1. Kelly throws a tetrahedral die n times and records the number on which it lands for each throw.

She calculates the expected frequency for each number to be 43 if the die was unbiased.

The table below shows three of the frequencies Kelly records but the fourth one is missing.

Number	1	2	3	4
Frequency	47	34	36	x

(a) Show that $x = 55$

(1)

Kelly wishes to test, at the 5% level of significance, whether or not there is evidence that the tetrahedral die is unbiased.

(b) Explain why there are 3 degrees of freedom for this test.

(1)

(c) Stating your hypotheses clearly and the critical value used, carry out the test.

(5)

Solution 1a

Number	1	2	3	4
Frequency	47	34	36	x
Freq of Unbiased Die	43	43	43	43

Total Frequency $= 47 + 34 + 36 + x = 117 + x$

Total Freq for unbiased die $= 4 \times 43 = 172$

Now $117 + x = 172 \Rightarrow x = 55$

Solution 1b

Degrees of freedom $\nu = 4 - 1 = 3$ since only constraint is that totals agree.

Solution 1c

H_0 : The die is unbiased

H_1 : The die is biased

Number	1	2	3	4
Frequency O_i	47	34	36	55
Freq of unbiased die E_i	43	43	43	43
$\frac{(O_i-E_i)^2}{E_i}$	0.37209	1.88372	1.13954	3.34884

$\sum \frac{(O_i-E_i)^2}{E_i} = \boxed{6.74419}$

$\chi^2_{(\boxed{3},\boxed{0.05})} = \boxed{7.815}$

degrees of freedom

5% level of significance

Critical Value

ν	0.995	...	0.100	$\boxed{0.050}$	...
1					
2			4.605	5.991	
$\boxed{3}$	0.072		6.251	7.815	
⋮					

Test Statistic $= \boxed{6.74419}$

Since Test Statistic $= 6.74419 < 7.815 =$ Critical Value, there is insufficient evidence to reject H_0

The test is inconclusive. (Result not significant)

This is consistent with die being unbiased

2. On a weekday, a garage receives telephone calls randomly, at a mean rate of 1.25 per 10 minutes.

(a) Show that the probability that on a weekday at least 2 calls are received by the garage in a 30-minute period is 0.888 to 3 decimal places.

(2)

(b) Calculate the probability that at least 2 calls are received by the garage in fewer than 4 out of 6 randomly selected, non-overlapping 30-minute periods on a weekday.

(2)

The manager of the garage randomly selects 150 non-overlapping 30-minute periods on weekdays.
She records the number of calls received in each of these 30-minute periods.

(c) Using a Poisson approximation show that the probability of the manager finding at least 3 of these 30-minute periods when exactly 8 calls are received by the garage is 0.664 to 3 significant figures.

(4)

(d) Explain why the Poisson approximation may be reasonable in this case.

(1)

The manager of the garage decides to test whether the number of calls received on a Saturday is different from the number of calls received on a weekday. She selects a Saturday at random and records the number of telephone calls received by the garage in the first 4 hours.

(e) Write down the hypotheses for this test.

(1)

The manager found that there had been 40 telephone calls received by the garage in the first 4 hours.

(f) Carry out the test using a 5% level of significance.

(4)

Solution 2a

Let X = no. of calls received in a 30 minute period

3×10

since $30 = \underline{3} \times 10$

$\Rightarrow X \sim Po(3 \times 1.25)$

$\Rightarrow X \sim Po(3.75)$

$P(X \geqslant 2) = 1 - P(X \leqslant 1) = 1 - \underbrace{0.1117}_{} = 0.8883 = 0.888$ (3dp)

$x = 1$ and $\lambda = 3.75$
use calculator to find $P(X \leqslant 1)$ for $X \sim Po(3.75)$

Solution 2b

Let Y = no. of 30 min periods when at least 2 calls are received

$Y \sim B(6, 0.888)$

$P(Y<4) = P(Y \leq 3)$

$= \boxed{P(Y=0)} + \boxed{P(Y=1)} + \boxed{P(Y=2)} + \boxed{P(Y=3)}$

$= \boxed{(1-0.888)^6} + \boxed{\underbrace{6}_{\frac{6!}{5!1!}}(1-0.888)^5 \times 0.888} + \boxed{\underbrace{15}_{\frac{6!}{4!2!}}(1-0.888)^4 \times 0.888^2}$

