Assignment

Roll.no: 86 Subject: Probability & Statistic 2

Name: Kshiti Vartak Section: B

1] $\theta_x = 200$ $P(X \geq 200) = 0.025$ $P(X \leq 200) = 0.025$

Considering Poisson (θ_1) we have:

$P(X \geq 200) = 0.025 \Rightarrow P(X > 199.5) = 0.025$ ($\because$ C.C)

$P\left(Z > \frac{199.5 - \theta_1}{\sqrt{\theta}}\right) = 0.025 \Rightarrow \frac{199.5 - \theta_1}{\sqrt{\theta_1}} = 1.96$

$\Phi(1.96) = 0.975 = 1 - 0.025$.

For θ_1: $(199.5 - \theta_1)^2 = 1.96^2 \theta_1 \Rightarrow$ 173.67 or 229.17

Lower bound $\theta_1 = 173.67$

Considering Poisson(θ_1) we have:

$P(X \leq 200) = 0.025 \Rightarrow P(X < 200.5) = 0.025$ ($\because$ C.C.)

$P\left(Z < \frac{200.5 - \theta_2}{\sqrt{\theta_2}}\right) = 0.025 \rightarrow \frac{200.5 - \theta_2}{\sqrt{\theta_2}} = -1.96$ $\Phi(-1.96) = 0.025$

For θ_2: $(200.5 - \theta_2)^2 = (-1.96)^2 \theta_2$

$= 174.60$ or 230.24

$\therefore$ Upper bound, $\theta_2 = 230.24$

2] i) $\bar{x} = \frac{\sum X_i}{n}$

$E[\sum X_i] = \sum E[X_i] = \sum \mu = n\mu$ ($\because$ iid)

$Var[\sum X_i] = \sum Var[X_i] = n\sigma^2$; ($\because$ iid)

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

$V[\bar{x}] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

ii) $s^2 = \frac{1}{n-1} [\sum X_i^2 - n\bar{x}^2]$

$E[s^2] = \frac{1}{n-1} (\sum E[X_i^2] - n E[\bar{x}^2])$

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

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

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

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

The variance of χ^2_k is $2k$

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

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

iv) $\sigma^2 = 100$, $n = 10$, $P(50 < s^2 < 150) = ?$

We know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ and we have χ^2_9

Let $Y = 9s^2/100$. Then, for $s^2 = 50$: $Y = 9(50)/100 = 4.5$
and for $s^2 = 150$: $Y = 9(150)/100 = 13.5$

Thus, $P(50 < s^2 < 150) = P(4.5 < Y < 13.5)$
$= P(Y < 13.5) - P(Y < 4.5)$
$= 0.8587 - 0.1245$
$= 0.7342$

3] i) The likelihood f^n is:

$L(p) = c[(1-p)^4]^{86}[4p(1-p)^3]^{75}[6p^2(1-p)^2]^{16}$
$\cdot[4p^3(1-p)]^3[p^4]^1$
$= c[(1-p)^{344}][p^{75}(1-p)^{225}][p^{32}(1-p)^{32}]$
$[p^6(1-p)^2][p^4]$
$= c(1-p)^{603}p^{117}$

c is a constant.

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

$d\ln L/dp = -603/(1-p) + 117/p = 0$

$\Rightarrow p = \frac{117}{(603+117)} = \frac{117}{720} = 0.1625$

Checking for maximam:

$d^2\ln L/dp^2 = -603/(1-p)^2 - 117/p^2 < 0$

$\therefore$ Max value for $\ln L$ is at $p = 0.1625$

ii) Goodness of fit test:

H_0: Prob. conform to Bin(4, p)

H_1: Prob. conform to Bin(4, p)

Using $\hat{p} = 0.1625$, the Binomid Prob are

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

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

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

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

$P(X=4) = p^4 = 0.00070$

The expected values are 88.55, 68.73, 20.00, 2.59 & 0.13.

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

nerating

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

$$\chi^2 = \Sigma(O_i - E_i)^2/E_i$$
$$= (86-88.55)^2/88.55 + (45-68.73)^2/68.43$$
$$+ (19-22.72)^2/22.72$$
$$= 0.074 + 0.572 + 0.608 = 1.254$$

This is less than 5% critical value of 3.841. We have insufficient evidence at 5% level to reject H_0. Hence, the model is a good fit.

4] i) $f(x) = \dfrac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$; $x>0$, $\lambda>0$ 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, $f(x) = \dfrac{c\beta^3}{(x+\beta)^4}$; $x>0$ and $\beta>0$ can be identified with $c=3$

$\lambda = \beta$

$\underset{[0.9}{E[X]} = \beta/2$ $\quad \underset{[0.9]}{var[x]} = 3/4\,\beta^2$

2:

DATE / /

4] ii] MOM estimator;

$\bar{x}_n = \beta/2 \Rightarrow \hat{\beta} = 2\bar{x}_n$

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 unbaised.

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

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

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

Differentiating the MSE w.r.t. b:

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

$\therefore b = 2n/(n+3) = 2/\left(1 + \frac{3}{n}\right)$ [Consider \= 0]

Taking double derivative

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

MSE has min value at $b = 2/(1 + 3/n)$

iv) $\hat{\beta} = 2\bar{x}_n$ is an unbaised estimator.

The estimator $b\bar{x}_n$ when $b = 2/\left(1 + \frac{3}{n}\right)$

is -vely biased.

DATE / /

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

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

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

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

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

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

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

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

iii) Sample : 1, 2, 3, 4, 60

$\Rightarrow \bar{x} = 14$, $s = 28$

mom estimates are :

$\hat{a} = 14 - \sqrt{3}(28) = -34.497$

$\hat{b} = 14 + \sqrt{3}(28) = 62.49$

For $U(a,b)$, the probability of a sample point being less than 'a' or greater than 'b' is zero and in our sample we have 60 that is greater than our estimate '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 as the range of x depends on b.

