample median — probability 10 — 0.216 30 — 0.784	 M1 A1
c	You could write £1 and £2 instead of 10 mins and 30 mins in the distributions for the sample mode and median.	B1
		13

5 Continuous distributions

1. Of the widgets produced by a factory, 10% are defective.

(a) Find the probability that in a box of 30 widgets more than 2 are defective.

(b) Using a suitable approximation find the probability that in a box of 80 widgets there will be between 6 and 12 (inclusive) defective widgets.

2. The average number of cars passing a point on a dual carriageway in 10 seconds between 8 a.m. and 9 a.m. is 3.5.

(a) Suggest a suitable model for the number of cars passing this point on the dual carriageway in 10 seconds between 8 a.m. and 9 a.m. Give a suitable value for any parameter or parameters required.

(b) Calculate the probability that in a period of 10 seconds between 8 a.m. and 9 a.m., the number of cars passing the point is
(i) 4 (ii) less than 5.

(c) Using a suitable approximation, find the probability that during a randomly selected minute between 8 a.m. and 9 a.m. fewer than 15 cars pass the point.

Some roadworks are due to take place to improve access to the dual carriageway.

(d) Explain why the model may be no longer applicable while the alterations are taking place.

(e) Explain why the parameters for the distribution may be different once the work has been completed.

3. The continuous random variable X has a rectangular distribution with the following probability density function, where k is a constant.

$$f(x) = \begin{cases} \frac{1}{7} & 3 \leqslant x \leqslant k \\ 0 & \text{otherwise} \end{cases}$$

(a) State the value of k.

(b) Hence find values for the mean, μ, and the variance, σ^2, of X.

(c) Calculate $P(X > \mu + \sigma)$.

4. On a production line, on average 8% of the boxes of chocolate are not filled properly.

(a) If 5 boxes are examined, find the probability that exactly one of them is not filled properly.

(b) If 25 boxes are examined, use a Poisson approximation to find the probability that not more than 2 of them are not filled properly.

(c) If 3000 bottles are examined, use a Normal approximation to find the probability that more than 250 of them are not filled properly.

© Oxford University Press 2009

5. A mail order company monitor the number of orders for a particular item in their catalogue.

(a) During February the number of orders follows a Poisson distribution with mean 2.5 per day.

For a particular day, what is

(i) the probability that two or fewer orders are received
(ii) the probability that four or more orders are received
(iii) the largest number of orders which is exceeded with probability greater than 0.01?

(b) During June, the item is on special offer, and the number of orders received on a day follows a Poisson distribution with mean 5.6. Use a suitable approximation to estimate the probability that in a particular week (5 days) more than 30 orders are received.

6. Diameters of surgical needles are measured correct to the nearest 0.01 cm. X is the continuous random variable which represents the measurement error for one of these surgical needles.

(a) Give the probability density function for X.

(b) Find $Var(X)$.

(c) Find the lower quartile of X.

7. A city based physiotherapy practice treats 25 people following an operation on a particular knee injury over a six month period. Physiotherapists know from previous research that the probability of a patient recovering full mobility after this injury is 0.3.

(a) Assuming the patients recover independently of one another find the probability that

(i) at least 10 will recover full mobility
(ii) exactly 5 will recover full mobility.

In the whole city, 157 people have had this operation on a knee injury.

(b) Using a suitable approximation, find the probability that the number of people who will recover is between 40 and 55 inclusive.

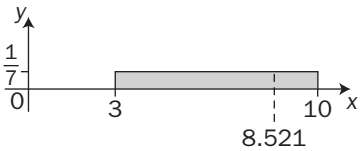
(c) Comment on the assumption of independence.

