SOLUTIONS TO
MATH10282
INTRO TO STATISTICS

Solutions to Question 1

ILOs addressed: present numerical summaries of a data set.

Suppose that we have the following sample of observations

-1.3, -0.59, 0.1, -1.4, -0.22, -0.35, -0.76, -0.2, 0.41, 0.32

The sample mean is

$$\frac{1}{10}\left(-1.3-0.59+0.1-1.4-0.22-0.35-0.76-0.2+0.41+0.32\right)=-0.399$$

(1 marks)

Level of difficulty = low

UNSEEN

The sample variance is

$$\frac{1}{9}\left[\left(-1.3-\overline{x}\right)^2+\left(-0.59-\overline{x}\right)^2+\cdots+\left(0.32-\overline{x}\right)^2\right]=0.3861211$$

(1 marks)

Level of difficulty = low

UNSEEN

Arrange the data as

-1.40, -1.30, -0.76, -0.59, -0.35, -0.22, -0.20, 0.10, 0.32, 0.41

The middle two numbers are -0.35 and -0.22. The median is their average which is -0.285. (1 marks)

Level of difficulty = low

UNSEEN

Note that $r=2.75$ and $r'=3$, so $Q(1/4)=x_{(2)}+0.75\left(x_{(3)}-x_{(2)}\right)=-0.895$. (1 marks)

Level of difficulty = low

UNSEEN

Note that $r=8.25$ and $r'=8$, so $Q(3/4)=x_{(8)}+0.25\left(x_{(9)}-x_{(8)}\right)=0.155$. (1 marks)

Level of difficulty = low

UNSEEN

The range of the data are

$$0.41 - (-1.4) = 1.81.$$

(1 marks)

Level of difficulty = low

UNSEEN

Note that $r = p(n+1)$ and $r^{'} = [p(n+1)]$ are

$$r = \begin{cases} 3m + \frac{3}{4}, & \text{if } n = 4m, \\ 3m, & \text{if } n = 4m-1, \\ 3m - \frac{3}{4}, & \text{if } n = 4m-2, \\ 3m - \frac{6}{4}, & \text{if } n = 4m-3 \end{cases}$$

and

$$r^{'} = \begin{cases} 3m, & \text{if } n = 4m, \\ 3m, & \text{if } n = 4m-1, \\ 3m-1, & \text{if } n = 4m-2, \\ 3m-2, & \text{if } n = 4m-3, \end{cases}$$

respectively. So,

$$r - r^{'} = \begin{cases} \frac{3}{4}, & \text{if } n = 4m, \\ 0, & \text{if } n = 4m-1, \\ \frac{1}{4}, & \text{if } n = 4m-2, \\ \frac{1}{2}, & \text{if } n = 4m-3. \end{cases}$$

Hence,

$$\text{thirdquartile} = \begin{cases} x_{(3m)} + \frac{3}{4}\left[x_{(3m+1)} - x_{(3m)}\right], & \text{if } n = 4m, \\ x_{(3m)}, & \text{if } n = 4m-1, \\ x_{(3m-1)} + \frac{1}{4}\left[x_{(3m)} - x_{(3m-1)}\right], & \text{if } n = 4m-2, \\ x_{(3m-2)} + \frac{1}{2}\left[x_{(3m-1)} - x_{(3m-2)}\right], & \text{if } n = 4m-3. \end{cases}$$

(4 marks)

Level of difficulty = medium

UNSEEN

Solutions to Question 2

ILOs addressed: define elementary statistical concepts and terminology such as unbiasedness; analyse and compare statistical properties of simple estimators.

Suppose $\widehat{\theta}$ is an estimator of θ.

(i) the bias of $\widehat{\theta}$ is $E(\widehat{\theta}) - \theta$. (1 marks)

Level of difficulty = low

SEEN

(ii) $\widehat{\theta}$ is an unbiased estimator of θ if $E(\widehat{\theta}) = \theta$. (1 marks)

Level of difficulty = low

SEEN

(iii) $\widehat{\theta}$ is an asymptotically unbiased estimator of θ if $\lim_{n\to\infty} E(\widehat{\theta}) = \theta$. (1 marks)

Level of difficulty = low

SEEN

(iv) the mean squared error of $\widehat{\theta}$ is $E(\widehat{\theta} - \theta)^2$. (1 marks)

Level of difficulty = low

SEEN

(v) $\widehat{\theta}$ is a consistent estimator of θ if $\lim_{n\to\infty} E(\widehat{\theta} - \theta)^2 = 0$. (1 marks)

Level of difficulty = low

SEEN

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Exp (λ) distribution. Consider the following estimators for $\theta = 1/\lambda$: $\widehat{\theta}_1 = (1/n)\sum_{i=1}^n X_i$ and $\widehat{\theta}_2 = (1/(n+1))\sum_{i=1}^n X_i$.

