Midsem solutions

Q1. $Y = p N(\mu_1, \sigma_1^2) + (1-p) N(\mu_2, \sigma_2^2)$.

Procedure to draw sample from the distribution of Y:

1. Let $q \sim$ Uniform $[0,1]$
2. If $q < p$, $r = \mu_1 + \sigma_1 z$

   else $r = \mu_2 + \sigma_2 z$

   where $z \sim$ sample from $N(0,1)$

Q2 $Y = 1/X$

$$F_Y(y) = P(Y \le y) = P(X \geq 1/y) = 1 - P(X < \tfrac{1}{y})$$
$$= 1 - F_X(1/y)$$

$$\therefore f_Y(y) = \frac{1}{y^2} f_X(1/y)$$

But $f_X(x) = 1/(b-a)$ if $a < x < b$
$= 0$ otherwise

$$\therefore f_Y(y) = \frac{1}{y^2(b-a)} \text{ if } \frac{1}{b} < y < \frac{1}{a} \quad \text{(PDF)}$$
$= 0$ otherwise

CDF of Y is

$$F_Y(y) = 1 - F_X\left(\frac{1}{y}\right) = 1 - \frac{\frac{1}{y} - a}{b-a}$$

$$= \frac{b - 1/y}{b-a}$$

Mean of Y is $\displaystyle\int_{1/b}^{1/a} \frac{1}{y^2} \frac{y}{b-a}\, dy$

$$= \frac{1}{b-a}\Big(\ln y\Big)_{1/b}^{1/a} = \frac{\ln\left(\frac{1}{a}\right) - \ln\left(\frac{1}{b}\right)}{b-a}$$

$$= E(Y)$$

Median:

$F_Y(y) = 1/2$

$\rightarrow \dfrac{b - 1/y}{b-a} = \dfrac{1}{2}$

$\rightarrow 2b - \dfrac{2}{y} = b - a \quad \therefore \quad b + a = \dfrac{2}{y}$

$\rightarrow y = 2/(a+b)$

$$\mathrm{Var}(Y) = E(Y^2) - (E(Y))^2$$

$$E(Y^2) = \int_{1/b}^{1/a} \cancel{\frac{1}{y^2}} \frac{\cancel{y^2}}{b-a}\, dy$$

$$= \frac{\frac{1}{a} - \frac{1}{b}}{b-a} = \frac{1}{ab}$$

$$\therefore \mathrm{Var}(Y) = \frac{1}{a} - \left(\frac{\ln(1/a) - \ln(1/b)}{b-a} \right)^2$$

Q3 $Y \sim \mathrm{Poiss}(\lambda)$

$$g(y) = \sqrt{y} = g(\lambda) + g'(\lambda)(y-\lambda) + g''(\lambda)\frac{(y-\lambda)^2}{2}$$

$$g'(y) = \frac{1}{2\sqrt{y}} \qquad g''(y) = \frac{1}{2}\left(\frac{-1}{2}\right) y^{-3/2}$$

$$\therefore g'(\lambda) = \frac{1}{2\sqrt{\lambda}} \qquad g''(\lambda) = \frac{-1}{4} \lambda^{-3/2}$$

$$\therefore \sqrt{Y} = \sqrt{\lambda} + \frac{1}{2}\lambda^{-1/2}(Y-\lambda) - \frac{1}{8}\lambda^{-3/2}(Y-\lambda)^2$$

$E(Y-\lambda) = 0 \qquad E\left((Y-\lambda)^2\right) = \lambda$

for Poisson r.v.

$$\therefore E(\sqrt{Y}) = \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}}$$

$$\begin{aligned} Var(\sqrt{Y}) &= E(Y) - \left(E(\sqrt{Y})\right)^2 \\ &= \lambda - \left(\sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}}\right)^2 \\ &= \lambda - \left[\lambda + \frac{1}{64\lambda} - \frac{1}{4}\right] \\ &= \frac{1}{4} - \frac{1}{64\lambda} \\ &\approx 1/4 \end{aligned}$$

Note that third order terms (or higher order terms) in the Taylor series expansion can be ignored for large λ. This is because

$g^{(n)}(\lambda) = O\left(\lambda^{-\frac{(2n-1)}{2}}\right)$ and

$E\left((Y-\lambda)^n\right)$ is $O\left(g^{(n)}(\lambda)\right)$.