We want b to be small, $b \geq max(x_i)$

maximum is attained at $b = max(x_i)$

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

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

i) 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) is the probability of rejecting H_0 when H_0 is true.

$P(\text{Type I error}) = P(X \geq 3 \mid p=0.1) + P(X=0 \mid p=0.1)$
$P(Y \geq 5 \mid p=0.1) + P(X=1 \mid p=0.1)$
$P(Y \geq 4 \mid p=0.1) + P(X=2 \mid p=0.1)$
$P(Y \geq 3 \mid p=0.1)$
$= (1-0.8891) + (0.2824)(1-0.9957) +$
$(0.6590) - 0.2824)(1-0.9744) +$
$(0.8891 - 0.6590)(1-0.8891)$
$= 0.14727 \approx 15\%$

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

$= P(X \geq 3 \mid p=0.3) + P(X=0 \mid p=0.3) P(Y \geq 5 \mid p=0.3)$
$+ P(X=1 \mid p=0.3) P(Y \geq 4 \mid p=0.3) +$
$P(X=2 \mid p=0.3) P(Y \geq 3 \mid p=0.3)$
$= (1-0.2528) + (0.0138)(1-0.7237) +$
$(0.0850 - 0.0138)(1-0.4925) + (0.2528 - 0.0850)$
$(1-0.2528)$
$= 0.91253 \approx 91\%$

iii) P(type II error) is the probability of not rejecting H_0 when H_0 is false.

$= 1 - 0.91253 = 0.08747 \approx 9\%$ (when $p = 0.3$)

iv) 120 missiles ~~were~~ fired by 10 space agencies in aggregate and 40 hits were recorded.
We assume that the sample comes from binomial distribution.

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

Here $n=120$, $X=40$, $\hat{p} = \frac{40}{120} = 0.3333$

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

$$= 0.2782$$

v) Test whether $p > 0.278$ at 10% level of significance
$H_0 : p = 0.278$ vs $H_1 : p > 0.278$

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

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

We ~~are~~ carried out a one-sided test. The value of test statistic is less than 1.2816 so we do not have sufficient evidence to reject H_0 at 10% level)

vi) As per the lower bound of Confidence interval there is only 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%.

vii) $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

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

we have, $\bar{x}_1 = 65$, $\bar{x}_2 = 70$; $S_1 = 54$, $S_2 = 70$,
$n_1 = 12$, $n_2 = 15$

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

$\dfrac{65 - 70 - 0}{68.459 * \sqrt{\frac{1}{12} + \frac{1}{15}}} = -0.2034$

The is within $\pm 2.060 (= t_{25;\,2.5\%})$ 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 insitudes.

7] i) $\sum X_{public} = 270 \Rightarrow (\bar{X}_{public}) = 30$

$\sum X_{private} = 315.5 \Rightarrow (\bar{X}_{private}) = 35.06$

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

$s_1^2 = (8614.5 - 9 \times 30^2)/8 = 64.81$

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

$s_2^2 = (11599.25 - 9 \times 35.06^2)/8 = 67.42$

ii) Here we have equal variances.

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

Test Statistic is $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-2}$

$\frac{64.81}{67.42} = 0.9542$

The values of $F_{8,8}$ at 5% level are 0.2256 and 4.433.

iii) $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 < \mu_2$

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

$s_p^2 = \frac{64.81 \times 8 + 67.42 \times 8}{16} = 65.86$

value of T.S. is $\frac{(35.06 - 30) - 0}{\sqrt{65.86(1/9 + 1/9)}} = 1.32$

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

Since $P(t_{16} > 1.32)$ is higher the 95% we do not reject.

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

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

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

iv] $\hat{\theta}$ is efficient

v] As sample size tends to ∞, MSE gets to 0 and then the estimator is consistent.

vi] a) Method of moments

b) Maximum likelihood method

c) Bootstrap method

9] i) Variable t_K is defined as $t_K \Rightarrow \frac{N(0,1)}{\sqrt{\chi^2_K / K}}$

ii) Mean and variance of t_K for $K > 2$ are 0 and $K(K-2)$ respectively.

iii) a) Confidence level $t_9 = 3.25$,

Confidence ~~leve~~ interval for μ is thus

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

$$\Rightarrow (42.83, 57.17)$$

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

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

critical value for given level of confidence is 2.58.

10] i] Type I error:
Let X be the random variable denoting the total number of claims on the portfolio. X thus follows Poi($n\mu$)
Null hypothesis H_0 is thus $X \sim Poi(8000)$

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

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

ii] Type II Error is the event of accepting the hypothesis when it is false.

iii]

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

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

iv) If $\hat{\mu}$ is the estimator of μ,
$\hat{\mu}$ follows $N(\mu, \mu/n)$
Hence $\frac{\hat{\mu} - \mu}{\sqrt{\hat{\mu}/n}}$ follows $N(0,1)$

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

Hence the confidence interval for μ is (2.7613, 3.0387) ii] New sample mean = (213200 − 11000 − 48000 + 27800 + 31600) / 12

$= (213200/12)$

$= 17767$

New $\Sigma x^2 = 4919860000 - 121000000 - 2304000000 + 756250000 + 992250000$

$= 4243360000$

New Sample S.D $= \left\{ \frac{1}{11} \left(4243360000 - \frac{(213200)^2}{12} \right) \right\}^{1/2}$

$= \left(\frac{1}{11} (455506666.7) \right)^{1/2}$

$= (41409696.97)^{1/2}$

$= 6435.$

A reduction in Sample S.D. was observed. and there was no change in sample mean

12] 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 the car