Probability and Statistics

Assignment (1)

Date / /
Page

Q1 $\dfrac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} \sim N(0,1)$

$\rightarrow$ $\hat{\lambda} = \dfrac{\Sigma x_i}{n}$; $P(z_{2.5\%} < z < z_{97.5\%}) = 0.95$

$\hat{\lambda} = 200$

$\Rightarrow P(-1.96 < z < 1.96) = 0.95$

$\Rightarrow P\left(-1.96 < \dfrac{\hat{\lambda}-\lambda}{\sqrt{\lambda/n}} < 1.96\right) = 0.95$

$\Rightarrow P\left(-1.96\sqrt{\dfrac{\lambda}{n}} - \hat{\lambda} < \lambda < 1.96\sqrt{\dfrac{\hat{\lambda}}{n}} - \hat{\lambda}\right) = 0.95$

$\Rightarrow P\left(\hat{\lambda} - 1.96\sqrt{\dfrac{\lambda}{n}} < \lambda < \hat{\lambda} + 1.96\sqrt{\dfrac{\hat{\lambda}}{n}}\right) = 0.95$

$\Rightarrow$ so, there is 95% confidence interval of λ for $\hat{\lambda} = 200$ and $n = 1$

$\Rightarrow$ 95% CI for $\lambda = \left(200 - 1.96\sqrt{200},\ 200 + 1.96\sqrt{200}\right)$

$= (172.281,\ 227.719)$

Q2 $\rightarrow$ X_i, $i = 1, 2, \dots n$ are iid.

Mean $= \mu$

Variance $= \sigma^2$

$\bar{x} = \dfrac{\Sigma x_i}{n}$

$= \quad s^2 = \frac{1}{n-1}\left(\Sigma x_i^2 - n\bar{x}^2\right)$

$=$ (i) $E(\bar{x}) = E\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n}\Sigma E(x_i)$

$= \frac{n\mu}{n} = \mu$

$= \quad V(\bar{x}) = V\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n^2}\Sigma V(x_i) = \frac{n\sigma^2}{n^2}$

$= \frac{\sigma^2}{n}$

(ii) $E(s^2) = \sigma^2$

$E(s^2) = E\left[\frac{1}{n-1}\left(\Sigma x_i^2 - n\bar{x}^2\right)\right]$

$= \frac{1}{n-1}\Sigma E(x_i^2) - \frac{n}{n-1}E(\bar{x}^2)$ — (1)

$=$

Now, $E(Y^2) - (E(Y))^2 = V(Y)$

$\therefore \quad E(x_i^2) - [E(x_i)]^2 = V(x_i)$

$E(x_i^2) = \sigma^2 + \mu^2$ — (2)

$\|^{ly}$ for $E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$ — (3)

$=$ From (2) & (3) Putting in (1),

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

$= \frac{n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}}{n-1} = \sigma^2$

(iii) Now, $X \sim N(\mu, \sigma^2)$

$$= \frac{(n-1) s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$= V\left(\frac{(n-1) s^2}{\sigma^2}\right) = \frac{(n-1)^2 V(s^2)}{\sigma^4}$$

$$= V\left(\chi^2_{n-1}\right) = \frac{(n-1)^2 V(s^2)}{\sigma^4}$$

$$= 2\cancel{(n-1)} = \frac{(n-1)^{\cancel{2}} V(s^2)}{\sigma^4}$$

$$= \left[V(s^2) = \frac{2\sigma^4}{(n-1)} \right]$$

(iv) $P(0.5\sigma^2 < s^2 < 1.5\sigma^2)$

$= P(s^2 < 1.5\sigma^2) - P(s^2 < 0.5\sigma^2)$

$= n = 10, \ \sigma^2 = 100$

$$P(s^2 < 1.5\sigma^2) = P(s^2 < 150)$$

$$= P\left(\frac{9s^2}{\sigma^2} < \frac{\cancel{150}^{3} \times 9}{\cancel{100}_{2}}\right)$$

$$= P(\chi^2_9 < 13.5)$$

$$= 0.8587$$

$= P(s^2 < 0.5\sigma^2) = P(s^2 < 50)$

$= P\left(\frac{9s^2}{\sigma^2} < \frac{50 \times 9}{100}\right)$

$= P(\chi^2_9 < 4.5)$

$= 0.1245$

$= P(50 < s^2 < 150) = 0.8587 - 0.1245$

$= \underline{\underline{0.7342}}$

Q3 let no. of claims be $x \xrightarrow{} $ random variable

$x \sim Bin(4, p)$

(i) $L = \prod_{i=1}^{180} f(x_i)$

$= (P(X=0))^{76} \times (P(X=1))^{75} \times (P(X=2))^{16} \times (P(X=3))^{2} \times (P(X=4))^{1}$

