Avani. Paresh. Khatri
Roll no. 40
Sec-A.

PAGE No
DATE / /

Probability & Statistics - 2

ASSIGNMENT. 1

Q.1 X ~ Poisson (θ)

Sol^n- Given $\frac{X-\theta}{\sqrt{\theta}} \sim N(0,1)$

An appropriate 95%. confidence interval.

$$P(Z_{2.5\%} < Z < Z_{97.5\%}) = 0.35$$

$$P(-1.96 < Z < 1.96) = 0.35$$ {1.96 from table book}

$$P\left(-1.96 < \frac{X-\theta}{\sqrt{\theta}} < 1.96\right) = 0.35$$

$$P(200 - 1.96\sqrt{200} < Z < 200 + 1.96\sqrt{200}) = 0.95.$$

$$P(172.28141 < X < 227.7185)$$

$$\boxed{\therefore \text{The CI is } (172.28141, 227.7185)}$$

PAGE No.
DATE / /

Q.2. (i)
Sol^n: 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.

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

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

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

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

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

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

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

$= \sigma^2$

$\therefore E(S^2) = \sigma^2$.

(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 & variance σ^2.
The variance of χ^2_k is $2k$.
Hence,

$Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1)$

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

$$Var[s^2] = \frac{2\sigma^4}{n-1}$$

(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 $Y = \frac{9s^2}{100}$ Then for $s^2 = 50$

$$Y = \frac{9(50)}{100}$$

$$Y = 4.5$$

and $s^2 = 150$

$$Y = \frac{9(150)}{100}$$

$$Y = 13.5$$

Thus,

$$\begin{aligned} P(50 < s^2 < 150) &= P(4.5 < Y < 13.5) \\ &= P(Y < 13.5) - P(Y < 4.5) \\ &= 0.8587 - 0.1245 \\ &= 0.7342. \end{aligned}$$

Q.3. Sol^n.

(i) The likelihood function is:

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

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

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

C is a constant

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

$$\frac{d\,Ln\,L}{dp} = \frac{-603}{(1-p)} + \frac{117}{p} = 0$$

$$p = \frac{117}{(603+117)}$$

$$= \frac{117}{720}$$

$$p = 0.1625$$

Checking for maximum:

$$\frac{d^2 LnL}{dp^2} = \frac{-603}{(1-p)^2} - \frac{117}{p^2} < 0$$

$\boxed{\therefore \text{ Maximum value for } Ln\,L \text{ is at } p = 0.1625}$.

(ii) Goodness of fit test:

We are testing the following hypothesis using a χ^2 goodness of fit test:

H_0 : the probabilities conform to a Bin(4,p) disⁿ.

H_1 : the probabilities do not conform to a Bin(4,p) disⁿ

Using $\hat{p} = 0.1625$ from part (i), the probabilities for this binomial distribution are:

$$P(X=0) = (1-p)^4 = 0.49197$$

$$P(X=1) = 4p(-p)^3 = 0.38183$$

$$P(X=2) = 6p^2(1-p)^2 = 0.11113$$

$$P(X=3) = 4p^3(1-p) = 0.01437$$

$$P(X=4) = p^4 = 0.000070$$

PAGE NO

DATE / /

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	Observed O_i	Expected E_i
0	86	88·55
1	75	68·73
2 +	19	20·00

The degrees of freedoms = 3−1−1 = 1.

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

$$= \frac{(86 - 88\cdot55)^2}{88\cdot55} + \frac{(75 - 68.73)^2}{68\cdot73} + \frac{(19 - 22\cdot72)^2}{22\cdot72}$$

$$= 0\cdot074 + 0\cdot572 + 0\cdot608$$

$$= 1\cdot254$$

This is less than the 5% critical value of 3·841. We have insufficient evidence at 5% level to reject H_0. Hence the model is a good fit.

Q.4. Sol^n:

The given pdf corresponds to two parameter version of the distribution given by

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

PAGE No.
DATE / /

$$E(x) = \frac{\lambda}{(\alpha-1)} \quad \text{and} \quad Var[x] = \frac{\alpha\lambda^2}{((\alpha-1)^2(\alpha-2))} \; ; \alpha>2$$

Thus the PDF of X,

$$f(x) = \frac{c\beta^3}{(x+\beta)^4} \; ; x>0 \text{ and } \beta>0$$

can be identified with $c=3$ (α known)

$\lambda=\beta$ (unknown)

$$E(X) = \frac{\beta}{2} = \theta.$$

$$V(X) = \frac{3}{4}\beta^2$$

Now,

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

$$\int_0^\infty \frac{c\beta^3}{(x+\beta)^4}\,dx = 1 \quad \text{implies}$$

$$\left[-\frac{c}{3}\cdot\frac{\beta^3}{(x+\beta)^3}\right]_0^\infty = 1 \quad \text{given } c=3.$$

(ii) The method of moment estimator :

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

$$\bar{X}_n = \frac{\beta}{2}$$

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

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

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

$$\begin{aligned} E(\hat{\beta}) &= E(2\bar{X}_n) \\ &= 2E(\bar{X}_n) \\ &= 2E(X) \end{aligned}$$

$$= 2 \times \frac{\beta}{2}$$

$$= \beta.$$

and Bias $= E(\hat{\beta}) - \beta = 0$

This estimator is unbiased.

