INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Subject: Probability and Statistics II

Chapter: Unit 1 & 2

Category: Assignment Solutions

1. ..

Solution 4: Let X be the amount of fixed benefit health insurance claims and Y the amount of indemnity based health insurance claim.

Then:
$X \sim N(900, 100^2)$ and $Y \sim N(1400, 300^2)$

We require

$P((Y_1+Y_2 + Y_3) > (X_1+X_2 + X_3+ X_4) + 900)$
$= P((Y_1+Y_2 + Y_3) - (X_1+X_2 + X_3+ X_4) > 900)$

So we need the distribution of $(Y_1+Y_2 + Y_3) - (X_1+X_2 + X_3 + X_4)$:

$(Y_1+Y_2 + Y_3) - (X_1+X_2 + X_3 + X_4) \sim N(3\times1400 - 4\times900,\ 3\times300^2+4\times100^2)$

i.e $(Y_1+Y_2 + Y_3) - (X_1+X_2 + X_3 + X_4) \sim N(600,310000)$

Therefore

$P((Y_1+Y_2 + Y_3) - (X_1+X_2 + X_3 + X_4)>900)$

$= P(Z > \frac{900-600}{\sqrt{310000}}) = P(Z > 0.54) = 1 - P(Z < 0.54) = 1- 0.70540 = 0.2946$

[4 Marks]

2. :

UNIT 1

ASSIGNMENT SOLUTIONS

(i) Sample mean: $\bar{X} = \frac{\sum X_i}{n}$

$E[\sum X_i] = \sum E[X_i] = \sum \mu = n\mu$; since they are identically distributed

$Var[\sum X_i] = \sum Var[X_i] = n\sigma^2$; since they are iid

$E[\bar{X}] = \mu.$

$Var[\bar{X}] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

(ii) Sample variance : $S^2 = \frac{1}{n-1}[\sum X_i^2 - n\bar{X}^2]$

$E[S^2] = \frac{1}{n-1}(\sum E[X_i^2] - n\,E[\bar{X}^2])$

$= \frac{1}{n-1}[\sum(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)]$

$= \frac{1}{n-1}[n(\sigma^2 + \mu^2) - \sigma^2 - n\mu^2] = \frac{1}{n-1}[(n-1)\sigma^2] = \sigma^2$

(iii) The sampling distribution of $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ when sampling from a normal population, with mean μ and variance σ^2

The variance of χ^2_k is 2k.

Hence, $Var[\frac{(n-1)S^2}{\sigma^2}] = 2(n-1) \Rightarrow Var[S^2] = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$

(iv) It is given that $\sigma^2 = 100$ and n = 10.

We need to compute P $(50<S^2<150)$

We know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ and in this case it is χ^2_9

Let $Y = 9\,S^2 / 100$. Then, for $S^2 = 50$: $Y = 9\,(50)/100 = 4.5$;

and for $S^2 = 150$: $Y = 9\,(150)/100 = 13.5$

Thus, P $(50<S^2<150)$ = P $(4.5< Y<13.5)$

= P (Y<13.5) - P (Y<4.5) [from page 165 of Tables]

= 0.8587 – 0.1245 = 0.7342

3. :

(i) The likelihood function is:

$$L(p) = C\,[(1-p)^4]^{86}\,[4p(1-p)^3]^{75}\,[6p^2(1-p)^2]^{16}\,[4p^3(1-p)]^2\,[p^4]^1$$

$$= C\,[(1-p)^{344}]\,[p^{75}(1-p)^{225}]\,[p^{32}(1-p)^{32}]\,[p^6(1-p)^2]\,[p^4]$$

$$= C\,(1-p)^{603}\,p^{117}$$

C is a constant.

