1. A chocolate manufacturer places special tokens in 2% of the bars it produces so that each bar contains at most one token. Anyone who collects 3 of these tokens can claim a prize.

Andreia buys a box of 40 bars of the chocolate.

(a) Find the probability that Andreia can claim a prize.

(2)

Barney intends to buy bars of the chocolate, one at a time, until he can claim a prize.

(b) Find the probability that Barney can claim a prize when he buys his 40th bar of chocolate.

(3)

(c) Find the expected number of bars that Barney must buy to claim a prize.

(1)

Solution 1a

Let X = no. of tokens Andreia wins

$$X \sim \text{Bin}(\overset{n=40}{40}, \underset{\text{probability bar has a token}}{0.02})$$

$$P(X \geq 3) = 1 - P(X \leq 2)$$

Andreia wins prize if more than or equal to 3 tokens are won.

$$= 1 - P(X=0) - P(X=1) - P(X=2)$$

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad q = 1-p$$

$$= 1 - 0.98^{40} - 40 \times 0.02 \times 0.98^{39} - \frac{40 \times 39}{2} \times 0.02^2 \times 0.98^{38}$$

$$= 1 - 0.4457 - 0.36384 - 0.14479$$

$$= 0.0457$$

Solution 1b (Method 1)

Let Y = no. of the bar when Barney wins

$$Y \sim NB(3, 0.02)$$

$$P(Y=40) = \binom{39}{2} \underbrace{0.02^2 \times 0.98^{37}}_{\text{37 failures \& two successes probability}} \times 0.02$$

$\binom{39}{2}$ ways of arranging 37 failures and two successes

probability of success on 40th purchase

Solution 1b (Method 2)

Let Y = no. of the bar when Barney wins

$Y \sim NB(3, 0.02)$

Substitute values in the formula

$$P(Y=y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$$

to give

$$P(Y=40) = \binom{40-1}{3-1} 0.02^3 \times (1-0.02)^{40-3}$$

$$= \binom{39}{2} 0.02^3 \times 0.98^{37}$$

$$= 0.0281$$

Negative Binomial Probability Function:

If $X \sim NB(r, p)$ then X can take values $r, r+1, r+2, \ldots$ with probability

$$P(X=x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Solution 1c

$Y \sim NB(3, 0.02)$

$$\Rightarrow E(Y) = \frac{3}{0.02} = 150$$

If $X \sim NB(r, p)$ then

Mean (or expected value) $= \mu = E(X) = \frac{r}{p}$

2. Indre works on reception in an office and deals with all the telephone calls that arrive. Calls arrive randomly and, in a 4-hour morning shift, there are on average 80 calls.

(a) Using a suitable model, find the probability of more than 4 calls arriving in a particular 20-minute period one morning.

(3)

Indre is allowed 20 minutes of break time during each 4-hour morning shift, which she can take in 5-minute periods. When she takes a break, a machine records details of any call in the office that Indre has missed.

One morning Indre took her break time in 4 periods of 5 minutes each.

(b) Find the probability that in exactly 3 of these periods there were no calls.

(2)

On another occasion Indre took 1 break of 5 minutes and 1 break of 15 minutes.

(c) Find the probability that Indre missed exactly 1 call in each of these 2 breaks.

(3)

Solution 2a

In 4 hours, average of 80 calls.

$\therefore$ in 20 mins, average of $\frac{20}{3}$ calls

$\Rightarrow$ If C = no. of calls in a 20 min period,

$C \sim Po\left(\frac{20}{3}\right)$

4 hours = 240 mins
To get 20 mins, divide 240 by 12
$\frac{240}{12}$ mins $\Rightarrow \frac{80}{12}$ calls $= \frac{20}{3}$

$P(C>4) = 1 - \underbrace{P(C \leq 4)} = 0.79437$

use your calculator to find this value
You will need to input $\lambda = \frac{20}{3}$ and $x = 4$

Solution 2b

Let X = no. of 5 minute periods with no calls.