(i) The bias of $\widehat{\theta}_1$ is

$$\begin{aligned}
E\left(\widehat{\theta}_1\right) - \theta &= E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \theta \\
&= \frac{1}{n}\sum_{i=1}^n E\left(X_i\right) - \theta \\
&= \frac{1}{n}\sum_{i=1}^n \theta - \theta \\
&= \theta - \theta \\
&= 0.
\end{aligned}$$

The bias of $\widehat{\theta_2}$ is

$$\begin{aligned} E\left(\widehat{\theta_2}\right) - \theta &= E\left(\frac{1}{n+1}\sum_{i=1}^{n} X_i\right) - \theta \\ &= \frac{1}{n+1}\sum_{i=1}^{n} E\left(X_i\right) - \theta \\ &= \frac{1}{n+1}\sum_{i=1}^{n} \theta - \theta \\ &= \frac{n\theta}{n+1} - \theta \\ &= -\frac{\theta}{n+1}. \end{aligned}$$

(2 marks)

Level of difficulty = medium

UNSEEN

(ii) The variance of $\widehat{\theta_1}$ is

$$\begin{aligned} Var\left(\widehat{\theta_1}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} \theta^2 \\ &= \frac{\theta^2}{n}. \end{aligned}$$

The variance of $\widehat{\theta_2}$ is

$$\begin{aligned} Var\left(\widehat{\theta_2}\right) &= Var\left(\frac{1}{n+1}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{(n+1)^2}\sum_{i=1}^{n} Var\left(X_i\right) \\ &= \frac{1}{(n+1)^2}\sum_{i=1}^{n} \theta^2 \\ &= \frac{n\theta^2}{(n+1)^2}. \end{aligned}$$

The mean squared error of $\widehat{\theta_1}$ is

$$MSE\left(\widehat{\theta_1}\right) = \frac{\theta^2}{n}.$$

The mean squared error of $\widehat{\theta}_2$ is

$$MSE\left(\widehat{\theta}_2\right) = \frac{n\theta^2}{(n+1)^2} + \left(\frac{\theta}{n+1}\right)^2 = \frac{\theta^2}{n+1}.$$

(3 marks)

Level of difficulty = medium

UNSEEN

Solutions to Question 3

ILOs addressed: define elementary statistical concepts and terminology such as confidence intervals and hypothesis tests.

(a) Suppose we wish to test $H_0 : \mu = \mu_0$ versus $H_0 : \mu \neq \mu_0$.

(i) the Type I error occurs if H_0 is rejected when in fact $\mu = \mu_0$; (1 marks)

Level of difficulty = low

SEEN

(ii) the Type II error occurs if H_0 is accepted when in fact $\mu \neq \mu_0$; (1 marks)

Level of difficulty = low

SEEN

(iii) the significance level is the probability of type I error. (1 marks)

Level of difficulty = low

SEEN

(b) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, where σ is unknown. The rejection region for the following tests are

(i) reject $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma \neq \sigma_0$ if $(n-1)s^2/\sigma_0^2 > \chi^2_{n-1,1-\alpha/2}$ or $(n-1)s^2/\sigma_0^2 < \chi^2_{n-1,\alpha/2}$; (1 marks)

Level of difficulty = low

SEEN

(ii) reject $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma < \sigma_0$ if $(n-1)s^2/\sigma_0^2 < \chi^2_{n-1,\alpha}$. (1 marks)

Level of difficulty = low

SEEN

(c) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$, where σ is unknown. Then,

(i) the required probability is

$$
\begin{aligned}
& \Pr\left(\text{Reject } H_0 \mid H_1 \text{ is true}\right) \\
= \; & \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,1-\alpha/2} \text{ or } \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,\alpha/2} \mid \sigma \neq \sigma_0\right) \\
= \; & \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,1-\alpha/2} \mid \sigma \neq \sigma_0\right) + \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,\alpha/2} \mid \sigma \neq \sigma_0\right) \\
= \; & 1 - \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha/2} \mid \sigma \neq \sigma_0\right) + \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,\alpha/2} \mid \sigma \neq \sigma_0\right) \\
= \; & 1 - \Pr\left(\frac{(n-1)s^2}{\sigma^2} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,1-\alpha/2} \mid \sigma \neq \sigma_0\right) + \Pr\left(\frac{(n-1)s^2}{\sigma^2} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha/2} \mid \sigma \neq \sigma_0\right) \\
= \; & 1 - \Pr\left(\chi^2_{n-1} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,1-\alpha/2}\right) + \Pr\left(\chi^2_{n-1} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha/2}\right) \\
= \; & 1 - F_{\chi^2_{n-1}}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,1-\alpha/2}\right) + F_{\chi^2_{n-1}}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha/2}\right).
\end{aligned}
$$