© Oxford University Press 2009

5 Continuous distributions

Question Number	Solution	Marks
1 a	X = number defective	M1
	$X \sim B(30, 0.1)$	
	$P(X > 2) = 1 - P(X \leqslant 2)$	
	$= 1 - 0.9^{30} - 30 \times 0.9^{29} \times 0.1$	
	$- \frac{30 \times 29}{2 \times 1} \times 0.9^{28} \times 0.1^2$	M2
	$= 1 - 0.0424 - 0.1413 - 0.2277$	
	$= 0.589$ (3 dp)	A1
b	mean $= np = 80 \times 0.1 = 8$	
	variance $= np(1 - p) = 80 \times 0.1 \times 0.9 = 7.2$	
	Use $X \sim N(8, 7.2)$	M2
	For $P(6 \leqslant X \leqslant 12) = P(5.5 < x < 12.5)$	
	using continuity correction	
	$Z_1 = \frac{5.5 - 8}{\sqrt{7.2}} = -0.93$	M1
	$P(X < 6) = 1 - 0.8238 = 0.1762$	A1
	$Z_2 = \frac{12.5 - 8}{\sqrt{7.2}} = 1.68$	M1
	$P(X \leqslant 12) = 0.9535$	A1
	$P(6 \leqslant X \leqslant 12) = 0.9535 - 0.1762$	M1
	$= 0.777$ (3 dp)	A1
		12
2 a	$X \sim Po(3.5)$	B1
b	**i** $P(X = 4)$	M1
	$= \frac{e^{-3.5} 3.5^4}{4!} = 0.189$ (3 dp)	A1
	ii $P(X < 5) = P(X \leqslant 4) = 0.725$ (3 dp) (tables)	A1
c	In 1 minute, mean $= 6 \times 3.5 = 21$	
	For Poisson, variance = mean	
	Use $X \sim N(21, 21)$	M2
	For $P(X < 15) = P(x < 14.5)$	
	using continuity correction	
	$z = \frac{14.5 - 21}{\sqrt{21}} = -1.42$	M1
	$P(X < 15) = 0.078$ (3 dp)	A1
d	The number of cars arriving every 10 seconds might no longer be random, e.g. due to traffic controls or queries.	B1
e	The average number of cars passing may change due to more cars using the improved route.	
	Faster speeds may be possible too.	B1
		10

© Oxford University Press 2009

3	$f(x)=\begin{cases}\frac{1}{7} & 3\leqslant x\leqslant k\\ 0 & \text{otherwise}\end{cases}$	
a	$k = 10$	B1
b	By symmetry (or otherwise)	
	$\text{mean} = \frac{10+3}{2} = 6.5$	M1 A1
	$\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2$	M1
	$= \int_3^{10} \frac{1}{7}x^2\,\text{d}x - 6.5^2$	M1
	$= \frac{1}{7}\left[\frac{x^3}{3}\right]_3^{10} - 42.25$	M1
	$= \frac{1}{21}(1000-27)-42.25$	
	$= 4.083$ (3 dp)	A1
c	$\sigma = \sqrt{4.083} = 2.021$	M1
	$\text{P}(X > \mu + \sigma) = \text{P}(X > 6.5 + 2.021)$	
	$= \text{P}(X > 8.521)$	
	graph of f(x)	
		
	Using graph or otherwise	M2
	$\text{P}(X > \mu + \sigma) = \frac{10-8.521}{7} = 0.211$ (3 dp)	A1
		11
4 a	X = number of boxes not filled properly	
	$X \sim \text{B}(5, 0.08)$	M1
	$\text{P}(X = 1) = 5 \times 0.92^4 \times 0.08$	M1
	$= 0.287$ (3 dp)	A1
b	For 25 boxes, mean = $np = 25 \times 0.08 = 2$	M1
	Also variance = $np(1-p) = 25 \times 0.08 \times 0.92 = 1.84$	
	which is close to mean, so use $X \sim \text{Po}(2)$	M1
	From tables	
	$\text{P}(X \leqslant 2) = 0.677$ (3 dp)	A1
c	For 3000 boxes, mean = $3000 \times 0.08 = 240$	M1
	variance = $240 \times 0.92 = 220.8$	M1
	Use N(240, 220.8)	M1
	For $\text{P}(X > 250) = \text{P}(X > 250.5)$	
	Using continuity correction	
	$Z = \frac{250.5-240}{\sqrt{220.8}} = 0.71$	M1
	$\text{P}(X > 250) = 0.239$ (3 dp)	A1
		11

