Tisha Kothari
A – 41

PROBABILITY & STATISTICS ASSIGNMENT.

(1) $X \sim Poi(\theta)$

$$\frac{X-\theta}{\sqrt{\theta}} \sim N(0,1) \qquad X = 200$$

$$\therefore \theta = X \pm 1.96 \left(\sqrt{\theta/n}\right)$$

(2) $X_i \sim N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

i) $$E(\bar{X}) = \frac{E(\Sigma X_i)}{n} - \frac{n\mu}{n}$$

$$= \mu$$

$$Var(\bar{X}) = \frac{Var(\Sigma X_i)}{n}$$

$$= \frac{1}{n^2}(n\sigma^2)$$

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

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

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

$$= \sigma^2$$

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

$$= \frac{2\sigma^4}{n-1}$$

iv) $$(50 < s^2 < 150)\left(\frac{9\times 50}{100} < \chi^2_9 < \frac{9 \times 150}{100}\right)$$

$$= (4.5, 13.5)$$

(3) $n = 4$ $\quad Bin(4, p)$ $\quad$ 180 Policies.

i)

No of claims	0	1	2	3	4
No of policies	86	75	16	2	1

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

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

$$= (1-p)^{603} (p)^{117} \text{ constant}$$

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

$$\frac{d}{dp} \ell(p) = \frac{-603}{1-p} + \frac{117}{p}$$

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

$$117 - 117p = 603p$$

$$\therefore p = 0.1625$$

ii) $P(X=0) = (1-p)^4 = 0.492$

$P(X=1) = 4p(1-p)^3 = 0.3818$

$P(X=2) = 6p^2(1-p)^2 = 0.111$

$P(X=3) = 4p^3(1-p) = 0.0144$.

$P(X=4) = 0.0007$

	0	1	2	3	4
O_i	86	75	16	2	1
E_i	88.56	68.72	19.98	2.59	0.126

By clubbing,

	0	1	2
O_i	86	75	19
E_i	88.56	68.72	22.71
$C_i = \frac{\Sigma(O_i - E_i)}{E_i}$	0.0757	0.5739	0.6061

$\Sigma c_i = 1.2557 - \chi_1^2$

$\therefore 1.2537 < 3.841$

$\therefore$ Do not reject H_0.

(4) i) $f(x) = \dfrac{c\beta^3}{(x+\beta)^4}$; $x > 0$, $\beta > 0$

$$\int_0^\infty f(x)\,dx = 1$$

$$c\beta^3 \int_0^\infty \frac{1}{(x+\beta)^4} = 1$$

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

$$\frac{c\beta}{-3}\left[0 - \beta^{-3}\right] = 1$$

$$c = 3$$

$$E(X) = \int_0^\infty x.f(x)\,dx$$

$$= 3\beta^3 \int_0^\infty \frac{x}{(x+\beta)^4}\,dx$$

$$= 3\beta^3 \int_0^\infty \frac{x+\beta}{(x+\beta)^4} - \frac{\beta}{(x+\beta)^4}\,dx$$

$$= 3\beta^3 \left[\frac{(x+\beta)^{-2}}{-2} - \frac{\beta(x+\beta)^{-3}}{-3}\right]_0^\infty$$

$$= 3\beta^3 \left[0 - 0 - \left[\left(\frac{\beta^{-2}}{-2}\right) - \left(\frac{\beta^{-2}}{-3}\right)\right]\right]$$

$$= 3\beta^3 \left[\frac{\beta^{-2}}{2} - \frac{\beta^{-2}}{3}\right]$$

$$= \frac{\beta^3}{2}\left[\beta^{-2}\right] = \frac{\beta}{2}$$

$$E(X^2) = \int_0^\infty x^2 \cdot f(x)\, dx$$

$$= 3\beta^3 \int_0^\infty \frac{x^2}{(x+\beta)^4}\, dx$$

Using Pareto distribution

$$\frac{3\beta^2}{5} = V(x) \quad \text{where } \alpha = 3 \ \& \ \lambda = \beta$$

ii) $\beta = X_1, X_2, \ldots, X_n$

$$E(X) = \bar{X}$$

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

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

$$E(\hat{\beta}) = E(2\bar{X}_n)$$

$$= 2E\left[\frac{\Sigma X_i}{n}\right]$$

$$= \frac{2}{n} \Sigma E(X_i)$$

$$= \frac{2}{n}(n)\left(\frac{\beta}{2}\right)$$

$$= \beta$$

$\therefore$ Unbiased.