Using the fact that the individual values of X_i are independent:

$$Var(\hat{\beta}) = Var(2\bar{X}_n)$$
$$= 4\,Var(\bar{X}_n)$$
$$= 4\,\frac{Var(X)}{n}.$$
$$= \cancel{4} \times \frac{3}{\cancel{4}} \beta^2 / n$$
$$= \frac{3\beta^2}{n}.$$

Hence, $MSE(\hat{\beta}) = \frac{3\beta^2}{n} + 0$

This estimator is consistent since MSE tends to zero as $n \to \infty$

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

$$= b^2 \times \frac{3}{4}\frac{\beta^2}{n} + (E(b\bar{X}_n) - \beta)^2$$
$$= b^2 \times \frac{3}{4}\frac{\beta^2}{n} + \left(\beta\left(\frac{b}{2} - 1\right)\right)^2$$
$$= b^2 \times \frac{3}{4}\frac{\beta^2}{n} + \left[\beta\left(\frac{b}{2} - 1\right)\right]^2$$

PAGE No
DATE / /

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

Differentiating the MSE respect to b:

$$\frac{d(MSE)}{db} = \frac{\beta^2}{n}\left(\frac{3}{2}b + n\left(\frac{b}{2}-1\right)\right).$$

Setting this equal to 0 gives $b = \frac{2n}{n+3}$

$$b = \frac{2}{(1+\frac{n}{3})}.$$

Differentiating the MSE a second time with respect to b we obtain:

$$\frac{d^2 MSE}{db^2} = \frac{\beta^2}{n}\left(\frac{3}{2} + \frac{n}{2}\right)$$ which is positive

$\therefore$ A minimum value for the MSE is at

$$\boxed{b = \frac{2}{(1+3/n)}}.$$

(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 $b = \frac{2}{(1+3/n)}$ is negatively biased estimator as $b < 2$ for $n \geq 1$.

As $n \to \infty$, b tend to 2 so, the estimator in (iii) is also consistent, in view of (iii)

Q.5. Sol^n:

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

$$b = 2E(X) - a$$
$$= 2\bar{x} - a$$

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

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

$$= \cancel{2\times}\frac{(\bar{x} - a)^2}{3}$$

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

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

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

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

$\bar{x} = 12$; $S = 21.27$.

Method of moments estimates using above formulae are:

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

For $U(a,b)$, the probability of a sample point being less than 'a' or greater than 'b' is zero and

PAGE NO.
DATE / /

We have a sample value ~~of~~ 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 w.r.t 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 $\hat{b} = \max(x_i)$

Q.6.(i) P(Type I error) is the probability of rejecting

Sol^n 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

$$P(\text{Type I error}) = P(X \geq 3 \mid p = 0.1) + P(X = 0 \mid p = 0.1) P(Y \geq 5 \mid p = 0.1) + P(X = 1 \mid p = 0.1) P(Y \geq 4 \mid p = 0.1) + P(X = 2 \mid p = 0.1) P(Y \geq 3 \mid p = 0.1).$$

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

$$= 0.14727 \approx 15\ \%.$$

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

$= p(X \geq 3 \mid p = 0.3) + P(X=0 \mid p = 0.3)\, P(Y \geq 5 \mid p = 0.3)$
$P(X = 1 \mid p = 0.3) P(Y \geq 4 \mid p = 0.3) +$
$P(X = 2 \mid p = 0.3)\, P(Y \geq 3 \mid p = 0.3)$.

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

$= 0.91253 \approx \underline{91\%}$

(iii) P (type II error) is the probability of accepting Ho when Ho is false.

$= 1 - 0.91253$
$= 0.08747 \approx \underline{9\%}$ (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 $\hat{p} = \frac{40}{120} = \frac{1}{3} = 0.3$

∵ Our calculate $Z_{10\%} = 1.2816$

Lower bound of 90% right-tailed confidence interval for p is

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

$$= 0.333 - 1.218 \sqrt{\frac{0.333(1-0.333)}{120}}$$

$$= \underline{0.2782}$$

(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:

$\frac{X - np_0}{\sqrt{np_0 q_0}} \sim N(0,1)$ 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 ~~100~~ 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 .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.

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

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

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 $S_p^2 = \frac{1}{25}\left(11(2916) + 14(4900)\right)$

$S_p^2 = 4027.04$

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

This is within ± 2.060 $(= t_{25, 2.5\%})$ so we have insufficient evidence to reject H_0 at the ~~same~~ 5% level.

$\therefore$ It is reasonable to conclude that there is no significant difference in the mean scores for the population associated with the two institutes.