(3 marks)

Level of difficulty = medium

UNSEEN

(ii) the required probability is

$$
\begin{aligned}
& \Pr\left(\text{Reject } H_0 \mid H_1 \text{ is true}\right) \\
= \; & \Pr\left(\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,\alpha} \mid \sigma < \sigma_0\right) \\
= \; & \Pr\left(\frac{(n-1)s^2}{\sigma^2} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha} \mid \sigma < \sigma_0\right) \\
= \; & \Pr\left(\chi^2_{n-1} < \frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right) \\
= \; & F_{\chi^2_{n-1}}\left(\frac{\sigma_0^2}{\sigma^2}\chi^2_{n-1,\alpha}\right).
\end{aligned}
$$

Level of difficulty = medium

UNSEEN

Solutions to Question 4

ILOs addressed: define elementary statistical concepts and terminology such as confidence intervals and hypothesis tests; conduct statistical inferences, including confidence intervals and hypothesis tests, in simple one and two-sample situations; sampling distributions.

(a) Let $\mathbf{X} = (X_1, \ldots, X_n)$, with $X_1, \ldots, X_n$ an independent random sample from a distribution F_X with unknown parameter θ. Let $I(\mathbf{X}) = [a(\mathbf{X}), b(\mathbf{X})]$ denote an interval estimator for θ.

(i) $I(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence interval if

$$\Pr\left(a(\mathbf{X}) < \theta < b(\mathbf{X})\right) = 1 - \alpha;$$

(1 marks)

Level of difficulty = low

SEEN

(ii) the coverage probability of $I(\mathbf{X})$ is

$$\Pr\left(a(\mathbf{X}) < \theta < b(\mathbf{X})\right);$$

(1 marks)

Level of difficulty = low

SEEN

(iii) the coverage length of $I(\mathbf{X})$ is $b(\mathbf{X}) - a(\mathbf{X})$. (1 marks)

Level of difficulty = low

SEEN

(b) Suppose $X_1, \ldots, X_n$ are counts of defective items produced on n different days assumed to be a random sample. The number of defectives on any given day is modeled as a Poisson random variable with parameter λ, which is the unknown population mean defectives per day.

(i) Then

$$\begin{aligned}
E\left(\overline{X}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \lambda \\
&= \frac{1}{n} n\lambda \\
&= \lambda
\end{aligned}$$

and

$$\begin{aligned}
Var\left(\overline{X}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \lambda \\
&= \frac{1}{n^2} n\lambda \\
&= \frac{\lambda}{n}.
\end{aligned}$$

(3 marks)

Level of difficulty = medium

UNSEEN

(ii) Since $\left(\overline{X} - \lambda\right) / \left(\sqrt{\lambda/n}\right)$ has the standard normal distribution,

$$\Pr\left(-z_{\alpha/2} < \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} < z_{\alpha/2}\right) = 1 - \alpha$$

which is equivalent to

$$\Pr\left(\frac{\left(\overline{X} - \lambda\right)^2}{\lambda/n} < z_{\alpha/2}^2\right) = 1 - \alpha$$

which is equivalent to

$$\Pr\left(\frac{n\left(\overline{X}^2 + \lambda^2 - 2\overline{X}\lambda\right)}{\lambda} < z_{\alpha/2}^2\right) = 1 - \alpha$$

which is equivalent to

$$\Pr\left(n\left(\overline{X}^2 + \lambda^2 - 2\overline{X}\lambda\right) < z_{\alpha/2}^2 \lambda\right) = 1 - \alpha$$

which is equivalent to

$$\Pr\left(n\overline{X}^2 + n\lambda^2 - \left(2n\overline{X} + z_{\alpha/2}^2\right)\lambda\right) = 1 - \alpha$$

which is equivalent to

$$\Pr\left(\lambda_L < \lambda < \lambda_U\right) = 1 - \alpha,$$

where

$$\lambda_L = \overline{X} + \frac{z_{\alpha/2}^2}{2n} - \sqrt{\frac{\overline{X} z_{\alpha/2}^2}{n} + \frac{z_{\alpha/2}^4}{4n^2}}$$

and

$$\lambda_U = \overline{X} + \frac{z_{\alpha/2}^2}{2n} + \sqrt{\frac{\overline{X} z_{\alpha/2}^2}{n} + \frac{z_{\alpha/2}^4}{4n^2}}.$$

Hence, $[\lambda_L, \lambda_U]$ is a $100(1-\alpha)\%$ confidence interval for λ. (4 marks)

Level of difficulty = medium

UNSEEN

Solutions to Question 5

ILOs addressed: analyse and compare statistical properties of simple estimators.