$= \left({}^4C_0\, p^0 (1-p)^4\right)^{76} \times \left({}^4C_1\, p^1 (1-p)^3\right)^{75} \times \left({}^4C_2\, p^2 (1-p)^2\right)^{16} \times \left({}^4C_3\, p^3 (1-p)\right)^2 \times \left({}^4C_4\, p^4\right)^1$

$\Rightarrow L = K \times p^{117} \times (1-p)^{603}$

$K \rightarrow$ constant

$= \log(L) = \log(k) + 117 \log p + 603 \log(1-p)$

= differentiating on both sides w.r.t (p)

$$= \frac{dl}{dp} = 0 + \frac{117}{p} - \frac{603}{(1-p)}$$

$$= \frac{dl}{dp} = \frac{117}{p} - \frac{603}{1-p}$$

$$0 = \frac{117}{p} - \frac{603}{1-p} \Rightarrow 117 - 117p = 603p$$

$= 117 = 720p$

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

$= \boxed{\hat{p} = 0.1625}$

(ii) Goodness of fit test

→ Ho : no. of claims confirms both binomial model

H₁ : no. of claims doesn't confirm to the binomial model.

x	0	1	2	3	4
O_i	86	75	16	2	1
e_i	89.542	68.724	19.998	2.574	0.108

$= p = 0.1625$

$= P(X=0) = {}^4C_0\, p^0 (1-p)^4 = 0.4819$

$= P(X=1) = {}^4C_1\, p^1 (1-p)^3 = 0.3818$

$= P(X=2) = {}^4C_2\, p^2 (1-p)^2 = 0.111$

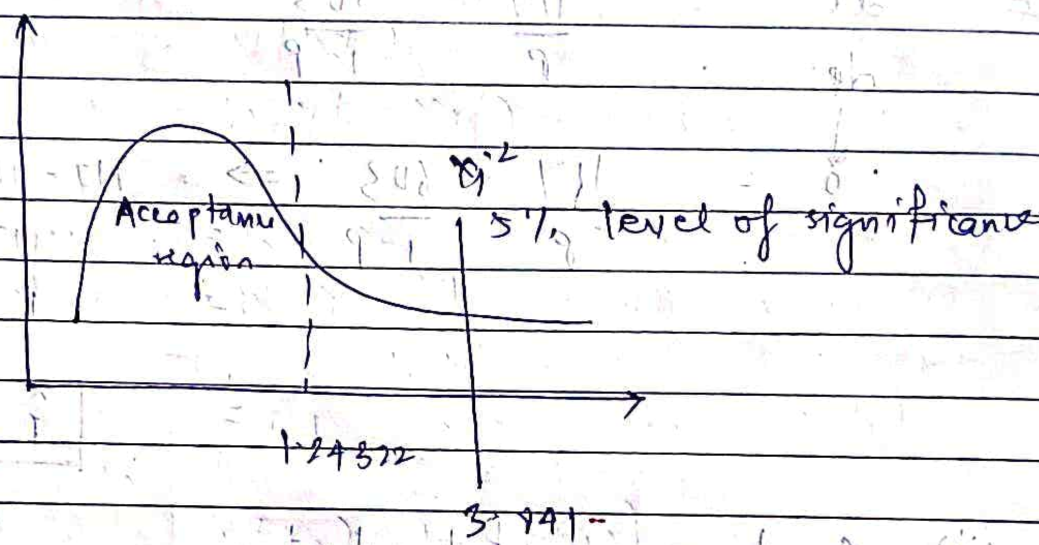
|||ly $P(X=3) = 0.0143$

$P(X=4) = 0.0006$

Oi	86	75	19
Ei	88·542	68·724	22·68

$\Rightarrow$ $T \cdot S = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{5-1-1-2}$

