SHIVAM PATEL 200070077 DS 203 Assn 3

Ans 1. Repeated question from Assignment 2, Q 9b. Question left unanswered as per instructions given in lecture.

Ans 2. $\mu = E(X_i)$ $\sigma^2 = \text{var}(X_i)$

$$Y = \frac{\sum_{i=1}^{n} (X_i - \mu)}{\sqrt{n\sigma^2}}$$

$$E(Y) = E\left(\frac{\sum_{i=1}^{n} (X_i - \mu)}{\sqrt{n\sigma^2}}\right) = \frac{\sum_{i=1}^{n} E(X_i) - n\mu}{\sqrt{n\sigma^2}}$$

$$= \frac{n\bar{X} - n\mu}{\sqrt{n\sigma^2}} = \frac{n(\mu - \mu)}{\sqrt{n\sigma^2}} = 0.$$

$$\text{var}(Y) = E\left[\left(\frac{\sum_{i=1}^{n} (X_i - \mu)}{\sqrt{n\sigma^2}}\right)^2\right] - (E(Y)) \times (E(Y))$$

$$= \frac{\sum_{i=1}^{n} E(X_i - \mu)^2}{n\sigma^2} - 0 \qquad [\because E(Y) = 0]$$

Ans 2 contd.

$$Var(Y) = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n\sigma^2} - 0$$

$$= \frac{n \times \cancel{\sigma^2} Var(X)}{n\sigma^2} - 0$$

$$= 1$$

$\therefore \quad E(Y) = 0$

and $\quad Var(Y) = 1.$

Ans 3. Random sample size → 25.

Values in ascending order →

79, 80, 83, 86, 87, 87, 88, 91, 92, 94, 96, 97, 98, 98,

99, 103, 104, 104, 107, 108, 109, 109, 111, 119, 122.

Sample mean $= \dfrac{\Sigma x}{n} = \dfrac{\Sigma x}{25} = 98.04.$

Sample variance $(S^2) = \dfrac{\Sigma (x_i - \bar{x})^2}{(n-1)} = \cancel{133.465}\ \boxed{133.7067}$

Sample std. dev $(s) = \cancel{\sqrt{133.465}} = \cancel{11.639}\ \cancel{11.329}$

$= \sqrt{133.7067} = 11.563$

Sample range → $[79, 122]$ → 43.

Median $= x_{13} = 98.$

Upper quartile $= \frac{3}{4} \times (n+1)^{th}$ term $= 19.5^{th}$ term

So we avg of 19^{th} & 20^{th} term

Upper quartile $= \dfrac{108+109}{2} = 108.5.$

Similarly Lower quartile $= \dfrac{6^{th} + 7^{th} \text{ term}}{2} = 87.5.$

Ans 3 contd.

95 % confidence intervals.

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

z value = 1.960

s = std dev of samples = 11.563.

n → sample size

$\bar{x}$ → sample mean = 98.04.

$$CI = 98.04 \pm 1.960 \times \frac{11.563}{\sqrt{25}}$$

$$= 98.04 \pm 4.5326$$

$$= (93.507, 102.573)$$

~~= 102.5~~

Ans4. Confidence interval

$$CI = \bar{x} \pm z\left(\frac{s}{\sqrt{n}}\right)$$

$\bar{x} \rightarrow$ sample mean

$z \rightarrow$ confidence level value

$s \rightarrow$ sample std dev

$n \rightarrow$ sample size.

z value for any CI (90%, 95% etc) is fixed.

CASE I. n_1, σ_1

CASE II $n_2 = 2n_1$, $\sigma_2 = 2\sigma_1$

CI in case I $= \bar{x} \pm z \frac{\sigma_1}{\sqrt{n_1}}$

CI in CASE II $= \bar{x} \pm z\left(\frac{2\sigma_1}{\sqrt{2n_1}}\right)$

$= \bar{x} \pm z (\sqrt{2})\left(\frac{\sigma_1}{\sqrt{n_1}}\right)$.

Hence, the range of CI in case II is $\sqrt{2}$ times CASE I.

Thus, we expect the estimate of population to be more accurate in CASE I, i.e. n_1 & σ_1 of population I.

Ans 5. (X_1, X_2) random samples from a population distribution $N(0, \sigma^2)$.