Let X and Y be uncorrelated random variables. Suppose that X has mean 2θ and variance 4. Suppose that Y has mean θ and variance 2. The parameter θ is unknown.

(i) The biases and mean squared errors of $\widehat{\theta}_1 = (1/4)X + (1/2)Y$ and $\widehat{\theta}_2 = X - Y$ are:

$$
\begin{aligned}
Bias\left(\widehat{\theta}_1\right) &= E\left(\widehat{\theta}_1\right) - \theta \\
&= E\left(\frac{X}{4} + \frac{Y}{2}\right) - \theta \\
&= \frac{E(X)}{4} + \frac{E(Y)}{2} - \theta \\
&= \frac{2\theta}{4} + \frac{\theta}{2} - \theta \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
Bias\left(\widehat{\theta}_2\right) &= E\left(\widehat{\theta}_2\right) - \theta \\
&= E\left(X - Y\right) - \theta \\
&= E(X) - E(Y) - \theta \\
&= 2\theta - \theta - \theta \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
MSE\left(\widehat{\theta}_1\right) &= Var\left(\widehat{\theta}_1\right) \\
&= Var\left(\frac{X}{4} + \frac{Y}{2}\right) \\
&= \frac{Var(X)}{16} + \frac{Var(Y)}{4} \\
&= \frac{4}{16} + \frac{2}{4} \\
&= \frac{3}{4},
\end{aligned}
$$

and

$$
\begin{aligned}
MSE\left(\widehat{\theta}_2\right) &= Var\left(\widehat{\theta}_2\right) \\
&= Var\left(X - Y\right) \\
&= Var(X) + Var(Y) \\
&= 4 + 2 \\
&= 6.
\end{aligned}
$$

(8 marks)

Level of difficulty = medium

UNSEEN

(ii) Both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased. The MSE of $\widehat{\theta}_1$ is smaller than the MSE of $\widehat{\theta}_2$. So, we prefer $\widehat{\theta}_1$. (4 marks)

Level of difficulty = medium

UNSEEN

(iii) The bias of $\widehat{\theta}_c$ is

$$
\begin{aligned}
E\left(\widehat{\theta}_c\right) - \theta &= E\left(\frac{c}{2}X + (1-c)Y\right) - \theta \\
&= \frac{c}{2}E(X) + (1-c)E(Y) - \theta \\
&= \frac{c}{2}2\theta + (1-c)\theta - \theta \\
&= 0,
\end{aligned}
$$

so $\widehat{\theta}_c$ is unbiased.

The variance of $\widehat{\theta}_c$ is

$$
\begin{aligned}
Var\left(\widehat{\theta}_c\right) &= Var\left(\frac{c}{2}X + (1-c)Y\right) \\
&= \frac{c^2}{4}Var(X) + (1-c)^2 Var(Y) \\
&= \frac{c^2}{4}4 + 2(1-c)^2 \\
&= c^2 + 2(1-c)^2.
\end{aligned}
$$

Let $g(c) = c^2 + 2(1-c)^2$. Them $g^{'}(c) = 6c - 4 = 0$ if $c = 2/3$. Also $g^{''}(c) = 6 > 0$. So, $c = 2/3$ minimizes the variance of $\widehat{\theta}_c$. (8 marks)

Level of difficulty = medium

UNSEEN

Solutions to Question 6

ILOs addressed: analyse statistical properties of simple estimators.

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with the common probability mass function

$$p(x) = \theta(1-\theta)^{x-1}$$

for $x = 1, 2, \ldots$ and $0 < \theta < 1$. This probability mass function corresponds to the geometric distribution, so $E(X_i) = 1/\theta$ and $Var(X_i) = (1-\theta)/\theta^2$.

