MH1820 Introduction to Probability and Statistical Methods

Tutorial 10 (Week 11) $\boxed{\text{Solution}}$

Problem 1 (Bias and Standard Error of Parameter Estimators)

Let D_θ, $0 \leq \theta \leq 1$, be the discrete distribution with the following PMF:

x	0	1	2	3
$f(x)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

and $f(x) = 0$ otherwise. Let $X_1, \ldots, X_n$ be an i.i.d random sample drawn from D_θ and let $\overline{X}$ denote the sample mean. We consider the following estimators for θ.

$$\begin{aligned}
\widehat{\theta}_1(n) &= -\frac{1}{2}\overline{X} \\
\widehat{\theta}_2(n) &= \frac{7-(X_1+X_2+X_3)}{6} \\
\widehat{\theta}_3(n) &= \frac{7-3\overline{X}}{6} \\
\widehat{\theta}_4(n) &= \frac{1}{16}\left(17-\frac{3}{n}\sum_{i=1}^{n} X_i^2\right)
\end{aligned}$$

(a) Which of the these estimators are unbiased?

(b) For each of these estimators, compute the standard error.

(c) The following observations for $X_1, \ldots, X_n$ are given (here $n = 10$):

$$3, 0, 2, 1, 3, 2, 1, 0, 2, 1$$

For each *unbiased* estimator from above, substitute the observations into the estimator to obtain an estimation for θ.

(d) If the unknown parameter θ occurs in the formula for the standard error $SE(\widehat{\theta})$, we can replace θ by $\widehat{\theta}$ to get an *estimated* standard error, denoted by $\widehat{SE}(\widehat{\theta})$. For each estimator found in part (c), compute its etimated standard error.

Solution (a) For any $1 \leq i \leq n$,

$$\mathbb{E}(X_i) = \sum_{x=0}^{3} x f(x) = 0\frac{2\theta}{3} + 1\frac{\theta}{3} + 2\frac{2(1-\theta)}{3} + 3\frac{1-\theta}{3} = \frac{7}{3} - 2\theta,$$

$$\mathbb{E}(X_i^2) = \sum_{x=0}^{3} x^2 f(x) = 0^2\frac{2\theta}{3} + 1^2\frac{\theta}{3} + 2^2\frac{2(1-\theta)}{3} + 3^2\frac{1-\theta}{3} = \frac{17-16\theta}{3}.$$

Note that $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$ for any random variable X. We have

$$\mathbb{E}(\hat{\theta_1}(n)) = -\frac{1}{2}\mathbb{E}(\overline{X}) = -\frac{1}{2}(\frac{7}{3} - 2\theta) = \theta - \frac{7}{6},$$

$$\mathbb{E}(\widehat{\theta_2}(n)) = \frac{7}{6} - \frac{1}{6}(\mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3)) = \frac{7}{6} - \frac{1}{2}(\frac{7}{3} - 2\theta) = \theta,$$

$$\mathbb{E}(\widehat{\theta_3}(n)) = \frac{7}{6} - \frac{3}{6}\mathbb{E}(\overline{X}) = \frac{7}{6} - \frac{1}{2}(\frac{7}{3} - 2\theta) = \theta,$$

$$\mathbb{E}(\widehat{\theta_4}(n)) = \frac{1}{16}\left(17 - \frac{3}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2)\right) = \frac{1}{16}\left(17 - 3\frac{17-16\theta}{3}\right) = \theta.$$

The estimators $\widehat{\theta_2}(n), \widehat{\theta_3}(n), \widehat{\theta_4}(n)$ are unbiased.