Mean of first order statistic $E(X_{(1)})$ can be found out as given $\Rightarrow$

CDF of $F(\min(X_1, X_2) \le x)$

$$= 1 - F(\min(X_1, X_2) > x)$$

$$= 1 - F(X_1 > x)\, F(X_2 > x)$$

$$= 1 - \frac{1}{2^2}\left(1 - \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right)^2$$

$$\left[\text{where } \operatorname{erf}(x) \quad \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt\right] \quad [\mu = 0]$$

$$\text{PDF } f(x) = \frac{d}{dx}(F) = \frac{-1}{4}\left(\frac{1}{2} - \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2/2\sigma^2}\, dt\right)\frac{2}{\sqrt{\pi}} e^{-x^2/2\sigma^2}$$

$$E(X_{(1)}) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \int \left[\frac{-x}{8}\frac{2}{\sqrt{\pi}} e^{\frac{-x^2}{2\sigma^2}} + \frac{x}{2\sqrt{\pi}}\left(\int_0^x e^{-t^2/2\sigma^2}\, dt\right)\left(\frac{2}{\sqrt{\pi}} e^{-x^2/2\sigma^2}\right)\right] dx$$

$$= \int_{-\infty}^{0}\left[\frac{-1}{4\sqrt{\pi}}\, x\, e^{-\frac{x^2}{2\sigma^2}} + 0\right] dx$$

$$\left[\text{Let } \frac{x^2}{2\sigma^2} \to \alpha \qquad \frac{x\,dx}{\sigma^2} = d\alpha \qquad -\infty < x \le \infty \qquad \alpha \ge 0\right]$$

$$\boxed{E(X_{(1)}) = \frac{-\sigma^2}{4\sqrt{\pi}}}$$

Ans 6. $\bar{x}$ and s^2 are unbiased estimator of mean & variance respectively, as

$E(\bar{x}) = \mu$ where $x = \frac{1}{n}\sum_{i=1}^{n} x_i$

$E(s^2) = \sigma$ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_i)^2$

Sample standard deviation (S) is given as

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

We know that sq. rt. f^n is a concave downward $f(n)$ $y(x) = \sqrt{x}$.

So, by applying Jensen's Inequality, we get that

$E(S) = E$ (std. dev of samples) $< \sqrt{E(\text{var})}$

as $E(S) = E(\sqrt{s^2}) < \left(\sqrt{E(s^2)} = \sigma\right)$

[can be shown by quadratic inequality]

→ By quadratic inequality also, we have

$$\sqrt{\frac{\alpha_1^2 + \alpha_2^2 + \cdots \alpha_n^2}{n}} > \frac{\alpha_1 + \alpha_2 + \alpha_3 \cdots \alpha_n}{n}$$

where $\alpha_i = (X_i - \mu)$

Hence, sample standard deviation is not an unbiased estimator of σ.

Ans 7. Likelihood f^n of $\theta \rightarrow$

$$P_X(\theta) \underset{(X=x)}{} = \begin{cases} \frac{3}{5}\theta & x=0 \\ \frac{2}{5}\theta & x=1 \\ \frac{3}{5}(1-\theta) & x=2 \\ \frac{2}{5}(1-\theta) & x=3 \end{cases}$$

$$L((x_1, x_2 --- x_n); \theta) = P_{x_1 x_2 -- x_n}(x_1, x_2, ---x_n; \theta)$$

$$L(2,3,2,1,0,0,3,2,1,1;\theta) = \left(\frac{3}{5}(1-\theta)\right)^3 \left(\frac{2}{5}(1-\theta)\right)^2 \left(\frac{2}{5}\theta\right)^3 \left(\frac{3}{5}\theta\right)^2$$

$$\text{Likelihood } f^n = \frac{2^5 3^5}{5^{10}} \theta^5 (1-\theta)^5$$

Maximum likelihood ~~estimattion~~ estimation $\rightarrow$

differentiating $L(\theta)$, we get.

$$\frac{dL(\theta)}{d\theta} = \frac{2^5 3^5}{5^{10}} \left(5\theta^4(1-\theta)^5 \cancel{+} -5\theta^5(1-\theta)^4\right)$$

For ~~MF~~ MLE,

$$\cancel{do}\ \frac{d(L(\theta))}{d\theta} = 0.$$

$$\cancel{+}\ \theta = \frac{1}{2}.$$

$$\boxed{\text{Hence } \hat{\theta}_{ML} = \frac{1}{2}.}$$

Ans 8. $f(x|\theta) = \begin{cases} \dfrac{\theta}{(1+x)^{\theta+1}} & 0 < x < \infty \\ 0 & \text{o.w.} \end{cases}$