(i) The likelihood function of θ is

$$L(\theta) = \prod_{i=1}^{n} \left\{ \theta(1-\theta)^{X_i - 1} \right\} = \theta^n (1-\theta)^{\sum_{i=1}^{n} X_i - n}.$$

(4 marks)

Level of difficulty = medium

UNSEEN

(ii) The log likelihood function of θ is

$$\log L(\theta) = n \log \theta + \left(\sum_{i=1}^{n} X_i - n \right) \log(1-\theta).$$

The first and second derivatives of this with respect to θ are

$$\frac{d \log L(\theta)}{d\theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} X_i - n}{1-\theta}$$

and

$$\frac{d^2 \log L(\theta)}{d\theta^2} = -\frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} X_i - n}{(1-\theta)^2},$$

respectively. Note that $d \log L(\theta)/d\theta = 0$ if $\theta = n/\sum_{i=1}^{n} X_i$ and that $d^2 \log L(\theta)/d\theta^2 < 0$ for all $0 < \theta < 1$. So, it follows that $\widehat{\theta} = n/\sum_{i=1}^{n} X_i$ is the maximum likelihood estimator of θ. (4 marks)

Level of difficulty = medium

UNSEEN

(iii) By the invariance principle, the maximum likelihood estimator of $\psi = 1/\theta$ is $\widehat{\psi} = (1/n)\sum_{i=1}^{n} X_i$. (4 marks)

Level of difficulty = medium

UNSEEN

(iv) The bias of $\widehat{\psi}$ is

$$\begin{aligned}
E\left(\widehat{\psi}\right) - \psi &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \psi \\
&= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) - \psi \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\theta} - \psi \\
&= \psi - \psi \\
&= 0.
\end{aligned}$$

The variance of $\widehat{\psi}$ is

$$\begin{aligned}
Var\left(\widehat{\psi}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \frac{1-\theta}{\theta^2} \\
&= \frac{1-\theta}{n\theta^2} \\
&= \frac{\psi^2 - \psi}{n}.
\end{aligned}$$

The mean squared error of $\widehat{\psi}$ is

$$MSE\left(\widehat{\psi}\right) = \frac{\psi^2 - \psi}{n}.$$

(4 marks)

Level of difficulty = medium

UNSEEN

(v) The maximum likelihood estimator of ψ is unbiased and consistent. (4 marks)

Level of difficulty = medium

UNSEEN

Solutions to Question 7

ILOs addressed: analyse statistical properties of simple estimators.

Consider the linear regression model with zero intercept:

$$Y_i = \beta X_i + e_i$$

for $i = 1, 2, \ldots, n$, where $e_1, e_2, \ldots, e_n$ are independent and identical normal random variables with zero mean and variance σ^2 assumed known. Moreover, suppose $X_1, X_2, \ldots, X_n$ are known constants.

(i) The likelihood function of β is

$$L(\beta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \prod_{i=1}^{n} \exp\left\{-\frac{e_i^2}{2\sigma^2}\right\} = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\sum_{i=1}^{n} \frac{(Y_i - \beta X_i)^2}{2\sigma^2}\right\}.$$

(4 marks)

Level of difficulty = medium

UNSEEN

(ii) The log likelihood function of β is

$$\log L(\beta) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta X_i)^2.$$

The normal equation is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta X_i) X_i = 0.$$

Solving this equation gives

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

This is an mle since

$$\frac{\partial^2 \log L(\beta)}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 < 0.$$

(4 marks)

Level of difficulty = medium

UNSEEN

(iii) The bias of $\widehat{\beta}$ is

$$\begin{aligned} E\widehat{\beta} - \beta &= E\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} - \beta \\ &= \frac{\sum_{i=1}^n X_i E\left(Y_i\right)}{\sum_{i=1}^n X_i^2} - \beta \\ &= \frac{\sum_{i=1}^n X_i \beta X_i}{\sum_{i=1}^n X_i^2} - \beta \\ &= 0, \end{aligned}$$

so $\widehat{\beta}$ is indeed unbiased. (4 marks)

Level of difficulty = medium

UNSEEN

(iv) The variance of $\widehat{\beta}$ is

$$\begin{aligned} Var\widehat{\beta} &= Var\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n X_i^2 Var\left(Y_i\right)}{\left(\sum_{i=1}^n X_i^2\right)^2} \\ &= \frac{\sum_{i=1}^n X_i^2 \sigma^2}{\left(\sum_{i=1}^n X_i^2\right)^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n X_i^2}. \end{aligned}$$

(4 marks)

Level of difficulty = medium

UNSEEN

(v) The estimator is a linear combination of independent normal random variable, so it is also normal with mean β and variance $\frac{\sigma^2}{\sum_{i=1}^n X_i^2}$. (4 marks)

Level of difficulty = medium

UNSEEN