Taking logs and differentiating with respect to p and setting equal to zero gives:

$d\text{Ln } L/dp = -\,603/(1-p) + 117/p = 0$
$\Rightarrow p = 117/\,(603+117) = 117/720 = 0.1625$
Checking for maximum:
$d^2\text{Ln } L/dp^2 = -\,603/\,(1-p)^2 - 117/p^2 < 0$
$\Rightarrow$ Maximum value for Ln L is at $p = 0.1625$

(ii)

Goodness of fit test:
We are testing the following hypotheses using a $\chi 2$ goodness of fit test:
H_0: the probabilities conform to a Bin (4, p) distribution
H_1: the probabilities do not conform to a Bin (4, p) distribution

Using $\hat{p} = 0.1625$ from part (a), the probabilities for this binomial distribution are:
$P(X=0) = (1-p)^4 = 0.49197$
$P(X=1) = 4p\,(1-p)^3 = 0.38183$
$P(X=2) = 6p^2\,(1-p)^2 = 0.11113$
$P(X=3) = 4p^3\,(1\text{-}p) = 0.01437$
$P(X=4) = p^4 = 0.00070$
The expected values are 88.55, 68.73, 20.00, 2.59 and 0.13.
Combining the expected values less than 5 to third group, we get

No of Claims	0	1	2 or more
Observed O_i	86	75	19
Expected E_i	88.55	68.73	22.72

The degrees of freedom = 3-1-1=1.

$\chi 2 = \sum (O_i - E_i)^2 / E_i$

$= (86 - 88.55)^2 / 88.55 + (75 - 68.73)^2 / 68.73 + (19 - 22.72)^2 / 22.72$
$= 0.074 + 0.572 + 0.608 = 1.254$

This is less than the 5% critical value of 3.841.We have insufficient evidence at 5% level to reject Ho. Hence, the model is a good fit.

4. :

(i)

From the Formulae and Tables for Actuarial Examinations this pdf corresponds to two parameter version of the distribution given by

$$f(x) = \frac{\alpha \lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \; ; x > 0 , \lambda > 0 \text{ and } \alpha > 0$$

$E[X] = \lambda / (\alpha - 1)$ and $Var[X] = \alpha \lambda^2 / ((\alpha - 1)^2 (\alpha - 2))$; $\alpha > 2$

Thus the pdf of X, $f(x) = \frac{c\beta^3}{(x+\beta)^4}$; $x > 0$ and $\beta > 0$ can be identified with c=3 (α known) $\lambda = \beta$ (unknown).

$E[X] = \beta/2$

[0.5]

$Var[X] = 3/4\, \beta^2$

[0.5]

If the students obtain the results using the first principles full credit to be given

$$f(x) = \frac{\alpha \lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \; ; x > 0 , \lambda > 0 \text{ and } \alpha > 0$$

$\int_0^{\infty} \frac{c\beta^3}{(x+\beta)^4} dx = 1$ implies $\left[- \frac{c}{3} \frac{\beta^3}{(x+\beta)^3} \right]_0^{\infty} = 1$ giving $c = 3$

Mean and Variance

(ii)

The method of moments estimator:

Equating sample mean to E[X] from (i) gives:

$\bar{X}_n = \beta / 2 \Rightarrow$ The moments estimator $\hat{\beta} = 2\bar{X}_n$

The mean square error of $\hat{\beta}$:

$MSE[\hat{\beta}] = Var[\hat{\beta}] + (Bias[\hat{\beta}])^2$

$E[\hat{\beta}] = E[2\bar{X}_n] = 2E[\bar{X}_n] = 2E[X] = 2(\beta/2) = \beta$ and $Bias = E[\hat{\beta}] - \beta = 0$

—This estimator is unbiased.

Using the fact that the individual values of *Xi* are independent:
Var $[\hat{\beta}]$ = Var $[2\,\bar{X}_n]$ = 4 Var $[\,\bar{X}_n]$ = 4 Var[X] / n = 4 (3/4 β^2) / n= 3 β^2 / n

Hence, MSE $[\hat{\beta}]$ = 3 β^2/ n + 0 = 3 β^2/ n
This estimator is consistent since MSE tends to zero as n $\rightarrow \infty$

(iii)