Likelihood f^n $= \prod_{i=1}^{n} f_X(x_i ; \theta)$

$$= \prod_{i=1}^{n} \frac{\theta}{(1+x_i)^{\theta+1}} = \frac{\theta^n}{(\pi(1+x_i))^{\theta+1}}$$

MLE of θ $\rightarrow$ We take log to simplify calculations.

$$\text{Log } L(\theta) = n \log \theta - (\theta+1) \Sigma \log(1+x_i)$$

differentiating once,

$$\frac{n}{\theta} - \Sigma \log(1+x_i) = 0.$$

$$\therefore \hat{\theta}_{MLE} = \frac{n}{\sum_{i=n}^{n} \log(1+x_i)}$$

Ans 9. $f(x|\theta) = \theta x^{(\theta-1)}$

$0 \leq x \leq 1$
$0 < \theta < \infty$

Likelihood $f^n = \prod f(x_i|\theta)$

$= \theta^n (\prod x_i)^{\theta-1}$.

$\text{Log}(L(\theta)) = n \log \theta + (\theta - 1) \Sigma (\log x_i)$

diff once wrt θ, and equating to zero.

$\frac{n}{\theta} + \Sigma \log(x_i) = 0$

$\therefore \hat{\theta}_{MLE} = \frac{-n}{\Sigma \log(x_i)}$ [-ve sign is justified as $\log x_i < 0$]

$\therefore \hat{\theta}_{MLE} = \frac{-1}{\left(\overline{\log(x_i)}\right)}$

$\text{Var}(\hat{\theta}_{MLE}) = \text{Cov}\left(\frac{-1}{\overline{\log x_i}}, \frac{-1}{\overline{\log x_i}}\right)$

$= E\left(\left(\frac{1}{n}\Sigma\left(\frac{1}{\log x_i} - \frac{1}{\overline{\log x_i}}\right)\right), \left(\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\log x_j} - \frac{1}{\overline{\log x_j}}\right)\right)$

$= \frac{1}{n^2}\sum_{i=1}^{n} E\left(\frac{1}{\log x_i} - \mu\right)^2 = \frac{\sigma^2}{n}$

Ans 9 contd

Here, σ^2 is the variance of the variable

$$\frac{1}{\log X_i}$$

$$Var(\hat{\theta}_{MLE}) = \frac{\sigma^2}{n}$$

As $n \to \infty$

$$Var(\hat{\theta}_{MLE}) \to 0.$$

Ans 10. Bernoulli(θ) R.Vs $\quad$ m observations.

$$\text{Likelihood } f^n = (\theta)^{\Sigma Y_i} (1-\theta)^{m-\Sigma y_i}$$

$$\lambda(y_1, y_2 \dots y_n; \theta) = \frac{(\theta_a)^{\Sigma Y_i}(1-\theta_a)^{m-\Sigma y_i}}{(\theta_b)^{\Sigma y_i}(1-\theta_b)^{m-\Sigma y_i}}$$

$$\theta_a < \theta_0 \quad , \quad \theta_b \in (-\infty, \theta_0] \cup [\theta_0, \infty)$$

Let the critical value be c, such that
H_0 is accepted if $\lambda(\theta) < c$, and rejected otherwise.

$$\text{Log}(\lambda) = (\Sigma y_i)\left(\log(\theta_a) - (\log\theta_b)\right) + (m - \Sigma y_i)\left(\log(1-\theta_a) - \log(1-\theta_B)\right) \le c$$

$$\therefore \Sigma y_i \overbrace{\left(\log\theta_a - \log\theta_b + \log(1-\theta_B) - \log(1-\theta_a)\right)}^{q} + \underbrace{m\left(\log(1-\theta_a) - \log(1-\theta_B)\right)}_{p} < c$$

$$\therefore \Sigma y_i \le \frac{c-p}{q} \quad \text{for } H_0 \text{ to be accepted.}$$

$$\text{Hence } \sum_{i=1}^{m} y_i > \frac{c-p}{q} \quad \text{for } H_0 \text{ to be rejected.}$$

Thus, there exists a b such that the LRT will reject H_0 if

$$\sum_{i=1}^{m} Y_i > b$$

[$\because$ where $b = \dfrac{c-p}{q}$ as calculated above] .