Probability and Statistics

Homework 11 - Solutions

Q1: Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from the following distribution

$$f_X(x) = \begin{cases} \theta\left(x - \frac{1}{2}\right) + 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta \in [-2, 2]$ is an unknown parameter. We define the estimator $\hat{\Theta}_n$ as

$$\hat{\Theta}_n = 12\overline{X} - 6$$

to estimate θ.

a. Is $\hat{\Theta}_n$ an unbiased estimator of θ?
b. Is $\hat{\Theta}_n$ a consistent estimator of θ?
c. Find the mean squared error (MSE) of $\hat{\Theta}_n$.

Solution:
To answer all three questions, we first need to find the expected value of X, $\mathbb{E}[X]$.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\, dx \\
&= \int_0^1 x\left[\theta\left(x - \frac{1}{2}\right) + 1\right] dx \\
&= \int_0^1 \left(\theta x^2 - \frac{\theta}{2}x + x\right) dx \\
&= \left[\frac{\theta x^3}{3} - \frac{\theta x^2}{4} + \frac{x^2}{2}\right]_0^1 \\
&= \left(\frac{\theta}{3} - \frac{\theta}{4} + \frac{1}{2}\right) - (0) \\
&= \frac{4\theta - 3\theta}{12} + \frac{1}{2} \\
&= \frac{\theta}{12} + \frac{1}{2} = \frac{\theta + 6}{12}
\end{aligned}$$

a. Unbiasedness

An estimator $\hat{\Theta}_n$ is unbiased if $\mathbb{E}[\hat{\Theta}_n] = \theta$. We compute the expected value of the estimator.

$$\begin{aligned}
\mathbb{E}[\hat{\Theta}_n] &= \mathbb{E}[12\overline{X} - 6] \\
&= 12\mathbb{E}[\overline{X}] - 6
\end{aligned}$$

Since $X_1, \ldots, X_n$ are i.i.d., $\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum \mathbb{E}[X_i] = \mathbb{E}[X]$.

$$\begin{aligned}
\mathbb{E}[\hat{\Theta}_n] &= 12\mathbb{E}[X] - 6 \\
&= 12\left(\frac{\theta+6}{12}\right) - 6 \\
&= (\theta+6) - 6 \\
&= \theta
\end{aligned}$$

Since $\mathbb{E}[\hat{\Theta}_n] = \theta$, yes, $\hat{\Theta}_n$ is an unbiased estimator of θ.

b. Consistency

An estimator is consistent if it converges in probability to the true parameter as $n \to \infty$. That is, $\lim_{n\to\infty} P(|\hat{\Theta}_n - \theta| \geq \epsilon) = 0$.

By the Weak Law of Large Numbers (WLLN), the sample mean converges in probability to the population mean:

$$\overline{X} \xrightarrow{p} \mathbb{E}[X]$$

We know $\hat{\Theta}_n = g(\overline{X})$, where $g(t) = 12t - 6$. Since $g(t)$ is a continuous function, we can apply the Continuous Mapping Theorem:

$$\hat{\Theta}_n = g(\overline{X}) \xrightarrow{p} g(\mathbb{E}[X])$$

Substituting the value of $\mathbb{E}[X]$ we found:

$$g(\mathbb{E}[X]) = 12\mathbb{E}[X] - 6 = 12\left(\frac{\theta+6}{12}\right) - 6 = (\theta+6) - 6 = \theta$$

Therefore, $\hat{\Theta}_n \xrightarrow{p} \theta$. Yes, $\hat{\Theta}_n$ is a consistent estimator of θ.

c. Mean Squared Error (MSE)

The Mean Squared Error (MSE) is defined as $\text{MSE}(\hat{\Theta}_n) = \mathbb{E}[(\hat{\Theta}_n - \theta)^2]$. This can be decomposed as:

$$\text{MSE}(\hat{\Theta}_n) = \text{Var}(\hat{\Theta}_n) + (\text{Bias}(\hat{\Theta}_n))^2$$

From part (a), we know the estimator is unbiased, so $\text{Bias}(\hat{\Theta}_n) = \mathbb{E}[\hat{\Theta}_n] - \theta = 0$. Therefore, $\text{MSE}(\hat{\Theta}_n) = \text{Var}(\hat{\Theta}_n)$.

$$\begin{aligned}
\text{Var}(\hat{\Theta}_n) &= \text{Var}(12\overline{X} - 6) \\
&= 12^2\text{Var}(\overline{X}) \\
&= 144\text{Var}\left(\frac{1}{n}\sum X_i\right) \\
&= \frac{144}{n^2}\sum \text{Var}(X_i) \quad \text{(since } X_i \text{ are independent)} \\
&= \frac{144}{n^2}(n \cdot \text{Var}(X)) \quad \text{(since } X_i \text{ are identically distributed)} \\
&= \frac{144}{n}\text{Var}(X)
\end{aligned}$$

Now, we must find $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. First, we find $\mathbb{E}[X^2]$:

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_0^1 x^2 f_X(x)\,dx \\
&= \int_0^1 x^2 \left[\theta\left(x - \frac{1}{2}\right) + 1\right] dx \\
&= \int_0^1 \left(\theta x^3 - \frac{\theta}{2}x^2 + x^2\right) dx \\
&= \left[\frac{\theta x^4}{4} - \frac{\theta x^3}{6} + \frac{x^3}{3}\right]_0^1 \\
&= \frac{\theta}{4} - \frac{\theta}{6} + \frac{1}{3} \\
&= \frac{3\theta - 2\theta}{12} + \frac{4}{12} = \frac{\theta + 4}{12}
\end{aligned}$$

Now, we calculate the variance:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= \left(\frac{\theta + 4}{12}\right) - \left(\frac{\theta + 6}{12}\right)^2 \\
&= \frac{\theta + 4}{12} - \frac{\theta^2 + 12\theta + 36}{144} \\
&= \frac{12(\theta + 4)}{144} - \frac{\theta^2 + 12\theta + 36}{144} \\
&= \frac{(12\theta + 48) - (\theta^2 + 12\theta + 36)}{144} \\
&= \frac{12\theta + 48 - \theta^2 - 12\theta - 36}{144} \\
&= \frac{12 - \theta^2}{144}
\end{aligned}$$

(Note: This variance is non-negative since $\theta \in [-2, 2]$, so $\theta^2 \in [0, 4]$, and $12 - \theta^2 \geq 8$.)

Finally, we substitute this back into our MSE equation:

$$\begin{aligned}
\text{MSE}(\hat{\Theta}_n) &= \frac{144}{n}\text{Var}(X) \\
&= \frac{144}{n}\left(\frac{12 - \theta^2}{144}\right) \\
&= \frac{12 - \theta^2}{n}
\end{aligned}$$

The Mean Squared Error is $\text{MSE}(\hat{\Theta}_n) = \frac{12-\theta^2}{n}$.

Q2: Let $X_1, \ldots, X_4$ be a random sample from a *Geometric(p)* distribution. Suppose we observed $(x_1, x_2, x_3, x_4) = (2, 3, 3, 5)$. Find the likelihood function using

$$P(X_i = x_i; p) = p(1 - p)^{x_i - 1}$$

as the PMF.

The likelihood function, $L(p)$, is the joint probability mass function (PMF) of the observed data, treated as a function of the parameter p.

Since $X_1, \dots, X_4$ is a random sample, the observations are independent and identically distributed (i.i.d.). Therefore, the likelihood function is the product of the individual PMFs:

$$\begin{aligned} L(p) &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4 \mid p) \\ &= \prod_{i=1}^{4} P(X_i = x_i \mid p) \end{aligned}$$

Now, we substitute the given PMF for the Geometric distribution, $P(X_i = x_i; p) = p(1-p)^{x_i - 1}$, into the product:

$$\begin{aligned} L(p) &= \prod_{i=1}^{4} \left[p(1-p)^{x_i-1} \right] \\ &= \left[p(1-p)^{x_1-1} \right] \cdot \left[p(1-p)^{x_2-1} \right] \cdot \left[p(1-p)^{x_3-1} \right] \cdot \left[p(1-p)^{x_4-1} \right] \end{aligned}$$

We can group the p terms and the $(1-p)$ terms together:

$$\begin{aligned} L(p) &= (p \cdot p \cdot p \cdot p) \cdot (1-p)^{(x_1-1)+(x_2-1)+(x_3-1)+(x_4-1)} \\ &= p^4 \cdot (1-p)^{(x_1+x_2+x_3+x_4)-4} \\ &= p^4 \cdot (1-p)^{\sum_{i=1}^{4} x_i - 4} \end{aligned}$$

We are given the observed data $(x_1, x_2, x_3, x_4) = (2, 3, 3, 5)$. We can now plug these values into our expression.

The sample size is $n = 4$.

The sum of the observations is:

$$\sum_{i=1}^{4} x_i = 2 + 3 + 3 + 5 = 13$$

Plugging $n = 4$ and the sum $\sum x_i = 13$ into the likelihood function:

$$\begin{aligned} L(p) &= p^4 (1-p)^{13-4} \\ &= p^4 (1-p)^9 \end{aligned}$$

The likelihood function is $\mathbf{L(p) = p^4(1-p)^9}$, for $p \in (0, 1)$.

Q3: A data scientist is analyzing a large stream of incoming data, $X_1, X_2, \dots, X_n$, where each X_i is an independent draw from a distribution with mean μ and variance $\sigma^2 < \infty$.

Two different estimators are proposed to estimate the mean μ as the sample size n grows:

(a) **Estimator A (The Sample Mean):** $\hat{\mu}_{A,n} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(b) **Estimator B (A "Weighted" Mean):** $\hat{\mu}_{B,n} = \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right) X_i$

Which estimator, if any, is a consistent estimator for μ? Justify your answer by checking if the limit of its Mean Squared Error (MSE) converges to 0.

Solution: We will analyze the consistency of each estimator by calculating the limit of its MSE. Recall that $\text{MSE}(\hat{\Theta}) = \text{Var}(\hat{\Theta}) + (\text{Bias}(\hat{\Theta}))^2$. We also use the facts that $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

1. Analysis of Estimator A (Sample Mean)

This is the standard proof for the sample mean.