$+ \boxed{\underbrace{20}_{\frac{6!}{3!3!}}(1-0.888)^3 \times 0.888^3}$

$= 0.02163$

Solution 2c

Use the variable in part a:

Let X = no. of calls received in a 30 minute period

$X \sim Po(3.75)$

$$P(X=8) = \frac{e^{-3.75}(3.75)^8}{8!} = 0.02281$$

Let E = no. of 30 minute periods when exactly 8 calls are made

$E \sim B(150, 0.02281)$

$np = 150 \times 0.02281 = 3.4215\ldots$

use calculator with $x=2$ and $E \sim Po(3.4215)$

$\Rightarrow E \sim Po(3.4215) \Rightarrow P(E \geq 3) = 1 - P(E \leq 2)$

$= 0.664$

Extra

Solution 2c (alternative method not required but given only for information)

$E \sim B(150, 0.02281)$ Let $p = 0.02281$, $q = 1-p$

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

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

$= 1 - q^{150} - 150q^{149}p + \frac{150 \times 149}{2} q^{148} p^2$

$= 0.667524\ldots$

Extra

Solution 2d

Poisson approx. may be reasonable because:

n (the number of periods) is large and

p (the probability of receiving 8 calls) is small

Solution 2e

$H_0: \lambda = 30$

$H_1: \lambda \neq 30$

Explanation

If 1.25 calls received in 10 minutes, how many calls received in 4 hours?

4 hours = 240 minutes = 24 x 10 minutes

So number of calls in 4 hours $= 24 \times 1.25$
$= 30$

Solution 2f

$X \sim Po(30)$

$P(X \geqslant 40) = 1 - P(X \leqslant 39)$

$= 0.04625$

Two tailed test
So we use $\frac{5\%}{2} = 0.025$

Since $0.04625 > 0.025$, there is no evidence to reject H_0

There is insufficient evidence at the 5% level of significance that the number of calls is different on a Saturday.

3. A courier delivers parcels. The random variable X represents the number of parcels delivered successfully each day by the courier where $X \sim B(400, 0.64)$

A random sample $X_1, X_2, \ldots X_{100}$ is taken.

Estimate the probability that the mean number of parcels delivered each day by the courier is greater than 257

(4)

Solution 3

Central Limit Theorem
Suppose you take a sample of readings from any distribution with mean μ and variance σ^2.
- For large n, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
- n large : $n > 30$

$X \sim B(400, 0.64)$

$n = 400$
$p = 0.64$

$\mu = np = 400 \times 0.64 = \boxed{256}$

$\sigma = np(1-p)$

$= 400 \times 0.64 \times (1 - 0.64)$

$= 92.16$

$Var(\bar{X}) = \frac{92.16}{100} = \boxed{0.9216}$

sample size = 100

$\bar{X} \approx N(\boxed{256}, \boxed{0.9216})$

μ Var($\bar{X}$)

$P(\bar{X} > \boxed{257}) = P\left(Z > \frac{\boxed{257} - \boxed{256}}{\sqrt{\boxed{0.9216}}}\right)$

$= 0.149$

calculator

4. Members of a photographic group may enter a maximum of 5 photographs into a members only competition.
Past experience has shown that the number of photographs, N, entered by a member follows the probability distribution shown below.

n	0	1	2	3	4	5
$P(N = n)$	a	0.2	0.05	0.25	b	c

Given that $E(4N + 2) = 14.8$ and $P(N = 5 \mid N > 2) = \frac{1}{2}$

(a) show that $\text{Var}(N) = 2.76$

(6)

The group decided to charge a 50p entry fee for the first photograph entered and then 20p for each extra photograph entered into the competition up to a maximum of £1 per person. Thus a member who enters 3 photographs pays 90p and a member who enters 4 or 5 photographs just pays £1

Assuming that the probability distribution for the number of photographs entered by a member is unchanged,

(b) calculate the expected entry fee per member.

(3)

Bai suggests that, as the mean and variance are close, a Poisson distribution could be used to model the number of photographs entered by a member next year.

(c) State a limitation of the Poisson distribution in this case.