$1 \cdot 2432 \sim \chi^2_1$



$\Rightarrow$ We have insufficient evidence to reject Ho at 5% level of significance. Therefore no. of claims conforms to the binomial Model

Q4. $f(x) = \frac{c\beta^3}{(x+\beta)^4}$; $x > 0$; $\beta > 0$

1) Comparing the above $f(x)$ to P.D.F of pareto dis^n

$\Rightarrow$ For Pareto, PDF $= \frac{\alpha \lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}$

Thus, β corresponds to λ

Thus $c = 3$ (as c corresponds to α here)

$$E(X) = \frac{\lambda}{\alpha - 1}$$

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

$$= V(X) = \frac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = \frac{3\beta^2}{4}$$

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

$$\frac{\beta}{2} = \bar{X} \Rightarrow \boxed{\hat{\beta} = 2\bar{X}}$$

$$\text{Bias} = E(\hat{\beta}) - \beta = E(2\bar{X}) - \beta$$

$$= \frac{2}{n}\sum E(X_i) - \beta$$

$$= \frac{\cancel{2}}{\cancel{n}} \times \frac{\cancel{n}\beta}{\cancel{2}} - \beta = 0 \quad \text{Zero}$$

$= \therefore \{\text{Bias} = 0\}$ so unbiased.

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

$$= V(\hat{\beta})$$

$$= V(2\bar{X})$$

$$= 4V(\bar{X}) = \frac{4V(\sum X_i)}{n^2} = \frac{4 \times 3\beta^2 \cancel{n}}{4 \times n^{\cancel{2}}}$$

$$= \frac{3\beta^2}{n}$$

As n increases, MSE tends to 0

$\therefore \hat{\beta}$ is consistent.

iii) $\hat{\beta} = b\bar{x}$

$$E(\hat{\beta}_2) = E(b\bar{x})$$
$$= bE\left(\frac{\sum x_i}{n}\right)$$
$$= \frac{b}{n} \times \frac{\beta}{2} \times n = \frac{b\beta}{2}$$

$$\therefore V(\hat{\beta}_2) = V(b\bar{x})$$
$$= b^2 V\left(\frac{\sum x_i}{n}\right)$$
$$= \frac{b^2}{n^2} \times \frac{3}{4}\beta^2 \times n$$
$$= \frac{3}{4}\frac{b^2\beta^2}{n}$$

$$\therefore MSE = V(\beta_2) + (Bias)^2$$
$$= \frac{3b^2\beta^2}{4n} + \left(\frac{b\beta}{2} - \beta\right)^2$$
$$= \frac{3b^2\beta^2}{4n} + \left(\frac{b\beta - 2\beta}{2}\right)^2$$
$$= \frac{3b^2\beta^2}{4n} + \frac{b^2\beta^2 + 4\beta^2 - 4\beta^2 b}{4}$$
$$= \frac{3b^2\beta^2 + nb^2\beta^2 + 4n\beta^2 - 4n\beta^2 b}{4n}$$

$$\therefore MSE = \frac{\beta^2}{4n}\left(3b^2 + nb^2 + 4n - 4nb\right)$$

Date / /
Page

let MSE be M

$$\frac{dM}{db} = \frac{\beta^2}{4n}(6b + 2nb - 4n)$$

$$\frac{dM}{db} = 0$$

$$\therefore \frac{\beta^2}{4n}(6b + 2nb - 4n) = 0$$

$$= 6b + 2nb = 4n$$

$$= \cancel{\frac{dM}{db} = 0} \quad b(3+n) = 2n$$

$$b = \frac{2n}{3+n}$$

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

re differentiating,

$$\frac{d^2M}{db^2} = \frac{\beta^2}{4n}(6 + 2n)$$

$$\frac{\beta^2}{4n}(6 + 2n) > 0$$

$$= \therefore \frac{d^2M}{db^2} \text{ is minimum}$$

$$= \text{Thus } b = \frac{2}{1 + 3/n} \text{ minimises the MSE}$$

(iv) $E(\hat{\beta}_2) = \dfrac{b\beta}{2}$

$$= \frac{2 \times \beta}{2(1+3/n)} = \frac{n\beta}{n+3}$$

$\Rightarrow$ $Bias = E(\hat{\beta}_2) - \beta$

$$= \frac{n\beta - \beta(n+3)}{n+3}$$

$$= \frac{n\beta - n\beta - 3\beta}{n+3} = -\frac{3\beta}{n+3}$$

$\Rightarrow$ So, we've a -ve bias & hence it underestimates the true parameter β