MSE $[b\,\bar{X}_n]$ = Var $[b\,\bar{X}_n]$ + (Bias $[b\,\bar{X}_n]$)^2

$$= b^2 \frac{3}{4}\frac{\beta^2}{n} + (E[b\,\bar{X}_n] - \beta)^2$$

$$= b^2 \frac{3}{4}\frac{\beta^2}{n} + (\beta\,(\frac{b}{2} - 1))^2$$

$$= \frac{\beta^2}{n}\,(\frac{3}{4}b^2 + n\,(\frac{b}{2} - 1)^2)$$

Differentiating the MSE respect to b:
d (MSE)/db = $\frac{\beta^2}{n}$ $\left(\frac{3}{2}b + n\,(\frac{b}{2}-1)\right)$
Setting this equal to 0 gives b = 2n / (n + 3) = 2 / (1 + $\frac{3}{n}$)

Differentiating the MSE a second time with respect to b, we obtain:
d^2(MSE)/db^2 = $\frac{\beta^2}{n}$ $(\frac{3}{2} + \frac{n}{2})$ which is positive
=> a minimum value for the MSE is at b = 2 / (1 + 3/n).

(iv)

It is seen in (ii) that $\hat{\beta} = 2\,\bar{X}_n$ is an unbiased estimator. The estimator $b\bar{X}_n$ in (iii) when b = 2 / (1 + $\frac{3}{n}$) is negatively biased estimator as b < 2 for n $\geq$ 1.

As n $\rightarrow \infty$, b tends to 2. So, the estimator in (iii) is also consistent, in view of (ii)

5. :

i) $E[X] = \frac{(a+b)}{2}$

$b = 2E[X] - a = 2\bar{x} - a$

$$Var(X) = \frac{(b-a)^2}{12}$$

$$s^2 = \frac{(2\bar{x}-a-a)^2}{12} = \frac{(\bar{x}-a)^2}{3}$$

$$(\bar{x} - a)^2 = 3\,s^2$$

$$\hat{a} = \bar{x} - \sqrt{3}s$$

ii) $\hat{b} = 2\bar{x} - (\bar{x} - \sqrt{3}s)$

$\hat{b} = \bar{x} + \sqrt{3}s$

iii) Sample: 1, 2, 3, 4, 50

$\bar{x} = 12\,; s = 21.27$

Method of moments estimates using above formulae are:

$\hat{a} = 12 - \sqrt{3}(21.27) = -24.84; \hat{b} = 12 + \sqrt{3}(21.27) = 48.84$

For U (a, b),the probability of a sample point being less than 'a' or greater than 'b' is zero and we have a sample value 50 that is greater than our estimate of 'b'. This highlights a potential weakness of the method of moments.

iv) Likelihood for a sample of size *n* is $L(b) = \frac{1}{b^n}$ if $b \geq \max(x_i)$,otherwise $L = 0$

Differentiation with respect to *b* does not work because in the range of *x* depends on *b*

We must find *b* that maximizes $L(b)$ *for* $\max(x_i)$ given.. We want *b* to be as small as possible subject to the constraint that $b \geq \max(x_i)$.

Clearly the maximum is attained at $b = \max(x_i)$.

Hence $\hat{b} = \max(x_i)$.

6. :

i) P(Type I error) is the probability of rejecting H_0 when H_0 is true.

Let X be the no. of hits in first step 12 missiles and Y be the no. of hits in second step 12 missiles.

P(Type I error) = P (X≥ 3 | p = 0.1) + P (X = 0 | p = 0.1) P (Y ≥ 5 | p = 0.1) +
P (X = 1 | p = 0.1) P (Y ≥ 4 | p = 0.1) + P (X = 2 | p = 0.1) P (Y ≥ 3 | p = 0.1)

Using Actuarial Tables page 188 (probabilities for Binomial Distribution)

$$= (1 - 0.8891) + (0.2824)(1 - 0.9957) + (0.6590 - 0.2824)(1 - 0.9744) + (0.8891 - 0.6590)(1 - 0.8891)$$

$$= 0.14727 \approx 15\%$$

ii) Probability of rejecting the null hypothesis when p = 0.3

= P (X ≥ 3 | p = 0.3) + P (X = 0 | p = 0.3) P (Y ≥ 5 | p = 0.3) +
P (X = 1 | p = 0.3) P (Y ≥ 4 | p = 0.3) + P (X = 2 | p = 0.3) P (Y ≥ 3 | p = 0.3)