© Oxford University Press 2009

5	$X \sim \text{Po}(2.5)$	
a	**i** From tables $P(X \leqslant 2) = 0.544$ (3 dp)	A1
	ii From tables	
	$P(X \geqslant 4) = 1 - P(X \leqslant 3)$	M1
	$= 1 - 0.758$	
	$= 0.242$ (3 dp)	A1
	iii From tables	
	$P(X \leqslant 6) = 0.986$	
	So $P(X > 6) = 0.014 > 0.01$	
	$P(X \leqslant 7) = 0.996$	
	So $P(X > 7) = 0.004 < 0.01$	M2
	So the largest number of orders which is exceeded with probability more than 0.01 is 6.	A1
b	$X \sim \text{Po}(5.6)$	
	In 5 days mean = variance = $5 \times 5.6 = 28$	
	Use N(28, 28)	M2
	For $P(X > 30) = P(X > 30.5)$	
	using continuity correction	
	$Z = \dfrac{30.5 - 28}{\sqrt{28}}$	M1
	$= 0.47$	
	$P(X > 30) = 0.319$ (3 dp)	A1
		10
6 a	$f(x) = \begin{cases} 100 & -0.005 \leqslant x \leqslant 0.005 \\ 0 & \text{elsewhere} \end{cases}$	B2
b	By symmetry, mean = 0	M1
	$\text{Var}(X) = E(X^2) - E(X)^2$	M1
	$= \int_{-0.005}^{0.005} 100x^2 \, dx - 0^2$	M1
	$= \dfrac{100}{3}\left[x^3\right]_{-0.005}^{0.005}$	M1
	$= \dfrac{100}{3}(0.000\,000\,125 - -0.000\,000\,125)$	
	$= 0.0\,000\,083$ (2 sf)	A1
c	Lower quartile (by symmetry)	M1
	$= -0.0025$	A1
		9

© Oxford University Press 2009

7	X = number who recover $X \sim B(25, 0.3)$	 M1
a	**i** For $P(X \geqslant 10)$ $P(X \geqslant 10) = 1 - P(X \leqslant 9)$ $= 1 - 0.8106$ using Binomial tables $= 0.189$ (3 dp)	M2 M1 A1
	ii $P(X = 5) = \binom{25}{5} \times 0.3^5 \times 0.7^{20}$ $= 0.103$ (3 dp)	M1 A1
b	mean $= 157 \times 0.3 = 47.1$ variance $= 47.1 \times 0.7 = 32.97$ Use $X \sim N(47.1, 32.97)$ For $P(40 \leqslant X \leqslant 55) = P(39.5 < X < 55.5)$ using continuity correction	 M2
	$Z_1 = \dfrac{39.5 - 47.1}{\sqrt{32.97}} = -1.32$	M1
	$P(X < 40) = 0.0934$	A1
	$Z_2 = \dfrac{55.5 - 47.1}{\sqrt{32.97}} = 1.46$	M1
	$P(X \leqslant 55) = 0.9279$ $P(40 \leqslant X \leqslant 55) = 0.9279 - 0.0934$ $= 0.835$ (3 dp)	A1 M1 A1
c	It might be suspect if there were, for instance, 2 doctors performing these operations. The probability of success might then depend on which doctor performed the operation. Then you lose independence.	B1
		16

© Oxford University Press 2009

6 Hypothesis testing

1. The number of births between 6 a.m. and 6 p.m. in a large maternity unit may be modelled by a Poisson distribution with mean of 7.5 births per day.

(a) (i) Find the probability that a particular day has 6 births between 6 a.m. and 6 p.m.

(ii) Use an appropriate distributional approximation to find the probability that the number of births between 6 a.m. and 6 p.m. during a particular week exceeds 40.

(b) The number of births between 6 p.m. and 6 a.m. could also be modelled by a Poisson distribution, but the average rate could differ from the daytime rate. During the period between 6 p.m. and 6 a.m. on this particular day there were 3 births.

Test, at the 5% level of significance, whether the arrival rate between 6 p.m. and 6 a.m. is lower than the rate during the daytime.

2. A mill produces cloth in 50-metre lengths. It is common for lengths to contain faults that have to be treated before the cloth is sold. These faults are distributed over the cloth independently, at random and at a constant average rate.

(a) Name a distribution that will provide a suitable model for the number of faults in a 50-metre length of cloth.

(b) A new manager states that it is unacceptable for the mean number of faults per 50-metre length of cloth to exceed 1.5. The next 50-metre length of cloth produced contains 4 faults.
Carry out a hypothesis test, using the 5% significance level, to test the hypothesis that the mean number of faults per 50-metre length of cloth does not exceed 1.5.

(c) The results of the test are presented to the new manager, who states that a significance level of 5% is too high and that a level of 1% should be used in order to reduce the possibility of error.

Without any further calculations, comment on the new manager's statement. [(c) Edexcel Limited 1998]

3. A manufacturer of fruit sweets decides to introduce a new flavour to its selection. The new flavour makes up 15% of the total production. The sweets are sold in mixed flavour packets of 20. A child buys a packet of these sweets and discovers that it contains only one sweet of the new flavour.

(a) Stating your hypotheses clearly, test at the 5% level of significance whether or not there is evidence that the percentage of the new flavour in a packet is less than 15%.

(b) State, giving a reason, the number of the new flavour in a packet that would be required to obtain a significant result to the test in part **a**.