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

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

$$= \frac{\beta^2}{4n}\left[\frac{4n^2\cancel{(n+3)}}{(n+3)^2\cancel{(n+3)}} + 4n\left(\frac{3-n}{n+3}\right)\right]$$

$$= \beta^2\left[\frac{\cancel{4n^2} + 12n - \cancel{4n^2}}{n+3}\right]$$

$$= \frac{12n\beta^2}{n+3}$$

$\Rightarrow$ Since $n \to \infty$, MSE can't tend $\to 0$

$\Rightarrow$ Therefore it is inconsistent

Date / /
Page

Q5 (i) $X \sim U(a,b)$

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

$= \bar{x} = \frac{a+b}{2}$

$\bar{x} = \frac{a+b}{2}$

$= a = 2\bar{x} - b$ —— (1)

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

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

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

$= 12S^2 = (b-a)^2$

$= 2\sqrt{3}S = b-a$

$= 2\sqrt{3}S = 2\bar{x} - a - a$

$= 2\sqrt{3}S = 2\bar{x} - 2a$

$= \sqrt{3}S = \bar{x} - a$

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

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

$b = 2\bar{x} - \bar{x} + \sqrt{3}S$

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

(iii) Sample of observations : 1, 2, 8, 50, 150

$= \bar{x} = 42.2$

$S = 63.57$

$= \hat{a} = \bar{x} - \sqrt{3}S = -67.906$

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

$= 152.3064$

= So, $\hat{b}$'s value is very close to sample value but $\hat{a}$ is very far from the true sample value, so it is the ~~pro~~ potential weakness of method of Moments

(v) $x \sim U(0, b)$

$$L = \prod_{i=1}^{n} f(t_x)$$

$$= \left(\frac{1}{b}\right)^n$$

$$= \log L = -n \log b$$

$$\Rightarrow \frac{dl}{db} = -\frac{n}{b}$$

$\frac{dl}{db} = 0$, $n = 0$ or $b = \infty$

= which is not possible

$= \text{Also, } \frac{d^2 l}{db^2} = \frac{n}{b^2}$

$= \frac{n}{b^2} > 0$ which we have a MLE

= Thus, we say, $\hat{b} = \max(x_i)$

Q6 $H_0 : p = 0.1$

$H_1 : p > 0.1$

(i) P(Type 1 error) = P(rejecting H_0 | H_0 is true)

Date / /
Page

= let X be the no. of hits in first 12 mins
Y be the no. of hits in step 12 mins

= $P(\text{Type I Error}) = P(X \geq 3 / p = 0.1) + P(X = 0 / p = 0.1)$
$+ P(Y \geq 5 / p = 0.1) + P(X = 1 / p = 0.1) *$
$P(Y \geq 4 / p = 0.1) + P(X = 2 / p = 0.1) *$
$P(Y \geq 3 / p = 0.1)$

$= (1 - 0.8891) + 0.2824(1 - 0.9857) +$
$(0.6580 - 0.2824)(1 - 0.8744) +$
$(0.8891 - 0.6590)(1 - 0.8891)$
$= 0.14727 \approx 15\%$

(ii) Probability (rejecting null hypothesis when $p = 0.3$)
$= P(X \geq 3 / p = 0.3) + P(X = 0 / p = 0.3) P(X \geq 5 / p = 0.3)$
$+ P(X = 1 / p = 0.3) P(X \geq 4 / p = 0.3) + P(X = 2 / p = 0.3)$
$* P(Y \geq 3 / p = 0.3)$

$= (1 - 0.2528) + 0.0138(1 - 0.7237) +$
$(0.0410 - 0.0138)(1 - 0.4825) + (0.2528 - 0.0850)($
$1 - 0.2528) = 0.9125 \approx 91\%$

= (iii) P(type II error Accepting Ho when Ho is false)
$= 1 - 0.81253$
$0.08747 \approx 9\%$ (where $p = 0.3$)