Using Actuarial Tables page 188 (probabilities for Binomial Distribution)

$$= (1 - 0.2528) + (0.0138)(1 - 0.7237) + (0.0850 - 0.0138)(1 - 0.4925) + (0.2528 - 0.0850)(1 - 0.2528)$$

$$= 0.91253 \approx 91\%$$

iii) P(type II error) is the probability of accepting H_0 when H_0 is false.
$= 1 - 0.91253 = 0.08747 \approx 9\%$(when p=0.3)

iv) 10 Space agencies in aggregate fired 120 missiles and recorded 40 hits
Assuming that the sample comes from a binomial distribution, we know that the quantity

$$\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Here n = 120, X = 40. so$\hat{p} = \frac{40}{120} = \frac{1}{3} = 0.3333$

Using Actuarial Tables page162, $Z_{10\%} = 1.2816$
Lower bound of 90% right-tailed confidence interval for p is

$$\hat{p} - 1.2816\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.3333 - 1.2816\sqrt{\frac{0.3333(1-0.3333)}{120}} = 0.2782$$

v) Test whether p > 0.278 at 10% level of significance
H_0: P = 0.278 vs. H_1: p > 0.278
For one sample binomial model:

$$\frac{X-np_0}{\sqrt{np_0q_0}} \sim N(0,1)\text{with continuity correction.}$$

$$\frac{39.5 - 120(0.278)}{\sqrt{120(0.278)(0.722)}} = 1.2511$$

We are carrying out a one-sided test. The value of the test statistic is less than

1.2816 (the upper 10% point of the N (0, 1) distribution) so we do not have sufficient evidence to reject H_0 at 10% level.

vi) Lower bound of Confidence interval implies that there is only a 10% chance of 'p' ≤ 0.2782, whereas from the hypothesis test, 'p' could be less than or equal to 0.2780 with probability more than 10% (approx 10.5% corresponding to 1.2511).

The minor disconnect between Confidence interval and Hypothesis testing at the same level is due to

- use of sample proportion $\hat{p}$to estimate population variance in calculating confidence interval, and
- applying continuity correction in hypothesis testing

vii) Test whether there is a difference in the mean scores

We assume that the samples come from normal distributions with the same variance and that the samples are independent.

$H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$.

The pivotal quantity is: $\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

Given: $\bar{X}_1 = 65$; $\bar{X}_2 = 70$; $S_1 = 54$; $S_2 = 70$; $n_1 = 12$; $n_2 = 15$

The pooled variance is: $S_P^2 = \frac{1}{25}(11\,(2916) + 14\,(4900)) = 4027.04$

$\frac{65-70-0}{63.459 * \sqrt{\frac{1}{12}+\frac{1}{15}}}$ =-0.2034

This is within ± 2.060 $(= t_{25;2.5\%})$So we have insufficient evidence to reject H_0 at the 5% level. Therefore, it is reasonable to conclude that there is no significant difference in the mean scores for the populations associated with the two Institutes.

7. :

UNIT 1

ASSIGNMENT SOLUTIONS

i)

$$\sum X_{Public} = 270$$
$$(\bar{X}_{Public}) = 30$$
$$\sum X_{Private} = 315.5$$
$$(\bar{X}_{Private}) = 35.06$$
$$\sum X_{Public}^2 = 8614.5$$

(subscript 1 refers public hospitals)

$$S_1^2 = (8614.5 - 9 \times 30^2)/8 = 64.31$$
$$\sum X_{Private}^2 = 11599.25$$
$$S_2^2 = (11599.25 - 9 \times 35.06^2)/8 = 67.42$$

ii) Two sided t-test can be applied in case the samples come from populations with equal variances.