(1)

Solution 4a

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

This formula shows we need to find $E(N)$ and $E(N^2)$

From question

$$\boxed{E(4N+2) = 14.8} \Rightarrow \boxed{4E(N) + 2 = 14.8} \Rightarrow \boxed{E(N) = 3.2}$$

n	0	1	2	3	4	5
$P(N=n)$	a	0.2	0.05	0.25	b	c
n^2	0	1	4	9	16	25

$$E(N^2) = \sum_{n=0}^{5} n^2 P(N=n)$$

$$= \boxed{0a} + \boxed{1 \times 0.2} + \boxed{4 \times 0.05} + \boxed{9 \times 0.25} + \boxed{16b} + \boxed{25c}$$

$$= 2.65 + 16b + 25c$$

Solution 4a (continued)

n	0	1	2	3	4	5
$P(N=n)$	a	0.2	0.05	0.25	b	c
n^2	0	1	4	9	16	25

From last page,

$$\boxed{E(N)} = \boxed{3.2}$$

$$\boxed{E(N^2) = 2.65 + 16b + 25c}$$

From table, (and formula $E(N) = \sum nP(N=n)$)

$$E(N) = 0 \times a + 1 \times 0.2 + 2 \times 0.05 + 3 \times 0.25 + 4b + 5c$$

$$= 0.2 + 0.1 + 0.75 + 4b + 5c$$

$$\Rightarrow \boxed{E(N)} = \boxed{1.05 + 4b + 5c}$$

But $\boxed{E(N)} = \boxed{3.2}$

$$\Rightarrow \boxed{3.2} = \boxed{1.05 + 4b + 5c}$$

$$\Rightarrow 2.15 = 4b + 5c \quad (*)$$

Also, since $\boxed{P(N=5 \mid N>2) = \frac{1}{2}}$, we have

$$\boxed{\frac{c}{0.25 + b + c} = 0.5}$$

$$\Rightarrow c = 0.5 \times 0.25 + 0.5b + 0.5c$$

$$\Rightarrow c - b = 0.25$$

$$\Rightarrow c = 0.25 + b \quad (**)$$

Substitute $(**)$ in $(*)$:

$$2.15 = 4b + 5(0.25 + b)$$

$$\Rightarrow 2.15 - 1.25 = 9b$$

$\Rightarrow b = 0.1$. Substitute in $(**)$ $\Rightarrow c = 0.25 + 0.1 = 0.35$

$$\Rightarrow \boxed{E(N^2) = 2.65 + 16 \times 0.1 + 25 \times 0.35} = \boxed{13}$$

$$\Rightarrow Var(N) = \boxed{E(N^2)} - \boxed{E(N)}^2 = \boxed{13} - \boxed{3.2}^2 = 2.76$$

Solution 4b

Fee	$\boxed{0}$	$\boxed{50}$	$\boxed{70}$	$\boxed{90}$	$\boxed{100}$	$\boxed{100}$
$P(N=n)$	$\boxed{a}$	$\boxed{0.2}$	$\boxed{0.05}$	$\boxed{0.25}$	$\boxed{b}$	$\boxed{c}$

$$E(Fee) = \sum \boxed{(Fee)} \times \boxed{P(N=n)}$$

$$= (\boxed{0} \times \boxed{a}) + (\boxed{50} \times \boxed{0.2}) + (\boxed{70} \times \boxed{0.05}) + (\boxed{90} \times \boxed{0.25}) + \boxed{100}\boxed{b} + \boxed{100}\boxed{c}$$

$$= 0 + 10 + 3.5 + 22.5 + 100b + 100c$$

$b = 0.1$ $\quad$ $c = 0.35$

$$= 36 + 100 \times 0.1 + 100 \times 0.35$$

$$= 36 + 10 + 35$$

$$= 81p$$

Solution 4c

Poisson distribution will assign substantial probability to $N > 5$

5. Asha, Davinda and Jerry each have a bag containing a large number of counters, some of which are white and the rest are red.
Each person draws counters from their bag one at a time, notes the colour of the counter and returns it to their bag.

The probability of Asha getting a red counter on any one draw is 0.07

(a) Find the probability that Asha will draw at least 3 white counters before a red counter is drawn.

(2)

(b) Find the probability that Asha gets a red counter for the second time on her 9th draw.

(2)

The probability of Davinda getting a red counter on any one draw is p.
Davinda draws counters until she gets n red counters. The random variable D is the number of counters Davinda draws.

Given that the mean and the standard deviation of D are 4400 and 660 respectively,

(c) find the value of p.

(4)

Jerry believes that his bag contains a smaller proportion of red counters than Asha's bag. To test his belief, Jerry draws counters from his bag until he gets a red counter. Jerry defines the random variable J to be the number of counters drawn up to and including the first red counter.