(iv) 10 space agencies in aggregate fired 120 missiles and recorded 40 hits

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

$\because$ $n = 120$, $x = 40$

$$\hat{p} = \frac{40}{120} = \frac{1}{3} = 0.3333$$

$\Rightarrow$ $Z_{10\%} = 1.2816$

Lower bound of 90% right failed 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) $p > 0.278$ at 10% LOS

$H_0 = p = 0.278$ & $H_1 = p > 0.278$

$\Rightarrow$ One sample binomial Model

$$\frac{x - np_0}{\sqrt{np_0 q_0}} \sim N(0,1) \text{ with continuity correction}$$

$$= \frac{395 - 120(0.278)}{\sqrt{120 \times 0.278 \times 0.722}} = \cancel{1.2511}$$

Hence we can't reject H_0 at 10% LOS.

(vi) (iv) suggests that there is 10% chance that $P < 0.2782$ while in the hypothesis test could be less than or equal to 0.2782 in 90% of the cases.

(vii) $n_1 = 12$ $\quad \sigma_1 = 54$

$\bar{x}_1 = 65$ $\quad \sigma_2 = 70$

$\bar{x}_2 = 70$ $\quad n_2 = 15$

$\rightarrow H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Assumption: Pop^n variance of both samples are equal.

$$TS: \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_1 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

$$\therefore S_1 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

$$= 4027.04$$

$$= t_{cal} = \frac{(65-70)}{\sqrt{4027.04\left(\frac{1}{120}+\frac{1}{15}\right)}} = \frac{-5}{\sqrt{24.57}} = -0.2035$$

$= t_{tab} = \cancel{(-2060)} \; (-2.060, 2.060)$

Since, t_{cal} is in 95% of CI, we won't reject H_0.

4) i) Sample Mean for :-

Private Hospital $(\bar{x}_{pr}) = \frac{\sum x_i}{n} = \frac{315.5}{9}$

$= 35.056$

Public Hospitals $(\bar{x}_{pu}) = \frac{270}{9} = 30$

Sample Variance :

$= S^2_{pr} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = 61.403$

$= S^2_{pu} = 64.3125$

Date / /
Page

ii) For using equal mean, we can use a 2-sided t-test assuming that the popⁿ variation of both is equal.

For verifying this assumption

$= H_0 : \sigma_1^2 = \sigma_2^2 \; ; \; H_1 : \sigma_1^2 \neq \sigma_2^2$

$= \dfrac{S^2_{pr} / S^2_{pu}}{\sigma^2_{pr} / \sigma^2_{pu}} \sim F_{n_{pr}-1,\, n_{pu}-1}$

For 95% CI

$= 1\left[\dfrac{1}{x} < F_{8,8} < x \right] = 0.95$

$= 1\left[\dfrac{1}{4.433} < \dfrac{67.4036\,\sigma^2_{pu}}{64.3125\,\sigma^2_{pr}} < 4.433 \right]$

$=$ 95% of CI $= (0.2152, 4.229)$

Since at 5% LOS, CI includes 1, we do not reject Ho. Our assumption of equal population variance holds true in this case

$H_0 = \mu_{pr} = \mu_{pu} \; ; \; H_1 = \mu_{pr} > \mu_{pu}$

Date / /
Page

This is a one sided test

Since pop^n var is unknown we use t-dist^n

$$t_{cal} = \frac{(\bar{x}_{pr} - \bar{x}_{pu}) - (\mu_{pr} - \mu_{pu})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{5.056}{8.6756\sqrt{2/9}} = \frac{5.056 \times 3}{\sqrt{2} \times 8.6756}$$

$t_{tab} = t_{16, 5\%} = 1.746$

Since, $t_{cal} < t_{tab}$; we have insufficient advance to reject H_0 at 5% level of significance

Q8. An estimator whose mean value is equal to the ~~esti~~ true parameter value (θ) is known as Unbiased estimator

The difference b/w true value of estimator and expected value of parameter is Biased estimator

(iii) The mean square error of an estimator θ is defined by $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2$

(iv) They both would be considered to be consultant when MSE minimise such that when $n \to \infty$, $MSE \to 0$.

(v) (i) Method of Moments.
(ii) Maximum likelihood estimate

Q9 (i) Variable t_k is defined as $t_k \Rightarrow \frac{N(0,1)}{\sqrt{\chi^2_k / k}}$

while k denotes the degree of freedom and the 2 random variables $N(0,1)$ and χ^2_k are independent.