$X \sim B(4, \boxed{e^{-5/3}})$

If F = no. of calls in a 5 min period, then $F \sim Po\left(\frac{20}{3\times 4}\right) \Rightarrow F \sim P\left(\frac{5}{3}\right)$

$\Rightarrow P(F=0) = \frac{e^{-5/3}\left(\frac{5}{3}\right)^0}{0!} = e^{-5/3}$

(We have used: if $X \sim Po(\lambda)$ then $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$)

If $X \sim B(n,p)$
$\Rightarrow P(X=x) = \binom{n}{x}p^x(1-p)^{n-x}$

$\Rightarrow P(X=3) = \binom{4}{3}\left(e^{-5/3}\right)^3\left(1-e^{-5/3}\right)^1$

$\Rightarrow P(X=3) = 0.021861$

Solution 2c

$X \sim Po(t^5/_3)$ $\quad$ $Y \sim Po(5)$

P(Indre missed exactly one call in each of the 2 breaks)

$= \boxed{P(X=1)} \times \boxed{P(Y=1)}$

$= \boxed{\left(e^{-5/3} \times \left(5/3\right)\right)} \times \boxed{\left(e^{-5} \times 5\right)}$

$= 0.0106$

If $X \sim Po(\lambda)$ then

$P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

3. A biased spinner can land on the numbers 1, 2, 3, 4 or 5 with the following probabilities.

Number on spinner	1	2	3	4	5
Probability	0.3	0.1	0.2	0.1	0.3

The spinner will be spun 80 times and the mean of the numbers it lands on will be calculated. Find an estimate of the probability that this mean will be greater than 3.25

(6)

Solution 3

Let X = no. when spinner is spun

By symmetry in the table, $\mu = E(X) = 3$

Alternatively,

$$E(X) = \sum x_i p_i = (1\times 0.3) + (2\times 0.1) + (3\times 0.2) + (4\times 0.1) + (5\times 0.3)$$

$$= 0.3 + 0.2 + 0.6 + 0.4 + 1.5$$

$$= 3$$

$$E(X^2) = \sum x_i^2 p_i$$

$$= (1^2\times 0.3) + (2^2\times 0.1) + (3^2\times 0.2) + (4^2\times 0.1) + (5^2 + 0.3)$$

$$= 0.3 + 0.4 + 1.8 + 1.6 + 7.5$$

$$= 11.6$$

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

$$= 11.6 - 3^2 = 2.6$$

$$\bar{X} \approx \sim N\left(3, \left(\sqrt{\frac{2.6}{80}}\right)^2\right)$$

$$P(\bar{X} > 3.25) = 0.0827589\ldots$$

use calculator with $x = 3.25$, $\mu = 3$, $\sigma = \sqrt{\frac{2.6}{80}}$

4. Liam and Simone are studying the distribution of oak trees in some woodland. They divided the woodland into 80 equal squares and recorded the number of oak trees in each square. The results are summarised in Table 1 below.

Number of oak trees in a square	0	1	2	3	4	5	6	7 or more
Frequency	1	4	21	23	13	11	7	0

Table 1

Liam believes that the oak trees were deliberately planted, with 6 oak trees per square and that a constant proportion p of the oak trees survived.

(a) Suggest the model Liam should use to describe the number of oak trees per square.

(2)

Liam decides to test whether or not his model is suitable and calculates the expected frequencies given in Table 2.

Number of oak trees in a square	0 or 1	2	3	4	5	6
Expected frequency	5.53	14.89	24.26	22.24	10.87	2.21

Table 2

(b) Showing your working clearly, complete the test using a 5% level of significance. You should state your critical value and conclusion clearly.

(7)

Simone believes that a Poisson distribution could be used to model the number of oak trees per square. She calculates the expected frequencies given in Table 3.

Number of oak trees in a square	0 or 1	2	3	4	5	6 or more
Expected frequency	12.69	16.07	s	14.58	t	9.37

Table 3

(c) Find the value of s and the value of t, giving your answers to 2 decimal places.

(4)

(d) Write down hypotheses to test the suitability of Simone's model.

(1)

The test statistic for this test is 8.749

(e) Complete the test. Use a 5% level of significance and state your critical value and conclusion clearly.

(3)

(f) Using the results of these tests, explain whether the origin of this woodland is likely to be cultivated or wild.

(2)

Solution 4a

T = no. of oak trees in a square

$T \sim B(6, p)$ ← This is what is required for the question. Below is additional information student may find interesting.

Extra ↓

$$p = \frac{\text{Total no. of oaks}}{\text{Total no. planted}}$$

$$= \frac{(0\times 1) + (1\times 4) + (2\times 21) + (3\times 23) + (4\times 13) + (5\times 11) + (6\times 7)}{80 \times 6}$$

$$= \frac{264}{560} = 0.55 \Rightarrow p = 0.55$$

So $T \sim B(6, 0.55)$

The question gives the expected frequencies. But, if student was required to find them, below is an explanation:

If $X \sim B(n,p)$ then $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

$E_i = N \times P(X = x_i)$ [N = Total no. of observations]

$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$ [n = no. of pairs of values]

To calculate χ^2, all expected frequencies must be at least 5 so you may have to combine some of the E_i.