Q4 $$\hat{\beta} = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i x_i^2 - n\bar{x}^2}$$

$$E(\hat{\beta}) = \frac{\sum_i (x_i - \bar{x}) E(Y_i)}{\sum_i x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum_i (x_i - \bar{x})(\alpha + \beta x_i)}{\sum_i x_i^2 - n\bar{x}^2}$$

$$= \frac{\beta\left(\sum_i x_i^2 - \bar{x} \sum_i x_i\right)}{\sum_i x_i^2 - \bar{x}^2 n}$$

as $\sum_i (x_i - \bar{x}) = 0$

$$= \beta \frac{\sum_i x_i^2 - n\bar{x}^2}{\sum_i x_i^2 - n\bar{x}^2}$$

$$= \beta.$$

Hence estimate $\hat{\beta}$ is unbiased

$$E(\hat{\alpha}) = \sum_i \frac{E(Y_i)}{n} - \bar{x}\,E(\hat{\beta})$$

$$= \sum_i \frac{\alpha + \beta x_i}{n} - \bar{x}\beta = \alpha$$

$\therefore$ this is unbiased

$$\mathrm{Var}(\hat{\beta}) = \frac{\mathrm{Var}\left(\sum_i (x_i - \bar{x})\, Y_i\right)}{\underbrace{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)^2}_{D}}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2\, \mathrm{Var}(Y_i)}{D}$$ due to independence

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

$$= \sigma^2 \Big/ \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) \quad \text{as} \quad \sum_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$\mathrm{Var}(\hat{\alpha}) = \mathrm{Var}\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \hat{\beta}\bar{x}\right)$$

$$= \mathrm{Var}\left(\sum_{i=1}^{n} \frac{Y_i}{n} - \bar{x} \sum_{i=1}^{n} Y_i \overbrace{\left(\frac{(x_i - \bar{x})}{\sum x_i^2 - n\bar{x}^2}\right)}^{z_i}\right)$$

$$= \sum_i \mathrm{Var}\left[Y_i\left(\frac{1}{n} - \bar{x} z_i\right)\right]$$

$$= \sum_i \sigma^2\left(\frac{1}{n} - \bar{x} z_i\right)^2$$

$$= \sum_i \sigma^2\left(\frac{1}{n^2} + \bar{x}^2 z_i^2 - \frac{2\bar{x} z_i}{n}\right)$$

$\downarrow$ 0 as

$$\sum_i \bar{x}(z_i) \propto \sum_i \bar{x}(x_i - \bar{x})$$

$$= \bar{x}\sum_i x_i - n\bar{x}^2$$

$$= n\bar{x}^2 - n\bar{x}^2 = 0$$

$$= \frac{\sigma^2}{n} + \sigma^2 \sum_i \frac{\bar{x}^2 (x_i - \bar{x})^2}{\sum_j \left(x_i^2 - n\bar{x}^2\right)^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\left(\sum_i x_i^2 - n\bar{x}^2\right)}$$

$$= \cancel{\frac{\sigma^2}{n} + (\bar{x})^2 \frac{\sigma^2}{\sum_i x_i^2 - n\bar{x}^2}}$$

$$\cancel{= \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\left(\sum_i x_i^2 - n\bar{x}^2\right)}}$$

Q5 The key is to realise that S is directly proportional to the sample std. dev.

Here is how

arithmetic mean

$$\sum_{i,j} (x_i - x_j)^2 = \sum_i \sum_j (x_i - m + m - x_j)^2$$

$$= \sum_i \sum_j (x_i - m)^2 + (x_j - m)^2 + 2(x_i - m)(x_j - m)$$

$$= n \sum_i (x_i - m)^2 + n \sum_j (x_j - m)^2 + 0$$

as $\sum_i x_i - m = 0$

$$= 2n \sum_i (x_i - m)^2 = 2n(n-1) \frac{\sum_i (x_i - m)^2}{n-1}$$

$$= 2n(n-1) \times (\text{std. dev.})^2$$

Thus std deviation

$$= \left(\frac{8}{2n(n-1)} \right)^{1/2}$$

<u>Q6</u> $E(F_n(x)) = E\left(\frac{1}{n} \sum_{i=1}^{n} 1(X_i \le x_i) \right)$

$$= \frac{1}{n} \sum_{i=1}^{n} P(X \le x_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} F_X(x) = \frac{1}{n} \times n F_X(x) = F_X(x)$$

$$\mathrm{Var}(F_n(x)) = \mathrm{Var}\left(\frac{1}{n} \sum_i 1(X_i \le x) \right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(1(X_i \le x))$$

this is a Bernoulli r.v. with par $p = F_X(x)$

$$= \frac{1}{n^2} \times n \quad P(X_i \le x)(1 - P(X_i \le x))$$

$$= \frac{F_X(x)(1 - F_X(x))}{n}$$

$$\therefore \lim_{n \to \infty} E\left[(F_n(x) - F_X(x))^2 \right]$$

$$= \lim_{n \to \infty} E\left[(F_n(x) - E(F_n(x)))^2 \right]$$

$$= \lim_{n \to \infty} \operatorname{Var} F_n(x)$$

$$= \lim_{n \to \infty} \frac{F_X(x)\left(1 - F_X(x)\right)}{n} = 0$$