(c) State an assumption that has been made about how the packets are filled with different flavours of sweets in order to carry out the test in part **a**.

Another child also buys a packet of sweets and it also only has one sweet of the new flavour.

(d) By combining the results of the two children, test at the 5% level of significance whether or not the combined results provide evidence that the percentage of the new flavour is less than 15.

(e) Comment briefly on your answers to the tests in parts **a** and **d**.

© Oxford University Press 2009

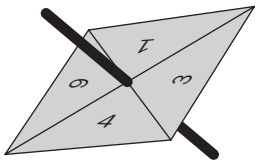
4. A carpet manufacturer has used a process which produced flaws in the carpet randomly at a rate of 0.4 per m^2. In an attempt to reduce the number of flaws produced, a new process is tried out. A randomly chosen section of carpet produced using this new process has an area of $8\,m^2$ and contains only 1 flaw.

(a) Stating your hypotheses clearly, test at the 10% level of significance whether or not the rate of occurrence of flaws using the new procedure has decreased.

The new procedure actually produces flaws at a rate of 0.2 per m^2.

(b) Find the probability that, using the test in part **a**, you would not reject the null hypothesis.

5.



A square spinner for use in a child's game has the four numbers 1, 3, 4 and 6 on its edges as shown. The manufacturer claims that the probability that the spinner lands on each of the four numbers is $\frac{1}{4}$. Before a box of spinners leaves the factory a randomly chosen spinner is tested by spinning it 25 times and recording the number of fours.

(a) Using a 5% significance level, find the critical regions for a two-tailed test of the hypothesis that the probability of the spinner landing on a four is $\frac{1}{4}$.

(b) Explain how a spinner could pass this test but still not satisfy the manufacturer's claim.

(c) Suggest an alternative test that the manufacturer could use to refine the testing procedure.

6. An official of the Department of Transport was studying the distribution of the number of vehicles in a queue at a set of temporary traffic lights. She modelled the number of vehicles as a Poisson random variable. She wished to test the null hypothesis that the parameter of the distribution was 6 against the alternative hypothesis that it was greater than 6. She intended to make one observation, and decided that the critical region would be 12 or greater.

(a) Determine the probability of a Type I error.

In fact, the value of the parameter was 10.

(b) Determine the probability of a Type II error.

She later changed her mind and decided that her alternative hypothesis should be that the parameter was not equal to 6. She decided to take the largest critical region for which the probability in each "tail" was less than 2.5%.

(c) Find the critical region which she should choose.

(d) Determine the actual significance level she would be using for the critical region.

© Oxford University Press 2009

7. A large college introduced a new procedure to try and ensure that staff arrived on time for the start of lectures. A recent survey by the students had suggested that in 15% of cases the staff arrived late for the start of a lecture. In the first week following the introduction of this new procedure a random sample of 35 lectures was taken and in only 1 case did the member of staff arrive late.

(a) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence that the new procedure has been successful.

A student complained that this sample did not give a true picture of the effectiveness of the new procedure.

(b) Explain briefly why the student's claim might be justified and suggest how a more effective check on the new procedure could be made.

© Oxford University Press 2009

6 Hypothesis testing

Question Number	Solution	Marks
1 a	$X \sim \text{Po}(7.5)$	
	i $P(X = 6) = P(X \leqslant 6) - (X \leqslant 5)$	
	$= 0.378 - 0.241$	M1
	$= 0.137$ (3 dp)	A1
	ii Average number of births in a week between 6 a.m. & 6 p.m.	
	$= 7.5 \times 7 = 52.5$	
	use N(52.5, 52.5)	M2
	For $P(X > 40) = P(x > 40.5)$	
	using continuity correction	
	$Z = \dfrac{40.5 - 52.5}{\sqrt{52.5}} = -1.66$	M1
	$P(X > 40) = 1 - 0.9515 = 0.049$ (3 dp)	A1
b	H_0: mean = 7.5	B1
	H_1: mean < 7.5	
	Under H_0 (from tables)	
	$P(X \leqslant 3) = 0.059$ (3 dp)	
	$= 5.9\%$	M1
	At the 5% level there is insufficient evidence to reject H_0.	
	The day rate is the same as the night rate.	A1
		9
2 a	Poisson distribution	B1
b	Using $X \sim \text{Po}(\lambda)$	
	H_0: $\lambda = 1.5$	
	H_1: $\lambda > 1.5$	M1
	Under H_0, from tables	
	$P(X \geqslant 4) = 1 - P(X \leqslant 3)$	
	$= 1 - 0.934$	M1
	$= 0.066$ (3 dp)	A1
	$= 6.6\%$	
	Insufficient evidence to reject H_0. There are 1.5 faults per 50 m of cloth on average.	A1
c	As far as the test goes, you can still not reject H_0.	B1
	But what does the manager mean by 'reducing the error'?	
	By reducing the significance level from 5% to 1%, you are making it harder to reject H_0.	
	So you are reducing the chance of rejecting H_0 when it is correct (Type I error), but increasing the chance of not rejecting H_0 when it is wrong (Type II error)	B1
		7