Bias:

$$E[\hat{\mu}_{A,n}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\mu) = \mu$$
$$\text{Bias}(\hat{\mu}_{A,n}) = E[\hat{\mu}_{A,n}] - \mu = \mu - \mu = 0 \quad \text{(Unbiased)}$$

Variance:

$$\begin{aligned}\text{Var}(\hat{\mu}_{A,n}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) \quad \text{(due to independence)} \\ &= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}\end{aligned}$$

MSE Limit:

$$\text{MSE}(\hat{\mu}_{A,n}) = \frac{\sigma^2}{n} + (0)^2 = \frac{\sigma^2}{n}$$
$$\lim_{n\to\infty}\text{MSE}(\hat{\mu}_{A,n}) = \lim_{n\to\infty}\frac{\sigma^2}{n} = 0$$

Conclusion for A: Since its MSE converges to 0, Estimator A is a **consistent estimator** of μ.

2. Analysis of Estimator B (Weighted Mean)

Bias:

$$\begin{aligned}E[\hat{\mu}_{B,n}] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right)X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right)E[X_i] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right)\mu = \frac{\mu}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right) \\ &= \frac{\mu}{n}\left(\sum_{i=1}^{n}1 + \sum_{i=1}^{n}\frac{1}{i}\right) = \frac{\mu}{n}(n + H_n) \\ &= \mu\left(1+\frac{H_n}{n}\right)\end{aligned}$$

where $H_n = \sum_{i=1}^{n} \frac{1}{i}$ is the n-th harmonic number.

Bias Limit: The bias is $\text{Bias}(\hat{\mu}_{B,n}) = E[\hat{\mu}_{B,n}] - \mu = \mu\left(\frac{H_n}{n}\right)$. As $n \to \infty$, $H_n \approx \ln(n) + \gamma$.

$$\lim_{n\to\infty} \text{Bias}(\hat{\mu}_{B,n}) = \lim_{n\to\infty} \mu\frac{\ln(n)}{n} = 0 \quad \text{(by L'Hôpital's rule)}$$

This estimator is *asymptotically unbiased*, but it is biased for any finite n.

Variance:

$$\begin{aligned}
\text{Var}(\hat{\mu}_{B,n}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right)X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\left(\left(1+\frac{1}{i}\right)X_i\right) \quad \text{(due to independence)} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(1+\frac{1}{i}\right)^2\text{Var}(X_i) = \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\left(1+\frac{2}{i}+\frac{1}{i^2}\right) \\
&= \frac{\sigma^2}{n^2}\left(\sum_{i=1}^{n}1 + 2\sum_{i=1}^{n}\frac{1}{i} + \sum_{i=1}^{n}\frac{1}{i^2}\right) \\
&= \frac{\sigma^2}{n^2}\left(n + 2H_n + \sum_{i=1}^{n}\frac{1}{i^2}\right)
\end{aligned}$$

MSE Limit: As $n \to \infty$, we know:

- $H_n \approx \ln(n)$
- $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$ (a finite constant)

The MSE is $\text{Var}(\hat{\mu}_{B,n}) + (\text{Bias}(\hat{\mu}_{B,n}))^2$. Let's look at the limit of the variance:

$$\lim_{n\to\infty} \text{Var}(\hat{\mu}_{B,n}) = \lim_{n\to\infty} \frac{\sigma^2}{n^2}\left(n + 2\ln(n) + \frac{\pi^2}{6}\right)$$

Distributing the $\frac{1}{n^2}$:

$$= \lim_{n\to\infty} \sigma^2\left(\frac{1}{n} + \frac{2\ln(n)}{n^2} + \frac{\pi^2}{6n^2}\right)$$

All terms go to 0 as $n \to \infty$. Thus, $\lim_{n\to\infty} \text{Var}(\hat{\mu}_{B,n}) = 0$.

Since $\lim_{n\to\infty} \text{Var}(\hat{\mu}_{B,n}) = 0$ and $\lim_{n\to\infty} \text{Bias}(\hat{\mu}_{B,n}) = 0$, we have:

$$\lim_{n\to\infty} \text{MSE}(\hat{\mu}_{B,n}) = \lim_{n\to\infty} \text{Var}(\hat{\mu}_{B,n}) + \lim_{n\to\infty} (\text{Bias}(\hat{\mu}_{B,n}))^2 = 0 + 0^2 = 0$$

Conclusion for B: Since its MSE also converges to 0, Estimator B is also a **consistent estimator** of μ.

Final Answer

Both Estimator A and Estimator B are consistent estimators for μ, as the MSE of both estimators converges to 0 as $n \to \infty$.

(Note: Although both are consistent, Estimator A converges much faster (MSE $= O(1/n)$) while Estimator B converges much slower (MSE $\approx O(\ln(n)/n^2)$ for the variance and $O((\ln(n)/n)^2)$ for the bias squared). Therefore, in practice, Estimator A is far superior.)

