PROBABLITY AND STATISTICS - 2
UNIT 1 AND 2

2. $x_i = 1, 2, \ldots n$

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

i) $\bar{x} = \frac{\Sigma x_i}{n}$

$$E(\bar{x}) = \frac{E(\Sigma x_i)}{n} = \frac{\Sigma E(x_i)}{n}$$

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

$$\boxed{E(\bar{x}) = \mu}$$

$$V(\bar{x}) = V\left(\frac{\Sigma x_i}{n}\right)$$

$$= \frac{1}{n^2} V(\Sigma x_i) = \frac{1}{n^2} \Sigma V(x_i)$$

$$= \frac{1}{n^2} \Sigma \sigma^2 = \frac{n\sigma^2}{n^2}$$

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

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

$V\left[\frac{(n-1)S^2}{\sigma^2}\right] = \chi^2_n \; V[\chi^2_{n-1}]$

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

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

iv) $n = 10, \; \sigma^2 = 100$

$P(0.5 < S^2 < 0.5) = P(S^2 < 0.5) - P(S^2 < -0.5)$

$P\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)0.5}{\sigma^2}\right) - P\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)(-0.5)}{\sigma^2}\right)$

$P(\chi^2_9 < 0.045) - P(\chi^2_9 < -0.045)$

$P(\chi^2_9 < 0.045) - P(\chi^2_9 > 0.045)$

$P(\chi^2_9 < 0.045) - [1 - P(\chi^2_9 < 0.045)$

$= 0.315$

3.

No. of claims	0	1	2	3	4
No. of policies	80	78	16	2	1

$$\text{Pdf} = {}^nC_r \, p^r q^{n-r}$$

$$L = \prod_{i=1}^{4} f(x_i)$$

$$= \prod_{i=1}^{4} {}^nC_{x_i} \, p^{x_i} q^{n-x_i}$$

$$= \sum {}^nC_{x_i} \cdot p^{\sum x_i} q^{\sum(n-x_i)}$$

$$\ln L = \cancel{\ln} \sum x_i \ln(p) + \sum(n-x_i)\ln(q) + \text{const.}$$

$$= \sum x_i \ln(p) + \sum(n-x_i)\ln(1-p) + \text{const.}$$

$$\frac{d \ln L}{dp} = \frac{\sum x_i}{p} - \frac{\sum(n-x_i)}{1-p} = 0$$

$$= \sum(x_i)\cdot(1-p) - [\sum(n-x_i)]p$$

$$= \sum x_i - p\sum x_i = p\sum(n-x_i)$$

$$\sum x_i - p\sum x_i = p4n - \sum x_i$$

iii) ~~Hem The above mention~~

5] $U(a,b)$.

i) To prove: $\hat{a} = \bar{x} - \sqrt{3}s$

Using MoM

Sample mean = Population mean

$$\bar{X} = \frac{1}{2}(a+b)$$

$$\hat{a} = 2\bar{X} - b \qquad b = 2\bar{X} - a$$

$$s^2 = \frac{1}{12}(b-a)^2$$

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

$$s\sqrt{12} = b - a$$

$$a = \cancel{b} - \cancel{s\sqrt{12}}$$

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

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

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

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

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

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

Hence Proved

ii) $\bar{X} - \frac{1}{2}(a+b)$

$$a = 2\bar{x} - b$$

$$s^2 = \frac{1}{12}(b-a)^2$$

$$12 \cdot s^2 = (b-a)^2$$

$$2s \cdot \sqrt{3} = b - a$$

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

Page No.
Date

$$2S\cdot\sqrt{3} = b - 2\bar{x} + b$$
$$= 2b - 2\bar{x}$$
$$= 2(b - \bar{x})$$
$$S\sqrt{3} = b - \bar{x}$$
$$\therefore b = \bar{x} + \sqrt{3}S.$$

iii) The above meondioned estimavors con't be used as the mean for a unijorm dist^n is ~~a~~ alwoogs 0.

iv) $U(0, b)$

$$f(x_i) = \frac{1}{b-0} = \frac{1}{b}$$

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{b} = \frac{1}{b^n} = b^{-n}$$

$$\ln L = -n \cdot \ln b$$

$$\frac{d\ln L}{db} = \frac{-n}{b} = 0$$

$$\hat{b} \neq \infty \quad \therefore \hat{b} = \infty$$

$$\frac{d^2 \ln L}{db} = \frac{n}{b^2} > 0$$

$\therefore$ It's minimum

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

8.] i) Unbiased estimator is an estimator such that it is equal to the True value of the parameter

ii) Bias is the difference between the expectation of the estimator and true value.