[1]

We are testing $H_0 : \sigma_1^2 = \sigma_2^2 \ vs \ H_1 : \sigma_1^2 \neq \sigma_2^2$

Test statistic is $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$

Value of statistic is $\frac{64.31}{67.42} = 0.9542$

$F_{8,8}$ values at 5% levels are 0.2256 and 4.433 . Since the value of the test statistic is between the above values, we have insufficient evidence to reject the hypothesis and conclude that the population variances are equal.

iii) We are testing $H_0 : \mu_1 = \mu_2 \ vs \ H_1 : \mu_1 < \mu_2$
[0.5]

Test statistic is $\frac{(\bar{X}_2-\bar{X}_1)-(\mu_2-\mu_1)}{\sqrt{S_P^2(1/n_1+1/n_2)}} \sim t_{n_1+n_2-2}$

Where $S_P^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)}{(n_1+n_2-2)}$

Using the values in section (I) above,

$$S_P^2 = \frac{64.31\times 8+67.42\times 8}{16} = 65.86$$

value of test statistic is

$$\frac{(35.06-30)-0}{\sqrt{65.86(1/9+1/9)}} = 1.32$$

$$P(t_{16} > 1.32) = 20.5\%$$

This is higher than 95% hence we do not have sufficient evidence to reject the hypothesis and hence conclude that the cost of claims in private hospitals is similar to that in public hospital

8. :

(i) $\hat{\theta}$ said to be unbiased when $E(\hat{\theta}) = \theta$

(ii) measure of the 'bias' is given by $E(\hat{\theta}) - \theta$

(iii) Mean Square Error (MSE) of this estimator $\hat{\theta} = \left(E(\hat{\theta}) - \theta\right)^2$

(iv) $\tilde{\theta}$ is efficient as an estimator with lower MSE is said to be more efficient than one with higher MSE.

(v) An estimator is termed as consistent if MSE converges to 0 as the sample size tends to ∞

(vi) θ can be estimated using:

a. Method of moments: the population moments are equated to the sample moments to estimate the parameters.

b. Maximum likelihood method: A maximum likelihood function $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$ is generated. A maximum likelihood estimate of the parameter is given by solution to $\frac{dL(\theta)}{d\theta} = 0$

c. Bootstrap method: This is computer intensive method that allows us to avoid making assumption about the sampling distribution by forming an empirical sampling distribution which is possible due to re-sampling based on the available sample.

9. :

i) Variable t_k is defined as $t_k \Rrightarrow \frac{N(0,1)}{\sqrt{\chi_k^2/k}}$

where k denotes the degrees of freedom
and the two random variables $N(0,1)$ and χ_k^2 are independent.

ii) Mean and variance of t_k for k>2 are 0
and k/(k-2) respectively.

iii)

a) We know that for a sample from a normal population,

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

For the given confidence level $t_9 = 3.25$, Confidence interval for μ is thus $\left(50 - 3.25 \times \sqrt{\frac{48.667}{10}}, 50 + 3.25 \times \sqrt{\frac{48.667}{10}}\right)$

i.e. $(42.83\ ,\ 57.17)$

b)

From (ii) above, we know that $t_{n-1} \sim N\left(0, \sqrt{\frac{n-1}{n-3}}\right) \sim N(0\,, \sqrt{\frac{9}{7}})$

i.e.

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0\,, \sqrt{\frac{9}{7}})$$

$$\frac{\bar{X}-\mu}{S/\sqrt{7n/9}} \sim N(0\,,1)$$

Critical value for given level of confidence is 2.58

Confidence interval for μ is thus $\left(50 - 2.58 \times \sqrt{\frac{48.667}{10} \times \frac{9}{7}}, 50 + 2.58 \times \sqrt{\frac{48.667}{10} \times \frac{9}{7}}\right)$

i.e. $(43.54\ ,\ 56.45)$

10. :

i) Type I error - Event of Rejecting the hypothesis when it is true

Let X be the random variable denoting the total number of claims on the portfolio. X thus follows Poi(nμ) i.e. Poi(3000) where μ is the Poisson parameter.