Q4: (For conceptual understanding) Often when working with maximum likelihood functions, out of ease we maximize the log-likelihood rather than the likelihood to find the maximum likelihood estimator. Why is maximizing $L(\mathbf{x};\theta)$ as a function of θ equivalent to maximizing $\log L(\mathbf{x};\theta)$?

Solution:

Maximizing $L(\mathbf{x};\theta)$ is equivalent to maximizing $\log L(\mathbf{x};\theta)$ because the **natural logarithm, $f(x) = \log(x)$, is a strictly increasing function** (also called a monotonically increasing function).

(A function $f(x)$ is strictly increasing if for any two values a and b, if $a > b$, then $f(a) > f(b)$)

Detailed proof:

Assume that maximising the likelihood is not equivalent to maximising the log-likelihood. This would imply that there exist two parameter values θ_1 and θ_2, such that:

$$L(\mathbf{x};\theta_1) > L(\mathbf{x};\theta_2)$$

and

$$\log L(\mathbf{x};\theta_1) \leq \log L(\mathbf{x};\theta_2)$$

But since the logarithm is strictly increasing, this leads to a contradiction. Therefore, our assumption must be false, and maximizing the likelihood is indeed equivalent to maximizing the log-likelihood. i.e $\theta_1 = \theta_2$.

Q5: Let X be one observation from a $N(0,\sigma^2)$ distribution.

a. Find an unbiased estimator of σ^2.

b. Find the log likelihood, $\log(L(x;\sigma^2))$, using

$$f_X(x;\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

as the PDF.

c. Find the Maximum Likelihood Estimate (MLE) for the standard deviation σ, $\hat{\sigma}_{ML}$.

Solution:

a. Find an unbiased estimator of σ^2

An estimator $T(X)$ is unbiased for σ^2 if $E[T(X)] = \sigma^2$.

We are given that $X \sim N(0, \sigma^2)$, which means:

- $E[X] = 0$
- $\mathrm{Var}(X) = \sigma^2$

By the definition of variance:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Substituting the known values:

$$\sigma^2 = E[X^2] - (0)^2$$

$$\sigma^2 = E[X^2]$$

This shows that the expected value of the statistic X^2 is exactly σ^2. Therefore, an unbiased estimator for σ^2 is $T(X) = X^2$.

b. Find the log likelihood, $\log(L(x; \sigma^2))$

Since we have only one observation $X = x$, the likelihood function L is equal to the PDF:

$$L(x; \sigma^2) = f_X(x; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

Let $\ell = \log L$ be the log-likelihood.

$$\ell(x; \sigma^2) = \log\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right)$$

$$\ell(x; \sigma^2) = \log(1) - \log(\sqrt{2\pi}) - \log(\sigma) + \log\left(\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\right)$$

$$\ell(x; \sigma^2) = 0 - \frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{x^2}{2\sigma^2}$$

The question asks for the log-likelihood as a function of σ^2. Let $\theta = \sigma^2$, which means $\sigma = \sqrt{\theta} = \theta^{1/2}$. We substitute this into the expression:

$$\ell(x; \theta) = -\frac{1}{2}\log(2\pi) - \log(\theta^{1/2}) - \frac{x^2}{2\theta}$$

$$\ell(x; \theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2\theta}$$

Substituting σ^2 back in for θ, we get the final answer:

$$\log L(x; \sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{x^2}{2\sigma^2}$$

c. Find the MLE for the standard deviation σ, $\hat{\sigma}_{ML}$

To find the MLE, we maximize the log-likelihood function. It is easier to first find the MLE for the variance, $\theta = \sigma^2$, and then use the invariance property of MLEs.

Let $\ell(\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2\theta}$. We take the derivative with respect to θ and set it to 0.

$$\frac{d\ell}{d\theta} = \frac{d}{d\theta}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2}\theta^{-1}\right]$$

$$\frac{d\ell}{d\theta} = 0 - \frac{1}{2\theta} - \frac{x^2}{2}(-1)\theta^{-2}$$

$$\frac{d\ell}{d\theta} = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$$

Set the derivative to 0 to find the maximum:

$$-\frac{1}{2\theta} + \frac{x^2}{2\theta^2} = 0$$

$$\frac{x^2}{2\theta^2} = \frac{1}{2\theta}$$

$$x^2(2\theta) = 2\theta^2$$

$$2\theta^2 - 2\theta x^2 = 0$$

$$2\theta(\theta - x^2) = 0$$

Since $\theta = \sigma^2$ must be greater than 0, the only valid solution is $\theta = x^2$. So, the MLE for the variance σ^2 is:

$$\hat{\sigma}^2_{ML} = X^2$$

By the **invariance property of MLEs**, if $\hat{\theta}$ is the MLE for θ, then the MLE for $g(\theta)$ is $g(\hat{\theta})$. Here, we want the MLE for the standard deviation σ. The function is $g(\theta) = \sqrt{\theta}$.

$$\hat{\sigma}_{ML} = g(\hat{\sigma}^2_{ML}) = \sqrt{\hat{\sigma}^2_{ML}}$$

$$\hat{\sigma}_{ML} = \sqrt{X^2} = |X|$$

The Maximum Likelihood Estimate for σ is $|X|$.