$$Var(\hat{\beta}) = Var(2\bar{X}_n)$$

$$= 4\,Var[\bar{X}_n]$$

$$= 4\,var\,X / n$$

$$= 4\left(\frac{3\beta}{4}\right) / n$$

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

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

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

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

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

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

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

Diff. the MSE wrt 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.

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

$= \frac{2}{1+3/n}$

Diff. MSE for the 2nd time wrt b.

$\frac{d^2(MSE)}{db^2} = \frac{\beta^2}{n}\left(\frac{3}{2} + \frac{n}{2}\right) > 0.$

$\therefore$ Min value for MSE at $b = \frac{2}{1+3/n}$

iv) As it seems in (ii) that $\hat{\beta} = 2\bar{X}_n$ is an unbiased estimator. The estimator $b\bar{X}_n$ in (iii) where $b = 2/1+3/n$. is very biased estimator for $b<2$, $n \geq 1$.

As $n \to \infty$, b tends to 2. so estimator in (iii) is also consistent as in (ii).

(5) (i) $E(X) = \bar{X}$

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

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

$$V(X) = s^2$$

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

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

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

$$2\bar{X} - b = b - 2\sqrt{3}\, s$$

$$2\hat{b} = 2\bar{X} - 2\sqrt{3}\, s$$

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

$$\hat{a} = 2\bar{X} - \bar{X} - \sqrt{3}\, s$$

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

(iii) Sample $\rightarrow$ 1, 2, 3, 4, 50

$\bar{x} = 12$ $\qquad$ $s = 21.27$

$$\hat{a} = 12 - \sqrt{3}(21.27) = -24.84$$

$$\hat{b} = 12 + \sqrt{3}(21.27) = 48.84$$

For $U(a,b)$, the probability of a sample pt being less than 'a' or greater than 'b' is zero and we have a sample value 50 that is greater than our estimate 'b'.

It highlights a potential weakness of MOM.

(iv) $X \sim U(0, b)$

$$f(x) = \frac{1}{b-a}$$

$$= \frac{1}{b}$$

$L = f(x_1) \cdot f(x_2) \cdots f(x_n)$

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

$= \frac{1}{b^n}$

$l = -n \log b$

$\frac{dl}{db} = -\frac{n}{b} < 0 \quad , b \to \infty$

$\frac{d^2 l}{db^2} = \frac{n}{b^2} > 0 \quad$, min at $b \to \infty$

So, using common sense, $L(\theta)$ is max when θ is minimum $b > \text{Max } X_i$

$\hat{b} = \text{Max } X_i$

(6) (i) P(type 1 error) = P(Reject H_0 | H_0 is T)

X be no of hits in 1st step 12 missiles

Y be no of hits in 2nd step 12 missiles

P(Type 1 error) $= P(X = 3 / p = 0.1)$
$+ P(X = 0 / p = 0.1) \cdot P(Y = 5 / p = 0.1)$
$+ P(X = 1 / p = 0.1) \cdot P(Y = 4 / p = 0.1)$
$+ P(X = 2 / p = 0.1) \cdot P(Y = 3 / p = 0.1)$

$= (1 - 0.8891) + (0.2824)(1 - 0.9957)$
$+ (0.6590 - 0.2824)(1 - 0.9944) +$
$(0.8891 - 0.6590)(1 - 0.8891)$

$= 0.14727 \approx 15\%$

(ii) $P(X = 3 / p = 0.3) + P(X = 0 / p = 0.3) \cdot P(Y = 5 / p = 0.3)$
$+ P(X = 1 / p = 0.3) \{ P(Y = 4 / p = 0.3) \} + P(X = 2 / p = 0.3)$
$P(Y = 3 / p = 0.3) \}$

$= 0.91253 \approx 91\%$

(iii) P(type II error) is the prob of accepting H_0 where H_0 is False.