(ii) For $k > 2$
$E(t_k) = 0$,
$V(t_k) = k/_{k-2}$

(iii) (a) $\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

$= P\left(t_{9,0.5\%} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{9,99.5\%}\right) = 0.99$

$= P\left(\bar{x} - t_{9,0.5\%} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{9,0.5\%} \frac{s}{\sqrt{n}}\right) = 0.99$

Date / /
Page

$t_{9, 0.5\%} = 3.25$

99% confidence interval for μ:

$$= \left(50 - 3.25 \times \sqrt{\frac{48.667}{10}},\ 50 + 3.25 \times \sqrt{\frac{48.667}{10}}\right)$$

$$= (42.83,\ 57.17)$$

(b) From (ii)

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

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

$$= \text{i.e. } \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim N\left(0, \sqrt{\frac{9}{7}}\right)$$

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

= Critical values level of confidence
2.58

= Confidence interval $= \left(50 - 2.58\sqrt{\frac{48.667 \times 9}{10 \times 7}},\ 50 + 2.58\sqrt{\frac{48.667 \times 8}{10 \times 7}}\right)$

Date / /
Page

Confidence interval = (43.54, 56.45)

Q10 $n = 1000$

$\lambda = 3$

$\longrightarrow$ $X \sim Poi(n\lambda)$

$X \sim Poi(3000)$

(i) ~~P(type error~~

$P(\text{type I error}) = P(\text{reject } H_0 / H_0 \text{ is true})$

$= P(X > 3100 / X \sim Poi(3000))$

= Using Normal approximations,

$= P(X > 3100) = P\left(Z > \dfrac{3100 - 3000}{\sqrt{3000}}\right)$

$= 1 - P(Z < 1.825)$

$= 3.39\%$

(ii) Type II error = event of accepting the hypothesis when it is false.

(iii) Power of test $= P(\text{rejecting } H_0 / H_0 \text{ if false})$

$= P[X > 3100 \mid X \sim Poi(n\lambda) - N(n\lambda, n\lambda)]$

$= 1 - P\left(Z < \dfrac{3100 - n\lambda}{\sqrt{n\lambda}}\right)$

The value of pass of test will depend on the value of parameter under after native hypothesis

(iv) $\hat{\mu}$ is estimator of μ, $\hat{\mu}$ followed $N(\mu, \frac{\hat{\mu}}{n})$

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

or, $P\left(-2.578 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.5758\right) = 0.99$

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

Q11

$n = 12$

$\Sigma X = 213,200$

$\Sigma X^2 = 4,919,860,000$

$\bar{X} = 17,767$

$s = 10,144$

= New Sample Mean

= $\Sigma X' = \Sigma X - 11,000 - 48000 + 27500 + 31500$

$=$ ~~213200~~ 213200

$\bar{X}' = \frac{213200}{12} = 17767$

$\therefore$ the new sample SD.

$$\sum X'^2 = 4919860000 - (11000)^2 - (48000)^2 + (27500)^2 + (31,500)^2$$
$$= 4,243,360,000$$

$$\therefore S'^2 = \frac{1}{n-1}\left(\sum X'^2 - n(\bar{X}')^2\right)$$
$$= \frac{1}{11}\left(4,243,360,000 - 12(17767)^2\right)$$
$$= \underline{41396775 \cdot 64}$$

$$\underline{\underline{S' = 6434 \cdot 0326}}$$

$\therefore$ The sample remains the same here too as the 2 values removed were outliers and the 2 values excluded were around the sample mean.