Note: The invariance property of the Maximum Likelihood Estimator (MLE) states that if θ is the MLE of a parameter θ, then $g(\theta)$ is the MLE of the function $g(\theta)$

Q6: Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where ϵ_i's are independent $N(0, \sigma^2)$ random variables. The estimators are:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}$$

where

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})$$

(a) **Show that $\hat{\beta}_1$ is a normal random variable.**

(b) **Show that $\mathbb{E}[\hat{\beta}_1] = \beta_1$.**

(c) **Show that Var$(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$.**

Solution

(a) **Show that $\hat{\beta}_1$ is a normal random variable.**
First, we simplify S_{xy}. Since $\sum(x_i - \bar{x}) = 0$:

$$S_{xy} = \sum(x_i - \bar{x})(Y_i - \bar{Y}) = \sum(x_i - \bar{x})Y_i - \bar{Y}\sum(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})Y_i$$

Now, substitute this into $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{S_{xx}} = \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S_{xx}}\right)Y_i$$

Let $k_i = (x_i - \bar{x})/S_{xx}$. Since x_i, $\bar{x}$, and S_{xx} are treated as fixed constants, k_i is a constant.

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i Y_i$$

Each Y_i is an independent normal random variable. Since $\hat{\beta}_1$ is a linear combination of independent normal variables, $\hat{\beta}_1$ is also a normal random variable.

(b) **Show that $\mathbb{E}[\hat{\beta}_1] = \beta_1$.**
Using the linear form from part (a) and the linearity of expectation:

$$\mathbb{E}[\hat{\beta}_1] = \mathbb{E}\left[\sum_{i=1}^{n} k_i Y_i\right] = \sum_{i=1}^{n} k_i \mathbb{E}[Y_i]$$

We know $\mathbb{E}[Y_i] = \mathbb{E}[\beta_0 + \beta_1 x_i + \epsilon_i] = \beta_0 + \beta_1 x_i + \mathbb{E}[\epsilon_i] = \beta_0 + \beta_1 x_i$.

$$\mathbb{E}[\hat{\beta}_1] = \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S_{xx}}\right)(\beta_0 + \beta_1 x_i)$$

$$\mathbb{E}[\hat{\beta}_1] = \frac{1}{S_{xx}}\left[\sum(x_i - \bar{x})\beta_0 + \sum(x_i - \bar{x})\beta_1 x_i\right]$$

The first sum is $\beta_0 \sum(x_i - \bar{x}) = 0$. The second sum is $\beta_1 \sum(x_i - \bar{x})x_i = \beta_1 \sum(x_i - \bar{x})^2 = \beta_1 S_{xx}$.

$$\mathbb{E}[\hat{\beta}_1] = \frac{1}{S_{xx}}[0 + \beta_1 S_{xx}] = \beta_1$$

(c) **Show that Var$(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$.**
Using the linear form $\hat{\beta}_1 = \sum k_i Y_i$. Since the Y_i's are independent:

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\sum_{i=1}^{n} k_i Y_i\right) = \sum_{i=1}^{n} k_i^2 \text{Var}(Y_i)$$

We are given $\mathrm{Var}(Y_i) = \sigma^2$.

$$\mathrm{Var}(\hat{\beta}_1) = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right)^2 \sigma^2 = \frac{\sigma^2}{(S_{xx})^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

By definition, $\sum (x_i - \bar{x})^2 = S_{xx}$.

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{(S_{xx})^2} \cdot S_{xx} = \frac{\sigma^2}{S_{xx}}$$

Q7: Again consider the simple linear regression model from Problem 1.

(a) **Show that $\hat{\beta}_0$ is a normal random variable.**
(b) **Show that $\mathbb{E}[\hat{\beta}_0] = \beta_0$.**
(c) **Show that Cov$(\hat{\beta}_1, Y_i) = \frac{x_i - \bar{x}}{S_{xx}} \sigma^2$.**
(d) **Show that Cov$(\hat{\beta}_1, \bar{Y}) = 0$.**
(e) **Show that Var$(\hat{\beta}_0) = \frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}} \sigma^2$.**

Solution

(a) **Show that $\hat{\beta}_0$ is a normal random variable.**
We have $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$. Substitute the linear forms for $\bar{Y}$ and $\hat{\beta}_1$:

$$\hat{\beta}_0 = \left(\frac{1}{n} \sum Y_i \right) - \left(\sum k_i Y_i \right) \bar{x} \quad \text{where } k_i = \frac{x_i - \bar{x}}{S_{xx}}$$

$$\hat{\beta}_0 = \sum \frac{1}{n} Y_i - \sum (k_i \bar{x}) Y_i = \sum_{i=1}^{n} \left(\frac{1}{n} - k_i \bar{x} \right) Y_i$$

$$\hat{\beta}_0 = \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}} \right) Y_i$$

Let $c_i = \left(\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{S_{xx}} \right)$. Since c_i is a constant, $\hat{\beta}_0$ is a linear combination of the independent normal variables Y_i. Therefore, $\hat{\beta}_0$ is also a normal random variable.