$= 1 - 0.91253$

$= 0.08747 \approx 9\%$. $(p = 0.3)$

(iv) 10 space agencies fire 120 missiles and record 40 hits.

$X \sim Bin(120, 1/3)$ $\quad \hat{p} = \frac{40}{120} = \frac{1}{3}$

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

$$P\left(-1.2816 < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < 1.2816\right)$$

At CI 90% $\quad p \in (0.3333\ 0.2782)$

(v) $p > 0.278$ at 10% LS

$H_0 : p = 0.278$

$H_1 : p > 0.278$

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

$$\frac{34.5 - 120(0.278)}{\sqrt{120(0.278)(0.722)}} = 1.2511$$

At TS < 1.2816

We have insufficient evidence to reject H_0 at LS 10%.

(vi) Lower Bound of CI implies that there is only a 10% chances "p" ≤ 0.2782, whereas

from the hypothesis test, 'p' could be less than equal to 0.2780. with prob more than 10%. (approx 10.5% corresponding to 1.25119)
The minor disconnect b/w CI & HT at some level is due to:
1) Applying continuity correction in HT.
2) Use of sample proportion $\hat{p}$ to estimate pop^n variance in calculating CI.

(viii) $H_0 : \mu_1 = \mu_2$ $\quad H_1 : \mu_1 \neq \mu_2$

PQ $\rightarrow$ $\dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$

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

$$= 4027.04$$

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

Since TS lies in A.R, we have insufficient evidence to reject H_0 at 5% LS.

(7) i) $\Sigma X_{Pub} = 270$, $\quad \bar{X}_{Pub} = 30$

$\Sigma X_{Pri} = 315.5$, $\quad \bar{X}_{Pri} = 35.06$

$\Sigma X^2_{Pub} = 8614.5$

$$S_1^2 = \frac{8614.5 - 9(30)^2}{8}$$

$$= 64.31$$

$$S_2^2 = \frac{11599.25 - 9(35.06)^2}{8} = 67.42$$

(ii) $H_0 : \sigma_1^2 = \sigma_2^2$ $\quad H_1 : \sigma_1^2 \neq \sigma_2^2$

$$TS = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{n_1-1, n_2-1}$$

$$= \frac{64.31}{67.42}$$

$$= 0.9542 \sim F_{8,8}$$

Since TS lies in AR, we have insufficient evidence to reject H_0 at LS 5%.

∴ Pop variance

(iii) $H_0 : \mu_1 = \mu_2$ $\quad H_1 : \mu_1 < \mu_2$.

$$TS = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{Sp^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1, n_2 - 2}$$

$$Sp^2 = \frac{64.31(8) + 67.42(8)}{16}$$

$$= 65.86$$

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

$$= 1.32$$

Since TS lies in AR we have insufficient evidence to reject H_0 at 5% LS.

∴ Cost of claims in private hospitals is similar to that of public hospitals.

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

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

(iii) MSE of this estimator $\hat{\theta} = [E(\theta) - \theta]^2$

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

(v) As estimator is termed as consistent if MSE converges to zero as sample size tends to ∞.

(vi) θ can be estimated using:

a) Method of Moments: The pop moments are equated to the sample moments to estimate the parameters.

b) Maximum Likelihood Method: A max likelihood f^n $L(\theta) = \prod_{i=1}^{n} f(x_i ; \theta)$ is generated. A MLE of parameter is given by sol to $\frac{dL(\theta)}{d\theta} = 0$

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

(i) $\sqrt{\frac{t_K}{K}}$

(ii) $\mu(t_K) = 0$ $\qquad var(t_K) = \frac{K}{K-2}$

(iii) $n = 10$ $\quad \bar{X} = 50$ $\quad s^2 = 48.667$ $\quad CI = 99\%$.