Null hypothesis H_0 is thus $X \sim Poi(3000)$

$P(reject\ H_0\ when\ H_0\ is\ true) = P(X > 3100\ when\ X \sim Poi(3000))$

Using normal approximation (as nλ is large enough)

$$X \sim N(3000{,}3000)$$

$P(X > 3100) = P\left(Z > \frac{3100-300}{\sqrt{3000}}\right) = 1 - P(Z < 1.825) = 3.39\%$

ii) Type II error - Event of Accepting the hypothesis when it is false

iii) Power of a test - Probability of Rejecting the hypothesis when it is false

In terms of μ it is given by:

$P(reject\ H_0\ when\ H_0\ is\ false) = P(X > 3100\ when\ X \sim Poi(n\mu) \sim N(n\mu, n\mu))$

$P(X > 3100) = 1 - \left(Z < \frac{3100-n\mu}{\sqrt{n\mu}}\right)$

The value of power of test will depend on the value of parameter under alternate hypothesis.

iv) If $\hat{\mu}$ is the estimator of μ (the poisson parameter), $\hat{\mu}$ follows $N(\mu, \hat{\mu}/n)$

Hence $\frac{\hat{\mu}-\mu}{\sqrt{\hat{\mu}/n}}$ follows N(0,1)

Or $P\left(-2.5758 < \frac{2.9-\mu}{\sqrt{2.9/1000}} < 2.5758\right) = 0.99$

Hence the confidence interval for μ is (2.7613, 3.0387)

11.:

<u>**Solution 5:**</u>

i) P(X=2) = 0.3 + 0.2 + 0 = 0.5

Required expectation is summation of y* P(Y=y| X=2)

= 1*0.3/0.5 + 2*0.2/0.5 +3*0/0.5

=3/5 + 4/5 = 7/5 =1.4

ii) For 5X - 4Y,
Mean is 5*E(X) – 4*E(Y) = 5*0 -4*0 =0
Variance is 5^2 * Var (X) + 4^2 * Var (Y) = 25*1 + 16*1 = 41
Hence required distribution is N(0,41)

iii) (2*Lambda *X) ~ Chi square distribution with (2 *alpha) degrees of freedom

P(X>50) = P(2*0.1*X > 2*0.1*50)

= P(Chi square >10)

with 2*10 =20 degrees of freedom

Hence,

Required Chi square expression is

P(Chi square > 10) where chi square distribution will have 20 degrees of freedom

Required chi square probability is equal to 1-0.0318 = 0.9682

Hence, probability of X greater than 50 is over 96.8%

iv) E(X) = alpha /lambda = 10/0.1 = 100

Var(X) = alpha /lambda^2 = 10/0.1^2 = 1000

Hence using Central Limit theorem, using normal approximation

(P X>50) = ~P(N(100,1000) >50)
~P(Z>(50-100)/sqrt(1000))
~P(Z(N(0,1)>-1.58114))
~Z(N(0,1)<1.58114))

For correct equation as above

From tables

x	phi(x)
1.58	0.94295
1.59	0.94408

Answers using interpolation are accepted though not expected.

There is over 94.3% probability that X is greater than 50

using normal approximation to underlying gamma distribution

v) Probability calculated using normal distributional assumption is lower as compared to the answer obtained using chi square for the underlying Gamma distribution. [1]

As gamma distribution is positively skewed and has thick tail compared to normal and it will tend to be more like normal only when alpha tends towards infinity. As in this case, value of alpha is only 10, so normal approximation is not truly able to capture the correct thick tail found for Gamma.

12. :

(i) The Cramér-Rao Lower Bound result holds under very general conditions except where the range of the distribution involves the parameter, such as the uniform distribution in this case.

This is due to a discontinuity, so the derivative in the formula doesn't make sense.

(ii) We have $Y = \max_i X_i$.

Since each X_i lies between 0 and θ, the support for Y will be 0 and θ.

For $0 < y < \theta$,

$$P[Y < y] = P\left[\max_i X_i < y\right]$$

$$= P\left[\bigcap_{i=1}^{n}(X_i < y)\right]$$

$$= \prod_{i=1}^{n} P[X_i < y] \qquad [\because\ X_i s \text{ are independent}]$$

$$= \prod_{i=1}^{n}\left[\int_0^y \frac{dx}{\theta}\right] = \left(\frac{y}{\theta}\right)^n$$

Thus the probability density function of Y will be:

$$g_Y(y) = \frac{d}{dy}P[Y < y] = \frac{n}{\theta^n}.y^{n-1} \quad for\ 0 < y < \theta$$