© Oxford University Press 2009

3 a	X is number of blue sweets $X \sim B(20, p)$ H_0: $p = 0.15$ H_1: $p < 0.15$ Under H_0, from tables $P(X \leqslant 1) = 0.176$ (3 dp) $= 17.6\%$ At the 5% level, insufficient evidence to reject H_0	 M2 M1 A1
b	$P(X = 0) = 0.039$ (3 dp) $= 3.9\%$ If there were no blue sweets you would reject H_0 at the 5% level	M1 A1
c	Assume that the blue sweets are randomly distributed among the packets	B1
d	Y is number of blue sweets in 2 packets Then $Y \sim B(40, 0.15)$ if $P = 0.15$ is true So $P(X \leqslant 2) = 0.0486$ $= 4.86\%$ At the 5% level there is insufficient evidence to reject H_0	 M2 M1 M1
e	Since they give different results, more importance should be given to the second since more packets of sweets were examined.	B1 **15**
4 a	X = number of faults $X \sim Po(\lambda)$ If process makes no difference you would get an average of $8 \times 0.4 = 3.2$ faults in $8\,m^2$ H_0: $\lambda = 3.2$ H_1: $\lambda < 3.2$ Under H_0 $P(X \leqslant 1) = P(0) + P(1)$ $= e^{-3.2} + 3.2e^{-3.2}$ $= 0.171$ (3 dp) $= 17.1\%$ At the 10% level, insufficient evidence to reject H_0. The new process has made no difference	 M2 M1 A1 A1
b	Under H_0 $P(X = 0) = e^{-3.2} = 0.041$ $= 4.1\%$ $P(X \leqslant 1) = 17.1\%$ So $X = 0$ is the only value, at the 10% level, which would make us reject H_0 Under H_1 $\lambda = 8 \times 0.2 = 1.6$ $P(X \geqslant 1) = 1 - P(X = 0)$ $= 1 - e^{-1.6}$ $= 1 - 0.202$ (3 dp) $= 0.798$ (3 dp)	 M1 M1 M1 M1 A1 **11**

© Oxford University Press 2009

5 a	X = number of sixes H_0: $X \sim B(25, 0.25)$ Need 2.5% in each tail From tables $P(X \leqslant 1) = 0.7\%$ $P(X \geqslant 12) = 1.07\%$ The critical region is $X \leqslant 1$ and $X \geqslant 12$ These values cause H_0 to be rejected.	 M1 B2
b	It may be that the other numbers do not occur with probability 0.25	B1
c	Perform the same test with each number.	B1 **7**
6	H_0: $\lambda = 6$ H_1: $\lambda > 6$ $\quad X \sim Po(\lambda)$	
a	Under H_0 from tables $P(X \geqslant 11) = 1 - P(X \leqslant 10)$ $= 0.0426$ $= 4.26\%$ This is the probability of rejecting H_0 when it is true. P(Type I error) = 4.26%	 M1 M1 A1
b	Under $X \sim Po(9)$ $P(X \leqslant 10) = 0.7060$ This is the probability of not rejecting H_0 when it is false. P(Type II error) = 0.7060	 M1 A1
c	From tables $P(X \leqslant 1) = 1.74\%$ $P(X \geqslant 12) = 1 - P(X \leqslant 11) = 2.01\%$ The critical region is $X \leqslant 1$ and $X \geqslant 12$	 M2 B2
d	Since $P(X \leqslant 1) = 0.7\%$ and $P(X \geqslant 12) = 2.01\%$ she would be using a 3.75% significance level.	 M1 A1 **12**
7 a	X = number not switched off in a sample of 30 $X \sim B(30, p)$ H_0: $p = 0.15$ H_1: $p < 0.15$ $P(X \leqslant 1) = 0.0480$ $= 4.8\%$ At the 5% level, reject H_0. The new procedure has been successful	 M1 M1 A1 A1
b	People try harder when they know they are being watched. Wait a while until sampling, and don't take all the samples in a cluster, when people might notice.	B1 B1 **9**