(a) $t_{0.005, 9} = 3.25$

$$\mu = \bar{X} \pm t_{0.5\%, 9} \frac{s}{\sqrt{n}}$$

$$= 50 \pm 3.25 \sqrt{\frac{48.667}{10}}$$

$$= 57.17, 42.83$$

(b) $t_{0.005, 9} = 2.5758$

$$\mu = 50 \pm 2.5758 \sqrt{\frac{48.667}{10}}$$

$$= 55.68, 44.32$$

(10) $n = 1000$ $\quad X \sim Poi(3)$ $\quad X \sim N(3000, 3000)$

If $P(X < 3100)$ then $X \sim Poi(3)$

else X does not follow $Poi(3)$.

(i) Type 1 error $= P(H_0 \text{ is reject} / H_0 \text{ is true})$

$= P(X > 3100, \lambda = 3)$

$\therefore X \sim N(3000, 3000)$

$$P\left(Z > \frac{3100-3000}{\sqrt{3000}}\right) = P(Z > 1.826)$$

$= 3.39\%$

(ii) Type 2 error $=$ P(H_1 accept / H_0 is false)

(iii) Power $= P(\text{reject } H_0 / H_0 \text{ false})$

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

(11) $n = 12$ $\quad \Sigma x = 213200$ $\quad \Sigma x^2 = 4919860000$

$\bar{x} = 17767$ $\quad S = 10144$

$X_1 = 11000$ $\quad X_2 = 48000$ $\quad X_1' = 27500$ $\quad X_2' = 31500$

$\Sigma x = 213200 = A + 11000 + 48000$

$A = 154200$

$\Sigma x' = A + x_1' + x_2' = 213200$

$\bar{x}' = 17767$

$\Sigma x^2 = 4919860000 = B + (11000)^2 + (48000)^2$

$B = 2474860000$

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

$= 6434.03$

(12) (i) The CRLB results holds under very general conditions except where range of dist^n

involves the parameter, such as uniform distribution in this case. This is due to a discontinuity so the derivative in the formula doesn't make any sense.

(ii) $Y = \text{Max } X_i$

Since X_i lies b/w 0 and θ, the support for Y will be 0 & θ. For $0 < y < \theta$.

$$P(Y < y) = P\left[\text{Max } X_i < y\right]$$
$$= P\left[\bigcap_{i=1}^{n} (X_i < y)\right]$$
$$= \prod_{i=1}^{n} P\left[X_i < y\right]$$
$$= \prod_{i=1}^{n} \left[\int_0^y \frac{dn}{\theta}\right]$$
$$= \left(\frac{y}{\theta}\right)^n$$

Thus PDF of Y will be $g_y(y) = \frac{d}{dy} P[Y < y]$

$$= \frac{n}{\theta^n} \cdot y^{n-1}$$

for $0 < y < \theta$

Now for $k \geq 0$

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

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

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

MSE of estimator $\hat{\theta}(c)$ is given by:

$$\text{MSE}[\hat{\theta}(c)] = E[(\hat{\theta}(c) - \theta)^2]$$
$$= E[(cY - \theta)^2]$$
$$= E[c^2y^2 - 2c\theta y + \theta^2]$$
$$= c^2E[y^2] - 2c\theta E(y) + \theta^2$$
$$= \frac{c^2 n\theta^2}{n+2} - \frac{c(2n)\theta^2}{n+1} + \theta^2$$

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

$$\left[\frac{c_u . n}{n+1} - 1\right]\theta = 0$$
$$c_u = \frac{n+1}{n}$$

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

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

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

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

$$= \frac{2n\theta^2}{n+2} > 0 \Rightarrow \min$$

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

As n becomes large $\hat{\theta}(c_u) = \frac{n+1}{n} y \to y$

||ly, $\hat{\theta}(c_M) = \frac{n+2}{n+1} y \to y$, the two estimators become one & same as $n \to \infty$.