Now, for any non-negative real number k,

$$E[Y^k] = \int_0^\theta y^k.\frac{n}{\theta^n}.y^{n-1}\,dy$$

$$= \frac{n}{n+k}\int_0^\theta \frac{n+k}{\theta^n}.y^{n+k-1}\,dy$$

$$= \frac{n}{n+k}.\frac{\theta^{n+k} - 0}{\theta^n}$$

$$= \frac{n\,\theta^k}{n+k}$$

(iii) Bias of the estimator $\hat{\theta}(c)$ is given as:

$$Bias[\hat{\theta}(c)] = E[\hat{\theta}(c) - \theta]$$

$$= E[c\,Y - \theta]$$

$$= c.E[Y] - \theta$$

$$= c.\frac{n\,\theta}{n+1} - \theta \quad [using\ the\ results\ in\ (b) with\ k = 1]$$

$$= \left(\frac{c\,n}{n+1} - 1\right).\theta$$

Mean Square Error of the estimator $\hat{\theta}(c)$ is given as:

$$MSE[\hat{\theta}(c)] = E\left[\{\hat{\theta}(c) - \theta\}^2\right]$$

$$= E[\{cY - \theta\}^2]$$

$$= E[c^2\,Y^2 - 2c\theta\,Y + \theta^2]$$

$$= c^2.E(Y^2) - 2c\,\theta\,E[Y] + \theta^2$$

$$= c^2.\frac{n\,\theta^2}{n+2} - c.\frac{2n\,\theta^2}{n+1} + \theta^2 \quad [using\ (ii)\ with\ k = 1\ \&\ k = 2]$$

(iv) For $\hat{\theta}(c)$ to be an unbiased estimator of θ, we need: $Bias[\hat{\theta}(c_u)] = 0$

$$\Rightarrow \left(\frac{c_u\,n}{n+1} - 1\right).\theta = 0 \quad \Rightarrow c_u = \frac{n+1}{n}.$$

(v) In order to minimize $MSE[\hat{\theta}(c)]$, we need:

$$0 = \frac{d}{dc}MSE[\hat{\theta}(c)] = \frac{d}{dc}\left[c^2.\frac{n\,\theta^2}{n+2} - c.\frac{2n\,\theta^2}{n+1} + \theta^2\right] = 2c.\frac{n\,\theta^2}{n+2} - \frac{2n\,\theta^2}{n+1}$$

Thus: $c_m = \frac{2n\,\theta^2}{n+1} \cdot \frac{n+2}{2n\,\theta^2} = \frac{n+2}{n+1}$.

$\left(\text{Note: } \frac{d^2}{dc^2} MSE[\hat{\theta}(c)] = \frac{d}{dc}\left[2c.\frac{n\,\theta^2}{n+2} - \frac{2n\,\theta^2}{n+1}\right] = \frac{2n\,\theta^2}{n+2} > 0 \quad \Rightarrow minimum\right)$

(vi) In order to minimize error in estimation, it is preferred to opt for an estimator which has lower mean square error among different competing estimators. So here $\hat{\theta}(c_m)$ will be preferred over $\hat{\theta}(c_u)$.

As n becomes large, $\hat{\theta}(c_u) = \frac{n+1}{n} Y \rightarrow Y$. Similarly $\hat{\theta}(c_m) = \frac{n+2}{n+1} Y \rightarrow Y$
Thus the two estimators becomes one and same as $n \rightarrow \infty$.

13.

i) Assumption : the sample is from a normal population. [0.5]

For a sample from normal distribution, we know that $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\bar{X} \sim N\left(2000, \frac{500^2}{10}\right)$ [0.5]

$P(\bar{X} < 1700) = P\left(Z < \frac{1700-200}{\frac{500}{\sqrt{10}}}\right)$ [1]

$= P(Z < -1.8974) = 1 - \Phi(1.8974) = 1 - 0.9711 = 2.889\ \%$ [1]

ii) Assumption : the sample is from a normal population. [0.5]

For a sample from normal distribution, we know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ [0.5]

$P(S^2 < 250^2) = P\left(\frac{(10-1)S^2}{500^2} < (10-1)\frac{250^2}{500^2}\right) = P(\chi^2_9 < 2.25) = 1.313\ \%$ [2]

iii) Assumption : the sample is from a normal population and hence the mean and variance are independent. [0.5]

$P(\bar{X} < 1700 \cap S^2 < 250^2) = P(\bar{X} < 1700).P(S^2 < 250^2) = 0.02889 \times 0.01313 = 0.038\%$ [0.5]

[7 Marks]

14.

i) The number of claims incurred by each policyholder follows the poisson distribution with mean 0.03. Therefore X, the number of claims for the 100 policyholders follows the Poi(3), X ~ Poi(3).