(d) Stating your hypotheses clearly and using a 10% level of significance, find the critical region for this test.

(5)

Jerry gets a red counter for the first time on his 34th draw.

(e) Giving a reason for your answer, state whether or not there is evidence that Jerry's bag contains a smaller proportion of red counters than Asha's bag.

(2)

Given that the probability of Jerry getting a red counter on any one draw is 0.011

(f) show that the power of the test is 0.702 to 3 significant figures.

(3)

Solution 5a

$P(\text{at least 3 whites}) = (1-0.07)^3$

Solution 5b

$X \sim NB(2, 0.07)$

$x=9$
$r=2$
$p=0.07$

9th draw 2nd red probability of red

Background in Brief
If $X \sim NB(r,p)$ if:
- only 2 possible outcomes
- constant prob p of success
- X is no. of trials until r^{th} success

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

$P(X=9) = \binom{9-1}{2-1} 0.07^2 (1-0.07)^{9-2}$

$= \binom{8}{1} 0.07^2\, 0.93^7 = 0.0236$

Solution 5c

Mean and Variance of Negative Binomial Distribution

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

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

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

$D \sim NB(n, p)$

r=n (in theory) Mean = 4400 in question

$$\text{Mean} = \frac{n}{p} = 4400$$

r=n (in the theory) Standard deviation is 660 in question

$$Var(D) = \frac{n(1-p)}{p^2} = 660^2$$

$$\frac{n}{p} = 4400 \quad (*)$$

$$\frac{n(1-p)}{p^2} = 660^2 \quad (**)$$

$$(**) \Rightarrow \boxed{\frac{n}{p}} \cdot \frac{(1-p)}{p} = 660^2$$

Using $(*)$ gives us $\boxed{4400} \cdot \frac{(1-p)}{p} = 660^2$

$$\Rightarrow 1 - p = 99p$$

$$\Rightarrow p = 0.01$$

Solution 5d

Geometric Distribution

A random variable X follows a geometric distribution if:

- there is sequence of independent trials with 2 outcomes
- constant probability p of success
- X is no. of trials until first success

$X \sim Geo(p) \quad P(X=x) = p(1-p)^{x-1}$

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

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

$H_0: p = 0.07$

$H_1: p < 0.07$

$J \sim Geo(0.07)$

$P(J \geq c) < 0.1$

$\Rightarrow (1 - 0.07)^{c-1} < 0.1$

$\Rightarrow c - 1 > \frac{\log 0.1}{\log 0.93} \Rightarrow c > 32.72 \quad \therefore$ critical region $J \geq 33$

Solution 5e

Since 34 is in the critical region, there is evidence to suggest that Jerry's bag contains a smaller propostion of red counters than Asha's bag.

Solution 5f

Power of test $= P(J \geq 33 \mid p = 0.011)$

$= (1 - 0.011)^{32}$

$= 0.7019$

6. The probability generating function of the random variable X is

$$G_X(t) = k(1 + 2t)^5$$

where k is a constant.

(a) Show that $k = \dfrac{1}{243}$ **(2)**

(b) Find $P(X = 2)$ **(2)**

(c) Find the probability generating function of $W = 2X + 3$ **(2)**

The probability generating function of the random variable Y is

$$G_Y(t) = \frac{t(1 + 2t)^2}{9}$$

Given that X and Y are independent,

(d) find the probability generating function of $U = X + Y$ in its simplest form. **(2)**

(e) Use calculus to find the value of $\text{Var}(U)$ **(6)**

Solution 6a

$$G_X(\boxed{t}) = k(1+2\boxed{t})^5$$

Since $G_X(1) = 1$, true for any pgf

$$G_X(\boxed{1}) = k(1+2\boxed{(1)})^5 = 1$$

$$\Rightarrow k(3)^5 = 1 \quad \Rightarrow k = \frac{1}{3^5} = \frac{1}{243}$$

Probability Generating Function of a discrete random variable X taking only non-negative integer values is

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

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

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

Solution 6b

Differentiating $G_X(t)$:

$$G_X'(t) = k(5)(2)(1+2t)^4 = 10k(1+2t)^4$$

$$\Rightarrow G_X''(t) = 10k(4)(2)(1+2t)^3 = 80k(1+2t)^3$$

$$\Rightarrow G_X''(0) = \boxed{\frac{80}{243}}$$

But $\boxed{P(X=2) = \boxed{\dfrac{G_X''(0)}{2}}} = \boxed{\dfrac{40}{243}}$

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

$G_X'(0) = P(X=1)$

$\boxed{\dfrac{G_X''(0)}{2} = P(X=2)}$ ... and so on ...