$T \sim B(6, 0.55)$

No. of oaks in a square	0 or 1	2	3	4	5 or 6
Observed	5	21	23	13	18
Expected	$\boxed{5.53}$	$\boxed{14.89}$	$\boxed{24.26}$	$\boxed{22.24}$	$\boxed{13.08}$

$E_1 = 80 \times P(T=0) = 80 \times (1-0.55)^6 = \boxed{0.66430}$, $E_7 = 80 \times 0.55^6 = \boxed{2.2145}$

$E_1 = 80 \times P(T=1) = 80 \times 6 \times 0.55 \times (1-0.55)^5 = \boxed{4.87154}$ 0 or 1

$E_2 = 80 \times P(T=1) = 80 \times \binom{6}{2} \times 0.55^2 \times (1-0.55)^4 = \boxed{14.8853}$

$E_3 = 80 \times P(T=2) = 80 \times \binom{6}{3} \times 0.55^3 (1-0.55)^3 = \boxed{24.2575}$

$E_4 = 80 \times P(T=3) = 80 \times \binom{6}{4} \times 0.55^4 (1-0.55)^2 = \boxed{22.2360}$

$E_5 = 80 \times P(T=4) = 80 \times 6 \times 0.55^5 (1-0.55) = \boxed{10.8709}$ 5 or 6

$E_6 = 80 \times P(T=5)$

↑ Extra

Solution 4b

No. of oaks in a square	0 or 1	2	3	4	5 or 6
Observed O_i	5	21	23	13	18
Expected E_i	5.53	14.89	24.26	22.24	13.08
$\frac{(O_i-E_i)^2}{E_i}$	0.051	2.51	0.0654	3.84	1.85
$\frac{O_i^2}{E_i}$	4.521	29.617	21.805	7.599	24.771

$\sum \frac{(O_i-E_i)^2}{E_i}$

Since p needs to be estimated, $(\hat{p}=0.55)$ we have degrees of freedom $\nu = 5-2 = 3$

Also $\boxed{\chi^2 = \sum \frac{(O_i-E_i)^2}{E_i} = 8.313}$

But critical value $\boxed{\chi^2 = 7.815}$ $(\boxed{\nu = 3})$

(Use χ^2 tables

ν	0.995	----	0.100	0.050	---
1	0.000	---	---	---	---
2	0.010	---	4.605	5.991	---
$\boxed{3}$	0.072	---	6.251	7.815	---)

Since $\boxed{\chi^2 = 8.313} > \boxed{7.815 = \chi^2}$, we have a significant result.

Therefore there is sufficient evidence at the 5% to suggest that the model is not suitable.

Solution 4c

$$\lambda = \frac{\sum fx}{\sum f} = \frac{(1\times 0)+(4\times 1)+(21\times 2)+(23\times 3)+(13\times 4)+(11\times 5)+(7\times 6)}{1+4+21+23+13+11+7}$$

$\Rightarrow \lambda = 3.3$

Let R = no. of oak trees in a square for Simone's model

$\Rightarrow R \sim Po(\boxed{3.3})$

$$P(X=x) = \frac{e^{-\boxed{\lambda}}\boxed{\lambda}^{\boxed{x}}}{x!} \quad \text{where } X \sim Po(\lambda)$$

Therefore

$$\boxed{P(\boxed{X=3})} = \frac{e^{-\boxed{3.3}} \times \boxed{3.3}^{\boxed{3}}}{\boxed{3}!} = \boxed{0.220912\ldots}$$

But since $\boxed{N=80}$ (total planted),

$$S = \boxed{N} \times \boxed{P(X=3)} = \boxed{80} \times \boxed{0.220912}$$

$\Rightarrow S = 17.67$

Also,

$$\boxed{P(\boxed{X=5})} = \frac{e^{-\boxed{3.3}} \times \boxed{3.3}^{\boxed{5}}}{\boxed{5}!} = \boxed{0.120286\ldots}$$

But since $\boxed{N=80}$,

$$t = \boxed{N} \times P(X=5) = \boxed{80} \times \boxed{0.120286}$$

$\Rightarrow t = 9.62$

Solution 4d

H_0: Poisson is a good fit (for no. of oak trees per square)

H_1: Poisson is not a good fit (for no. of oak trees per square)

Solution 4e

Test statistic is 8.749

Extra Start

The question gives the test statistic. But, if student was required to find it, below is an explanation