Since the poisson distribution only takes integer value P(X<6) = P(X<=5)
Using the poisson cumulative probability tables gives 0.91608

[2]

ii) Counting the numbers of trials up to and including the 4th success. This describes the variable (X) is Type 1 negative binomial distribution with k= 4 and p = 0.4

P(X=x) = $\binom{x-1}{3} 0.4^4 0.6^{x-4}$ x = 4,5,6,....

So P(X <7) = P(X=4) + P(X=5) + P(X=6)

P(X=4) = $\binom{3}{3} 0.4^4$ = 0.0256

Now using the iterative formula P(X=x) = $\frac{x-1}{x-4}$q P(X=x-1)

P(X=5) = $\frac{4}{1}$ ×0.6 × 0.0256 = 0.06144
P(X=6) = $\frac{5}{2}$ ×0.6 × 0.06144 = 0.09216

Hence, P(X <7) = 0.0256 + 0.06144 + 0.09216 = 0.1792

[2]

iii) Here the variable(X) is binomial distribution with n = 1000 and p = 0.015
Since n is large and p is small, hence poisson approximation can be used

Bin(1000,0.015) ~ Poi(15) (approximately)

Using the cumulative Poisson table gives

P(X <10) = P(X <=9) = 0.06985

15.

i) Wickets taken per 500 balls follow Poi(5) distribution. As the number of trials (balls) is very high and poisson parameter >= 5, we can use normal approximation to Poison Distribution. [0.5]

Thus the wickets take approximately follow N(5,5). [0.5]

Hence we need 'X' such that:

$P\left(Z > \frac{X-5}{\sqrt{5}}\right) = 0.95$ [0.5]

Critical value at 95% confidence is 1.65 [0.5]

Thus

$P\left(1.65 > \frac{X-5}{\sqrt{5}}\right) = 0.95$

Hence $X < 8.68$ [0.5]

As number of wickets can only take whole values, we need to truncate the number to lower whole number. Hence the team takes upto 8 wickets at 95 % confidence level. [0.5]

[3]

ii) For team B the runs in 50 ball will follow Bin(50,0.4) [0.5]

The mean and variance for this Binomial distribution are $50 \times 0.4 = 20$ *and* $50 \times 0.4 \times (1 - 0.4) = 12$ respectively [1]

For large number of trials and probability of success is close to 0.5 (or np>10), normal approximation can be applied & thus the runs per 50 balls follows approximately N(20,12) [1]

Probability of team B scoring 26 or more runs in 50 balls is thus $P\left(Z > \frac{26-20}{\sqrt{12}}\right) = 4.16\%$ [1]

The Poisson rate of taking wickets (by team A) is 1 per 100 balls i.e. 0.01 per ball. Hence, the wickets taken by team A in 50 balls has rate = 0.01 x 50 i.e it follows Poi(0.5) process. [1]

Probability of not taking any wicket is $\frac{0.5^0}{1} * e^{-0.5} = 60.65\%$ [1]

Hence probability of winning is 1-0.6065=39.34% [1]

Thus Team A has higher probability of winning. [0.5]

[7]

iii) Probability that team B bats for 30 balls = (Probability of waiting time (in terms of number of balls)> 30) x (Probability of A not scoring 26 runs in 30 balls) [0.5]

The waiting time has Exp(0.01) distribution, hence P(T>30)=exp(-0.01*30) = 74.08% [1]

Probability of A not scoring 30 runs = $P\left(Z < \frac{26-30\times0.4}{\sqrt{30\times0.4\times0.6}}\right) = 97.41\%$ [1]

Hence there is a 74.08% x 97.41% = 72.16% chance that team B will bat for 30 balls. [0.5]

[3]

[13 Marks]