Q.7. Sol^n:

(i) $\sum X_{Public} = 270$

$\therefore \bar{X}_{Public} = 30.$

$\sum X_{Private} = 315.5$

$\therefore \bar{X}_{Private} = 35.06$

$\sum X^2_{public} = 8614.5$

$\sum X^2_{private} = 11599.25$

$$S_1^2 = \frac{\sum X^2_{Public} - (n \times \bar{X}_{Public})}{n-1}$$

$$= \frac{8614.5 - (9 \times 30^2)}{8}$$

$$= 64.31$$

$$S_2^2 = \frac{11599.25 - (9 \times 35.06^2)}{8}$$

$$= 67.42$$

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

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

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

Value of Statistic is $\dfrac{64.31}{67.42} = 0.9542.$

$F_{8,8}$ values at 5% levels are 0.2256 and 4.433. Since the value of the test statistic is between the above results, 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
$H_1 : \mu_1 < \mu_2$

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

Where, $S_p^2 = \dfrac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{(n_1 + n_2 - 2)}$

Using the values in section (i) above,

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

$$S_p^2 = 65.86$$

Value of 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

---

Q.8. (i) Unbiased estimator :

Consideration of the sampling distribution of an estimator can give an indication of how good it is as an estimator.

$E(\theta) = \theta$

(ii) Bias : When θ used to estimate $\hat{\theta}$, and it doesn't give value same as θ.

(iii) Mean Square Error (MSE) of this estimator $\hat{\theta}$
→ Variance of estimator with respect to true parameter values.

$$MSE = Variance + Bias^2.$$

(iv) As MSE of $\hat{\theta}$ is ~~true~~ less, it is more consistent

(v) They would be termed as consistent, when MSE decreases or gradually becomes zero.

(vi) • Method of moments -
It is a method of estimation of population parameters
• Method of Likelihood Estimator -
It is a method of estimating the parameters of an assumed probability distribution.

Q.9. Sol^n:
(i) Variable t_k is defined as $t_k \Rightarrow \dfrac{N(0,1)}{\sqrt{\chi_k^2/k}}$.

Where k denotes the degree 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 $\frac{k}{k-2}$ respectively

PAGE No
DATE / /

(iii) (a) We know that for a sample from a normal population, $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

For ~~a~~ the given confidence level $t_q = 3.25$, 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. $\underline{(42.83, 57.17)}$

(b) From (ii) above we know that

$$t_{n-1} \sim N\left(0, \sqrt{\frac{n-1}{n-3}}\right) \sim N\left(0, \sqrt{\frac{9}{7}}\right).$$

i.e.

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

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

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. $\underline{(43.54, 56.45)}$.

Q.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 $Poi(n\mu)$

i.e. $Poi(3000)$ Where μ is 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 $n\lambda$ 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 $N\left(\mu, \frac{\hat{\mu}}{n}\right)$.

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 $\underline{(2.7613, 3.0387)}$.

Q.11 Revised Sample Mean $= \frac{2,13,200 + 11,000 + 48,000 + 27,500 + 31,500}{12}$

$= ₹\ 17,767.$

Revised $\Sigma x^2 = 4,919,860,000 - 1,21,000,000 - 2,304,000,000 + 756,250,000 + 992,250,000$

$= \underline{4,243,360,000.}$

Therefore: Revised sample ~~Mean~~ Standard Deviation

$$= \left\{\frac{1}{11}\left(4,243,360,000 - \frac{213,200^2}{12}\right)\right\}^{1/2}$$

$$= \left\{\frac{1}{11}(455506,666.7)\right\}^{1/2}$$

$= \underline{6435.}$

Comment: There has been no charge in the sample mean. However there has been a reduction in the Sample Standard Deviation from 10,144 to 6,435.

The mean remains unaltered as the total salary of the temporary employees who were being replaced is equal to the total salary of the permanent employees who replaced them. The reason for the reduction in standard deviation is that the salaries of the permanent employees who replaced temporary employees are closer to the sample mean.

Q.12. (i) The Cramer-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, 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 $0 < y < \theta$,

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

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

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

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

Thus the probability density function of y will be;

$$g_Y(y) = \frac{d}{dy} P[Y < y]$$

$$= \frac{n}{\theta^n} \cdot y^{n-1} \quad \text{for } 0 < y < \theta$$

Now, for any non-negative real number k,

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

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

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

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

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

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

$$= E[cY - \theta]$$

$$= c \cdot E(Y) - \theta$$

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

$$= \left(\frac{cn}{n+1} - 1\right) \cdot \theta$$

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

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

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

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

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

$$= c^2 \cdot \frac{n\theta^2}{n+2} - c \cdot \frac{2n\theta^2}{n+1} + \theta^2$$

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

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

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

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

$$= \cancel{4}\, 2c \cdot \frac{n\theta^2}{n+2} - \frac{2n\theta^2}{n+1}$$

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

(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, $\hat{\theta}(c_u) = \frac{n+1}{n} Y \to Y$.

Similarly $\hat{\theta}(c_m) = \frac{n+2}{n+1} Y \to Y$

Thus the two estimators becomes one and same as $n \to \infty$