(b) **Show that $\mathbb{E}[\hat{\beta}_0] = \beta_0$.**
By linearity of expectation:

$$\mathbb{E}[\hat{\beta}_0] = \mathbb{E}[\bar{Y} - \hat{\beta}_1 \bar{x}] = \mathbb{E}[\bar{Y}] - \mathbb{E}[\hat{\beta}_1] \bar{x}$$

From Problem 1(b), we know $\mathbb{E}[\hat{\beta}_1] = \beta_1$. We find $\mathbb{E}[\bar{Y}]$:

$$\mathbb{E}[\bar{Y}] = \mathbb{E}\left[\frac{1}{n} \sum Y_i \right] = \frac{1}{n} \sum \mathbb{E}[Y_i] = \frac{1}{n} \sum (\beta_0 + \beta_1 x_i)$$

$$\mathbb{E}[\bar{Y}] = \frac{1}{n} (n\beta_0 + \beta_1 \sum x_i) = \beta_0 + \beta_1 \left(\frac{\sum x_i}{n} \right) = \beta_0 + \beta_1 \bar{x}$$

Substitute these results back:

$$\mathbb{E}[\hat{\beta}_0] = (\beta_0 + \beta_1 \bar{x}) - (\beta_1) \bar{x} = \beta_0$$

(c) **Show that Cov($\hat{\beta}_1, Y_i$) = $\frac{x_i - \bar{x}}{S_{xx}}\sigma^2$.**
Using the linear form $\hat{\beta}_1 = \sum_{j=1}^{n} k_j Y_j$ (with index j):

$$\text{Cov}(\hat{\beta}_1, Y_i) = \text{Cov}\left(\sum_{j=1}^{n} k_j Y_j, Y_i\right) = \sum_{j=1}^{n} k_j \text{Cov}(Y_j, Y_i)$$

Since the Y_j's are independent, $\text{Cov}(Y_j, Y_i) = 0$ for $j \neq i$. The only non-zero term is when $j = i$:

$$\text{Cov}(Y_i, Y_i) = \text{Var}(Y_i) = \sigma^2$$

The sum collapses to the $j = i$ term:

$$\text{Cov}(\hat{\beta}_1, Y_i) = k_i \text{Var}(Y_i) = k_i \sigma^2 = \frac{x_i - \bar{x}}{S_{xx}}\sigma^2$$

(d) **Show that Cov($\hat{\beta}_1, \bar{Y}$) = 0.**
Using the result from part (c):

$$\text{Cov}(\hat{\beta}_1, \bar{Y}) = \text{Cov}\left(\hat{\beta}_1, \frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} \text{Cov}(\hat{\beta}_1, Y_i)$$

$$\text{Cov}(\hat{\beta}_1, \bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S_{xx}}\sigma^2\right)$$

Pull the constants out of the sum:

$$\text{Cov}(\hat{\beta}_1, \bar{Y}) = \frac{\sigma^2}{nS_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})$$

Since $\sum(x_i - \bar{x}) = 0$, the entire expression is zero.

$$\text{Cov}(\hat{\beta}_1, \bar{Y}) = 0$$

(e) **Show that Var($\hat{\beta}_0$) = $\frac{\sum_{i=1}^{n} x_i^2}{nS_{xx}}\sigma^2$.**
Use the variance formula for a difference:

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y} - \hat{\beta}_1\bar{x}) = \text{Var}(\bar{Y}) + \text{Var}(\hat{\beta}_1\bar{x}) - 2\text{Cov}(\bar{Y}, \hat{\beta}_1\bar{x})$$

Since $\bar{x}$ is a constant:

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \bar{x}^2\text{Var}(\hat{\beta}_1) - 2\bar{x}\text{Cov}(\bar{Y}, \hat{\beta}_1)$$

From part (d), $\text{Cov}(\bar{Y}, \hat{\beta}_1) = 0$, so the last term vanishes.

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{Y}) + \bar{x}^2\text{Var}(\hat{\beta}_1)$$

We need $\text{Var}(\bar{Y})$ and $\text{Var}(\hat{\beta}_1)$.

- From Problem 1(c): $\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$.
- $\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{n}\sum Y_i\right) = \frac{1}{n^2}\sum \text{Var}(Y_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$.

Substitute these in:

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2 \left(\frac{\sigma^2}{S_{xx}} \right) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

Find a common denominator:

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{S_{xx} + n\bar{x}^2}{nS_{xx}} \right)$$

Recall the identity $S_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$. This implies $S_{xx} + n\bar{x}^2 = \sum x_i^2$. Substitute this into the numerator:

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{\sum_{i=1}^n x_i^2}{nS_{xx}} \right) = \frac{\sum_{i=1}^n x_i^2}{nS_{xx}} \sigma^2$$

Q8: Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample with unknown mean $\mathbb{E}[X_i] = \mu$, and unknown variance $\text{Var}(X_i) = \sigma^2$. Suppose that we would like to estimate $\theta = \mu^2$. We define the estimator $\hat{\Theta}$ as

$$\hat{\Theta} = (\bar{X})^2 = \left[\frac{1}{n} \sum_{k=1}^{n} X_k \right]^2$$

to estimate θ. Is $\hat{\Theta}$ an unbiased estimator of θ? Why?