No. of oaks per square	0 or 1	2	3	4	5	6 or more
Expected frequency E_i	12.69	16.07	17.67	14.58	9.62	9.37
Observed O_i	5	21	23	13	11	7
$\frac{(O_i - E_i)^2}{E_i}$	4.660	1.512	1.608	0.171	0.197	0.599

Test statistic $= 4.660 + 1.512 + 1.608 + 0.171 + 0.197 + 0.599$

$= 8.748$ (rounding error)

Extra End

λ needed to be estimated

Degrees of freedom $= 6 - 2 = 4$

ν	---	0.100	0.050	--
1	--	---	---	--
2				
3	--	6.251	7.815	--
4	--	7.779	9.488	--

Critical value is $\boxed{9.488}$

Since Test statistic < Critical Value, this is not a significant result. Therefore, at the 5% level, there is ~~evidence to suggest that the model is suitable~~ insufficient evidence to reject H_0

do not reject H_0

Solution 4f

Poisson model has better fit. Suggests oaks occur at random.

Can also say: Binomial suggests deliberately planted

Therefore, forest likely to be wild not cultivated

5. Information was collected about accidents on the *Seapron* bypass. It was found that the number of accidents per month could be modelled by a Poisson distribution with mean 2.5

Following some work on the bypass, the numbers of accidents during a series of 3-month periods were recorded. The data were used to test whether or not there was a change in the mean number of accidents per month.

(a) Stating your hypotheses clearly and using a 5% level of significance, find the critical region for this test. You should state the probability in each tail.

(5)

(b) State P(Type I error) using this test.

(1)

Data from the series of 3-month periods are recorded for 2 years.

(c) Find the probability that at least 2 of these 3-month periods give a significant result.

(3)

Given that the number of accidents per month on the bypass, after the work is completed, is actually 2.1 per month,

(d) find P(Type II error) for the test in part (a)

(3)

Solution 5a

$H_0 : \mu = 7.5$

$H_1 : \mu \neq 7.5$ $\qquad X \sim Po(7.5)$

$P(X \leqslant 2) = \boxed{0.0203}$

$P(X \leqslant 13) = 0.9784 \Rightarrow P(X \geqslant 14) = \boxed{0.0216}$

This gives a critical region of $X \leqslant 2$ and $X \geqslant 14$.

$\lambda =$	...	7.5	...
$x=0$	---	0.0006	---
1	---	0.0047	---
2	---	0.0203	---
3	---	0.0591	---
4	---	0.1321	---
5		---	
6		---	
7		---	
8		---	
9		---	
10		---	
11		---	
12	---	0.9573	---
13	---	0.9784	---

Solution 5b

P(Type I Error)

$= \boxed{0.0203} + \boxed{0.0216}$

$= 0.0419$

Solution 5c

Let M = no. of 3 month periods with a significant result

$M \sim B(8, 0.0419)$

7.5 rounded up to the nearest whole number

result from part b

$P(M \geq 2) = 1 - P(M \leq 1)$

$= 1 - P(M=0) - P(M=1)$

$= 1 - (1-0.0419)^8 - 8(1-0.0419)^7(0.0419)$

$= 1 - 0.710046 - 0.248416$

$= 0.041537$

Solution 5d

$Y \sim Po(6.3)$ 3x2.1

$P(\text{Type II error}) = P(3 \leq Y \leq 13)$

$= P(Y \leq 13) - P(Y \leq 2)$

$= 0.9945147 - 0.049846$

using calculator

$= 0.9446\ldots$

≈ 0.945

6. The discrete random variable X has probability generating function

$$G_X(t) = k\ln\left(\frac{2}{2-t}\right)$$

where k is a constant.

(a) Find the exact value of k

(1)

(b) Find the exact value of Var(X)

(7)

(c) Find P($X = 3$)

(4)

Solution 6a

$$G(1) = 1 = k\ln\left(\frac{2}{2-1}\right)$$

$$\Rightarrow k\ln\left(\frac{2}{2-1}\right) = 1$$

$$\Rightarrow k\ln(2) = 1$$

$$\Rightarrow k = \frac{1}{\ln(2)}$$

Probability Generating Function

$$G_X(t) = \sum_x P(X=x)\,t^x = E(t^X)$$

$$G_X(0) = P(X=0) = 0$$

$$G_X(1) = \sum_x P(X=x)\,1^x = \sum_x P(X=x)$$

$$\Rightarrow G_X(1) = 1$$

Solution 6b

$$E(X) = G_X'(1)$$

$$Var(X) = G_X''(1) + G_X'(1) - [G_X'(1)]^2$$

$$G(t) = \frac{1}{\ln 2}\ln\left(\frac{2}{2-t}\right)$$

$$\Rightarrow G(t) = \frac{1}{\ln 2}\left(\ln 2 - \ln(2-t)\right)$$

$$\Rightarrow G(t) = 1 - \frac{\ln(2-t)}{\ln 2}$$

$$\Rightarrow G'(t) = \frac{1}{\ln 2}(2-t)^{-1} \qquad \Rightarrow G''(t) = \frac{1}{\ln 2}(2-t)^{-2}$$