$\dfrac{G_X'''(0)}{(3)(2)} = P(X=3)$

$\dfrac{G_X''''(0)}{(4)(3)(2)} = P(X=4)$ ...

Solution 6c

$G_{a+bX}(t) = t^a G_X(t^b)$

NOT in formula booklet

$$G_X(t) = \frac{1}{243}(1+2t)^5$$

So for $W = 3+2X$,

$$G_{3+2X}(t) = t^3 G_X(t^2) = t^3 \cdot \frac{1}{243}(1+2t^2)^5$$

So $$G_{3+2X}(t) = \frac{t^3}{243}(1+2t^2)^5$$

Solution 6d

If X and Y are two independent variables:

$G_{X+Y}(t) = G_X(t) \times G_Y(t)$

In formula booklet

$$G_X(t) = \frac{1}{243}(1+2t)^5$$

$$G_Y(t) = \frac{t(1+2t)^2}{9}$$

$$U = X+Y \Rightarrow G_U(t) = G_X(t) \times G_Y(t)$$

$$= \frac{1}{243}(1+2t)^5 \times \frac{t}{9}(1+2t)^2$$

$$= \frac{t}{2187}(1+2t)^7$$

Solution 6e

$$Var(U) = G_U''(1) + G_U'(1) - [G_U'(1)]^2$$

In formula booklet

$$G_U(t) = \frac{t}{2187}(1+2t)^7$$

product rule

$$\boxed{G_U'(t)} = \frac{t}{2187}(7)(2)(1+2t)^{7-1} + \frac{1}{2187}(1+2t)^7$$

$$= \underbrace{\frac{14t}{2187}(1+2t)^6}_{\text{product rule}} + \frac{1}{2187}(1+2t)^7$$

$$\boxed{G_U''(t)} = \frac{14t}{2187}(6)(2)(1+2t)^5 + \frac{14}{2187}(1+2t)^6 + \frac{(7)(2)}{2187}(1+2t)^6$$

Hence $\boxed{G_U'(1)} = \frac{14}{2187}(3)^6 + \frac{1}{2187}(3)^7 = \frac{3^6 \times 17}{3^6 \times 3} = \boxed{\frac{17}{3}}$

and $\boxed{G_U''(1)} = \frac{14 \times 12}{2187}(3)^5 + \frac{14}{2187}(3)^6 + \frac{14}{2187}(3)^6$

$$= \frac{3^6(14 \times 4 + 14 + 14)}{3^6 \times 3} = \frac{14 \times 6}{3} = \boxed{28}$$

Hence using formula for Var(U):

in formula booklet

$$Var(U) = \boxed{G_U''(1)} + \boxed{G_U'(1)} - [\boxed{G_U'(1)}]^2$$

$$= \boxed{28} + \boxed{\frac{17}{3}} - \left(\boxed{\frac{17}{3}}\right)^2$$

$$= \frac{9 \times 28 + 3 \times 17 - 17^2}{9} = \frac{14}{9}$$

7. A manufacturer has a machine that produces lollipop sticks.
The length of a lollipop stick produced by the machine is normally distributed with unknown mean μ and standard deviation 0.2

Farhan believes that the machine is not working properly and the mean length of the lollipop sticks has decreased.
He takes a random sample of size n to test, at the 1% level of significance, the hypotheses

$$H_0: \mu = 15 \qquad H_1: \mu < 15$$

(a) Write down the size of this test.

(1)

Given that the actual value of μ is 14.9

(b) (i) calculate the minimum value of n such that the probability of a Type II error is less than 0.05
Show your working clearly.

(6)

(ii) Farhan uses the same sample size, n, but now carries out the test at a 5% level of significance. Without doing any further calculations, state how this would affect the probability of a Type II error.

(1)

Solution 7a

Size of test $= 0.01$

The size of the test is the probability of being in the critical region

Solution 7b

Type I error: rejecting null hypothesis when it is actually true

Type II error: not rejecting null hypothesis when it is actually false

Size of test = Probability of Type I error (or probability of being in critical region given that H_0 is correct)