Solution An estimator is unbiased if $E[\hat{\Theta}] = \theta$. We must calculate $E[\hat{\Theta}] = E[(\bar{X})^2]$.

Recall the relationship for any random variable Y: $E[Y^2] = \text{Var}(Y) + (E[Y])^2$. Let's apply this to $Y = \bar{X}$:

$$E[\hat{\Theta}] = E[(\bar{X})^2] = \text{Var}(\bar{X}) + (E[\bar{X}])^2$$

We need to find the two components:

- $E[\bar{X}] = E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}\sum E[X_i] = \frac{1}{n}\sum \mu = \frac{1}{n}(n\mu) = \mu$. So, $(E[\bar{X}])^2 = \mu^2$.
- $\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum \text{Var}(X_i)$ (since X_i are independent). $\text{Var}(\bar{X}) = \frac{1}{n^2}\sum \sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$.

Substitute these back into the equation:

$$E[\hat{\Theta}] = \frac{\sigma^2}{n} + \mu^2$$

The parameter we are estimating is $\theta = \mu^2$. Since $E[\hat{\Theta}] = \mu^2 + \frac{\sigma^2}{n} \neq \mu^2$ (unless $\sigma^2 = 0$), the estimator is biased.

Answer: No, $\hat{\Theta}$ is **not an unbiased estimator** of θ. Its bias is $E[\hat{\Theta}] - \theta = \frac{\sigma^2}{n}$.

Q9: **Estimating the parameter of a uniform random variable I.** We are given i.i.d. observations $X_1, \ldots, X_n$ that are uniformly distributed over the interval $[0, \theta]$. What is the ML estimator of θ? Is it consistent? Is it unbiased or asymptotically unbiased?

Solution Let $X_{(n)} = \max(X_1, \ldots, X_n)$.

(1) **ML Estimator:** The PDF for a single observation is $f(x_i|\theta) = \frac{1}{\theta}$ for $0 \le x_i \le \theta$, and 0 otherwise. The joint likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \left(\frac{1}{\theta}\right)^n$$

This likelihood is only non-zero if all $x_i \le \theta$. This introduces the constraint that $\theta \ge x_i$ for all i, which means $\theta \ge \max(x_i) = x_{(n)}$.
We want to maximize $L(\theta) = \theta^{-n}$ subject to $\theta \ge x_{(n)}$. Since $L(\theta)$ is a decreasing function of θ, we make it as large as possible by choosing the smallest possible θ that satisfies the constraint. The smallest valid θ is $x_{(n)}$.

Therefore, the ML estimator is $\hat{\theta}_{ML} = X_{(n)}$.

(2) **Consistency: Yes, it is consistent.** An estimator is consistent if it converges in probability to θ. We check if $\lim_{n\to\infty} P(|\hat{\theta}_{ML} - \theta| > \epsilon) = 0$ for any $\epsilon > 0$. Since $\hat{\theta}_{ML} = X_{(n)} \le \theta$, this simplifies to $P(\theta - X_{(n)} > \epsilon) = P(X_{(n)} < \theta - \epsilon)$.

$$P(X_{(n)} < \theta - \epsilon) = P(X_1 < \theta - \epsilon, \dots, X_n < \theta - \epsilon)$$

Since they are i.i.d.:

$$= P(X_1 < \theta - \epsilon)^n = \left(\frac{\theta - \epsilon}{\theta}\right)^n = \left(1 - \frac{\epsilon}{\theta}\right)^n$$

As $n \to \infty$, $\left(1 - \frac{\epsilon}{\theta}\right)^n \to 0$ because $1 - \frac{\epsilon}{\theta} < 1$. Thus, the estimator is consistent.

(3) **Bias: It is biased, but asymptotically unbiased.** We must find $E[\hat{\theta}_{ML}] = E[X_{(n)}]$. Let $Y = X_{(n)}$. The CDF of Y is:

$$F_Y(y) = P(Y \le y) = P(X_1 \le y, \dots, X_n \le y) = P(X_1 \le y)^n = \left(\frac{y}{\theta}\right)^n$$

The PDF of Y is the derivative: $f_Y(y) = F'_Y(y) = \frac{ny^{n-1}}{\theta^n}$ for $0 \le y \le \theta$. The expectation is:

$$E[Y] = \int_0^\theta y f_Y(y)\, dy = \int_0^\theta y \left(\frac{ny^{n-1}}{\theta^n}\right) dy = \frac{n}{\theta^n} \int_0^\theta y^n\, dy$$

$$E[Y] = \frac{n}{\theta^n} \left[\frac{y^{n+1}}{n+1}\right]_0^\theta = \frac{n}{\theta^n} \left(\frac{\theta^{n+1}}{n+1}\right) = \frac{n}{n+1}\theta$$

- **Unbiased?** No. $E[\hat{\theta}_{ML}] = \frac{n}{n+1}\theta \neq \theta$.
- **Asymptotically Unbiased?** Yes. $\lim_{n\to\infty} E[\hat{\theta}_{ML}] = \lim_{n\to\infty} \frac{n}{n+1}\theta = \theta$.