$$E(X) = G'(1) = \frac{1}{\ln 2}(2-1)^{-1} = \frac{1}{\ln 2} \quad \text{and} \quad G''(1) = \frac{1}{\ln 2}(2-1)^{-2} = \frac{1}{\ln 2}$$

$$Var(X) = G''(1) + G'(1) - [G'(1)]^2 = \frac{1}{\ln 2} + \frac{1}{\ln 2} - \left(\frac{1}{\ln 2}\right)^2$$

$$= \frac{1}{\ln 2}\left(2 - \frac{1}{\ln 2}\right)$$

Solution 6c

$P(X=3) =$ co-efficient of t^3

So, by Maclaurin, need $G'''(0)$

$$G'''(t) = \frac{1}{\ln 2} \times \frac{2}{(2-t)^3}$$

$$P(X=3) = \frac{G'''(0)}{3!} = \frac{\frac{1}{4\ln 2}}{6} = \frac{1}{24\ln 2} = 0.0601$$

7. A spinner can land on red or blue. When the spinner is spun, there is a probability of $\frac{1}{3}$ that it lands on blue. The spinner is spun repeatedly.

The random variable B represents the number of the spin when the spinner first lands on blue.

(a) Find (i) $P(B = 4)$

(ii) $P(B \leqslant 5)$

(4)

(b) Find $E(B^2)$

(3)

Steve invites Tamara to play a game with this spinner.

Tamara must choose a colour, either red or blue.

Steve will spin the spinner repeatedly until the spinner first lands on the colour Tamara has chosen. The random variable X represents the number of the spin when this occurs.

If Tamara chooses red, her score is e^X

If Tamara chooses blue, her score is X^2

(c) State, giving your reasons and showing any calculations you have made, which colour you would recommend that Tamara chooses.

(5)

Solution 7ai

$B \sim Geo(1/3)$

$P(B=4)$

$= \left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right) = \frac{8}{81}$

Geometric Distribution measures no. of trials until a success

Probability of success $= p$

So $X \sim Geo(p)$

Mean $= \mu = E(X) = \frac{1}{p}$

Variance $= Var(X) = \sigma^2 = \frac{1-p}{p^2}$

$P(X=x) = \underbrace{p}_{\text{1 success}}\underbrace{(1-p)^{x-1}}_{x-1\text{ failures}}$

Solution 7aii

$P(B \leqslant 5)$

$= P(B=1) + P(B=2) + P(B=3) + P(B=4) + P(B=5)$

$= \left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^3 + \left(\frac{1}{3}\right) + \left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)$

$= \left(\frac{1}{3}\right)\left[1 + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^3 + \left(\frac{2}{3}\right)^4\right]$

$\sum_{k=1}^{n-1} ar^k = \frac{a(1-r^n)}{1-r}$

$= \left(\frac{1}{3}\right)\left[\frac{1-(2/3)^5}{1-2/3}\right] = 1 - (2/3)^5 = \frac{211}{243}$

Solution 7b

$$Var(B) = E(B^2) - (E(B))^2$$

$$\Rightarrow E(B^2) = Var(B) + (E(B))^2$$

$$= \frac{1-p}{p^2} + \left(\frac{1}{p}\right)^2$$

$$= \frac{1-p+1}{p^2}$$

$$= \frac{2-p}{p^2} = \frac{2-\frac{1}{3}}{(\frac{1}{3})^2} = 18 - 3 = 15$$

Solution 7c

Let R = no. of spin when it first lands on red

$$X = R \sim Geo\left(\frac{2}{3}\right)$$

$$E(e^X) = \sum_{x=1}^{\infty} e^x \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{x-1}$$

$$= \frac{2e}{3} \sum_{x=1}^{\infty} \left(\frac{e}{3}\right)^{x-1}$$

$$\sum_{n=1}^{\infty} ar^n = \frac{a}{1-r}, \; |r|<1$$

$$= \frac{2e}{3} \frac{1}{1-e/3} = \frac{2e}{3-e}$$

$$\Rightarrow E(e^X) = 19.297 > 15 = E(B^2)$$

So Tamara should choose red since it has the greater expected score