The sample standard devn reduces as the 2 values now excluded do not deviate very much from the sample mean causing the variance to reduce and hence the reduction of the sample standard deviation

Q12 $x = x_1, x_2, \ldots \ldots x_n$

$X \sim U(0, \theta)$

(i) LRCB for variance of unbiased estimates of θ doesnt apply here as the range of the dist^n involves the parameter hence the derivative in the formula doesn't make sense

(ii) $\hat{\theta}(c) = cy$, here c :- const
$y = \max x_i$

Since $y = \max(x_i)$

$F(y) = F(x_{max}) = F(\max(x_1, x_2, \ldots x_n))$

$= 1 - P(X \geq \max(x_1, x_2, \ldots, x_n))$

$= 1 - P(X \geq x)^n$

$= P(X < x)^n = [F(x)]^n$

$F(y) = \left(\frac{x - 0}{\theta - 0}\right)^n = \left(\frac{x}{\theta}\right)^n = \left(\frac{y}{\theta}\right)^n$

$= f(y) = F'(y) = \frac{d}{dy}\left(\frac{y}{\theta}\right)^n$

Date / /
Page

$= \frac{1}{\theta^n}[n y^{n-1}], \quad 0 < y < \theta$

$= E(y^k) = \int_0^\theta y^k f(y)\, dy$

$= \int_0^\theta y^k \left(\frac{n}{\theta^n} y^{n-1}\right) dy$

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

$= \frac{n}{\theta^n}\left[\frac{y^{n+k}}{k+n}\right]_0^\theta$

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

$= \boxed{E(y^k) = \frac{n\theta^k}{n+k}}$ ---- (1)

Date / /
Page

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

$= E(cy) - \theta$

$= c\,E(y) - \theta$

$=$ using (1) at $k = 1$

$= \text{Bias} = c\left[\frac{n\theta}{n+1}\right] - \theta$

$= \theta\left[\frac{cn}{n+1} - 1\right]$

$= \text{MSE}[\hat{\theta}(c)] = \text{var} + \text{Bias}^2$

$v(\hat{\theta}(c)) = v(cy) = c^2 v(y)$

$= c^2\left[E(y^2) - [E(y)]^2\right]$

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

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

$= \text{MSE} = c^2\left[\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{n+1}\right] +$

$\left[\theta\left(\frac{cn}{n+1} - 1\right)\right]^2$

$$= \frac{c^2 n\theta^2}{n+2} - \frac{c^2 n^2 \theta^2}{(n+1)^2} + \theta^2\left[\frac{c^2n^2}{(n+1)^2} + 1 - \frac{2cn}{n+1}\right]$$

$$\left\{ MSE = \frac{c^2 n\theta^2}{(n+2)} - \frac{c\,2n\theta^2}{n+1} + \theta^2 \right\}$$

(iv) For $\hat{\theta}(c)$ to become an unbiased estimator of θ

Bias $= 0$

$\Rightarrow E(\hat{\theta}(c)) - \theta = 0$

Hence,

$\Rightarrow \therefore E(\hat{\theta}(c)) = \theta$

$\Rightarrow E(cY) = \theta$

$\Rightarrow cE(Y) = \theta$

$\Rightarrow c\left(\frac{n\theta}{n+1}\right) = \theta$

$\Rightarrow cn = n+1$

(v) $MSE = \frac{c^2 n\theta^2}{n+2} - c\left(\frac{2n\theta^2}{n+1}\right) + \theta^2$

$\longrightarrow \frac{d(MSE)}{dc} = \frac{2cn\theta^2}{n+2} + \frac{2n\theta^2}{n+1} - \frac{2n\theta^2}{n+1} = 0$

$\Rightarrow \frac{2cn\theta^2}{n+2} = \frac{2n\theta^2}{n+1}$

$\Rightarrow \; Cn + C = n + \alpha$

$\Rightarrow \; C = \dfrac{n+\alpha}{n+1}$

$\Rightarrow \; \dfrac{d^2}{dC^2}(MSE) = \dfrac{2n\theta^2}{n+\alpha} > 0$

$\Rightarrow$ Minimised at $C = \dfrac{n+\alpha}{n+1}$

$C_M = \dfrac{n+\alpha}{n+1}$

(vi) $\hat{\theta}(C_m)$ is preferred over $\hat{\theta}(C_u)$ as in the latter's case it could so happen that some of the estimators are very large while other small, but on average give a true value.

$\Rightarrow$ In the former case, it can have a value which doesn't much deviate from the true value

$\Rightarrow$ For $n \to \infty$, both $\hat{\theta}(C_u)$ & $\theta(C_M)$ become equal a $\hat{\theta}(C_u) = C_u \cdot Y = \left(\dfrac{n+1}{n}\right)Y$ $\to Y_i$

$\Rightarrow$ Similarly $\hat{\theta}(C_m) = C_m Y \to Y$