Q10: **Estimating the parameter of a uniform random variable II.** We are given i.i.d. observations $X_1, \dots, X_n$ that are uniformly distributed over the interval $[\theta, \theta + 1]$. Find a ML estimator of θ. Is it consistent? Is it unbiased or asymptotically unbiased?

Solution Let $X_{(1)} = \min(X_1, \dots, X_n)$ and $X_{(n)} = \max(X_1, \dots, X_n)$.

(1) **ML Estimator:** The PDF for a single observation is $f(x_i|\theta) = \frac{1}{(\theta+1)-\theta} = 1$ for $\theta \le x_i \le \theta + 1$. The joint likelihood is $L(\theta) = \prod_{i=1}^n 1 = 1$, but only if all observations satisfy the constraints. The constraints are:

- $x_i \ge \theta$ for all $i \implies \theta \le \min(x_i) \implies \theta \le x_{(1)}$.
- $x_i \le \theta + 1$ for all $i \implies \theta \ge \max(x_i) - 1 \implies \theta \ge x_{(n)} - 1$.

So, $L(\theta) = 1$ if $X_{(n)} - 1 \le \theta \le X_{(1)}$, and $L(\theta) = 0$ otherwise.
The likelihood is 1 for any θ in the interval $[X_{(n)} - 1, X_{(1)}]$. This means the ML estimator is **not unique**. Any value in this interval maximizes the likelihood. We can choose any value, for example an endpoint. Let's analyze $\hat{\theta}_{ML} = X_{(1)}$. (Another valid choice is $\hat{\theta}_{ML} = X_{(n)} - 1$).

(2) **Consistency:** (Analyzing $\hat{\theta} = X_{(1)}$) **Yes, it is consistent.** We check if $\lim_{n\to\infty} P(|X_{(1)} - \theta| > \epsilon) = 0$. Since $X_{(1)} \ge \theta$, this simplifies to $P(X_{(1)} - \theta > \epsilon) = P(X_{(1)} > \theta + \epsilon)$.

$$P(X_{(1)} > \theta + \epsilon) = P(X_1 > \theta + \epsilon, \ldots, X_n > \theta + \epsilon)$$
$$= P(X_1 > \theta + \epsilon)^n$$

The probability for a single X_i is $P(X_i > \theta + \epsilon) = \frac{(\theta+1)-(\theta+\epsilon)}{(\theta+1)-\theta} = \frac{1-\epsilon}{1} = 1 - \epsilon$.

$$P(X_{(1)} > \theta + \epsilon) = (1 - \epsilon)^n$$

As $n \to \infty$, $(1 - \epsilon)^n \to 0$ because $1 - \epsilon < 1$. Thus, the estimator is consistent.

(3) **Bias:** (Analyzing $\hat{\theta} = X_{(1)}$) **It is biased, but asymptotically unbiased.** We need $E[X_{(1)}]$. Let $Z = X_{(1)}$. The CDF of Z is:

$$F_Z(z) = P(Z \le z) = 1 - P(Z > z) = 1 - P(X_1 > z, \ldots, X_n > z)$$
$$F_Z(z) = 1 - P(X_1 > z)^n$$

For $z \in [\theta, \theta + 1]$, $P(X_1 > z) = \frac{(\theta+1)-z}{1} = 1 + \theta - z$.

$$F_Z(z) = 1 - (1 + \theta - z)^n$$

The PDF is $f_Z(z) = F_Z'(z) = 0 - n(1 + \theta - z)^{n-1} \cdot (-1) = n(1 + \theta - z)^{n-1}$. The expectation is:

$$E[Z] = \int_\theta^{\theta+1} z \cdot n(1 + \theta - z)^{n-1}\, dz$$

(Using substitution $u = 1 + \theta - z$, $du = -dz$, $z = 1 + \theta - u$)

$$E[Z] = \int_1^0 (1 + \theta - u) n u^{n-1} (-du) = n \int_0^1 (1 + \theta - u) u^{n-1}\, du$$

$$E[Z] = n \int_0^1 ((1 + \theta) u^{n-1} - u^n)\, du = n \left[(1 + \theta)\frac{u^n}{n} - \frac{u^{n+1}}{n+1} \right]_0^1$$

$$E[Z] = n \left(\frac{1 + \theta}{n} - \frac{1}{n+1} \right) = (1 + \theta) - \frac{n}{n+1} = \theta + 1 - \frac{n}{n+1} = \theta + \frac{1}{n+1}$$

- **Unbiased?** No. $E[X_{(1)}] = \theta + \frac{1}{n+1} \ne \theta$.
- **Asymptotically Unbiased?** Yes. $\lim_{n\to\infty} E[X_{(1)}] = \lim_{n\to\infty} \left(\theta + \frac{1}{n+1}\right) = \theta$.