$Bias = E(g(x)) - \theta$

iii) MSE is the variance of the estimator wrt the true parameter value

$MSE = Var + Bias^2$

iv) $\tilde{\theta}$ with a positive bias ~~as~~ is preferred over the other as MSE rules over biasedness. Even though it has a positive bias it would more accurately represent the data.

v) When both the parameter have a lower MSE they will be termed as consistency

vi) θ can be estimated by either Method of Moments method or by Most likelihood method.

9] In sampling distribution $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_k$ is usu...

$\bar{x}$ - sample mean
μ - Population mean
s - sample variance
n - size of sample

ii) For t_k with $k > 2$
Mean : 0
Variance : $\frac{k}{k-2}$

iii) $n = 10$,
$\bar{x} = 50$
$s^2 = 48.067$

99% CI : $\bar{x} \pm t_{\alpha/2, n-1} (s/\sqrt{n})$

$= \bar{x} \pm t_{0.5\%, 9} (s/\sqrt{n})$

$= 50 \pm 3.250 (\sqrt{48.067/10})$
$= (42.8303, 57.1696)$

iv) $\mu = 0$, $\sigma^2 = \frac{9}{9-2} = 1.2857$

~~99% CI / $\mu \pm Z_{\alpha/2}(\sigma/\sqrt{n})$~~
$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

99% CI = $\bar{x} \pm Z_{\alpha/2} (\sigma/\sqrt{n})$
$= 139.444 \pm Z_{0.5\%} (\sqrt{1.2857/10})$
$= (139.44, 140.38)$

11] $n = 12$

$\Sigma x = 213200$

$\Sigma x^2 = 4919860000$

Sample mean : 17767

Sample s.d : 10144

Revise $\Sigma x = 213200 - 11000 - 48000 + 27500 + 31500$

$= 213200$

Revisem mean $= \frac{213200}{12} = 17767$

New s.d

$\Sigma x^2 = 4919860000 - (11000)^2 - (48000)^2 + (27500)^2 + (31500)^2$

$= 4243360000$

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

$= \frac{1}{11} (4243360000 - 12(17767)^2)$

$= 41396775.64$

$s = 6437.0326$

12.) The CRLB results hold under very general conditions except whe the range of distribution involves parameter such as the uniform distribution in this case

ii) $Y = \max X_i$

$P[Y < y] = P[\max X_i < y]$
$= P[\cap (X_i < y)]$
$= \prod_{i=1}^{n} P[X_i < y]$
$= \prod_{i=1}^{n} \int_0^y \frac{dx}{\theta} = \left(\frac{y}{\theta}\right)^n$

Pdf: $\frac{n}{\theta^n} y^{n-1}$ for $0 < y < \theta$

$E[y^k] = \int_0^\theta y^k \frac{n}{\theta^n} \cdot y^{n-1} \, dy$
$= \frac{n}{n+k} \int_0^\theta \frac{n+k}{\theta^n} y^{n+k-1} \, dy$
$= \frac{n}{n+k} \cdot \frac{\theta^{n+k}}{\theta^n} - 0$
$= \frac{n\theta^k}{n+k}$

iii) $\text{Bias}[\hat{\theta}(c)] = E[\hat{\theta}(c) - \theta]$
$= E[cY - \theta]$
$= c\,E(Y) - \theta$
$= c \cdot \frac{n\theta}{n+1} - \theta$
$= \left(\frac{cn}{n+1} - 1\right)\theta$

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

Page No.
Date

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

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

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

$$c = \frac{n+1}{n}$$

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

$$0 = \frac{d}{dc} MSE[\hat{\theta}(c)]$$

$$= \frac{d}{dc}\left[\frac{c^2 n\theta^2}{n+2} - \frac{c2n\theta^2}{n+1} + \theta^2\right]$$

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

Thus,

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

vi) In order to minimize error in estimation, it is preferred to opt for an estimator which has lower MSE θ, hence in this case $\hat{\theta}(c_m)$ will be preferred.

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

$$\hat{\lambda} = \frac{\sum y_i}{n} = 200$$

$P(-1.96 < Z < 1.96) = 0.95$

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

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

95% CI for λ when $\hat{\lambda} = 200, n = 1$

$(200 - 1.96\sqrt{200},\; 200 + 1.96\sqrt{200})$

$(172.28191,\; 227.7185)$