(b) Note that $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$ and $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ for every random variable X constants a, b. For $1 \le i \le n$, we have

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \frac{17-16\theta}{3} - (\frac{7}{3} - 2\theta)^2 = -4\theta^2 + 4\theta + \frac{2}{9}.$$

Moreover,

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n}\left(-4\theta^2 + 4\theta + \frac{2}{9}\right)$$

and

$$\begin{aligned}
\mathrm{Var}(X_i^2) = \mathbb{E}(X_i^4) - \mathbb{E}(X_i^2)^2 &= \sum_{x=0}^{3} x^4 f(x) - \frac{1}{9}(17-16\theta)^2 \\
&= \frac{113-112\theta}{3} - \frac{(17-16\theta)^2}{9} \\
&= -\frac{256}{9}\theta^2 + \frac{208}{9}\theta + \frac{50}{9}.
\end{aligned}$$

We obtain

$$\mathrm{Var}(\widehat{\theta_1}(n)) = (-1/2)^2\mathrm{Var}(\overline{X}) = \frac{1}{4n}\left(-4\theta^2 + 4\theta + \frac{2}{9}\right) = \frac{1}{n}\left(-\theta^2 + \theta + \frac{1}{18}\right),$$

$$\mathrm{Var}(\widehat{\theta_2}(n)) = (-1/6)^2(\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \mathrm{Var}(X_3)) = -\frac{1}{3}\theta^2 + \frac{1}{3}\theta + \frac{1}{54},$$

$$\mathrm{Var}(\widehat{\theta_3}(n)) = (-3/6)^2\mathrm{Var}(\overline{X}) = \frac{1}{4n}\left(-4\theta^2 + 4\theta + \frac{2}{9}\right) = \frac{1}{n}\left(-\theta^2 + \theta + \frac{1}{18}\right),$$

$$\mathrm{Var}(\widehat{\theta_4}(n)) = \left(\frac{3}{16n}\right)^2\sum_{i=1}^{n}\mathrm{Var}(X_i^2) = \frac{9}{256n}\left(-\frac{256}{9}\theta^2 + \frac{208}{9}\theta + \frac{50}{9}\right) = \frac{1}{n}\left(-\theta^2 + \frac{13}{16}\theta + \frac{25}{128}\right).$$

By definition, we have $SE(\widehat{\theta}(n)) = \sqrt{\mathrm{Var}(\widehat{\theta}(n))}$. Hence

$$
\begin{aligned}
SE(\widehat{\theta_1}(n)) &= \sqrt{\frac{1}{n}\left(-\theta^2+\theta+\frac{1}{18}\right)},\\
SE(\widehat{\theta_2}(n)) &= \sqrt{-\frac{1}{3}\theta^2+\frac{1}{3}\theta+\frac{1}{54}},\\
SE(\widehat{\theta_3}(n)) &= \sqrt{\frac{1}{n}\left(-\theta^2+\theta+\frac{1}{18}\right)},\\
SE(\widehat{\theta_4}(n)) &= \sqrt{\frac{1}{n}\left(-\theta^2+\frac{13}{16}\theta+\frac{25}{128}\right)}.
\end{aligned}
$$

(c) Given the data, the following are estimations of θ and standard errors using unbiased estimators $\widehat{\theta_2}(n), \widehat{\theta_3}(n)$ and $\widehat{\theta_4}(n)$:

- For $\widehat{\theta_2}(n)$: An estimation of θ is

$$\widehat{\theta_2} = (7 - x_1 - x_2 - x_3)/6 = (7-3-0-2)/6 = 1/3.$$

- For $\hat{\theta}_3(n)$: An estimation of θ is

$$\hat{\theta}_3 = (7 - 3\bar{x})/6 = (7 - 3\times 1.5)/6 = 5/12,$$

  here $\bar{x} = 1.5$ is obtained from the given sample.

- For $\hat{\theta}_4(n)$: An estimation for θ is

$$\widehat{\theta_4} = \frac{1}{16}\left(17 - \frac{3}{10}\sum_{i=1}^{10} x_i^2\right) = \frac{71}{160}.$$

(d)

- For $\widehat{\theta_2}(n)$: The estimated standard error is

$$\widehat{SE}(\hat{\theta}_2(n)) = \sqrt{-\frac{1}{3}\hat{\theta}_2^{\,2} + \frac{1}{3}\hat{\theta}_2 + \frac{1}{54}} \approx 0.304.$$

- For $\hat{\theta}_3(n)$: The estimated standard error is

$$\widehat{SE}(\hat{\theta}_3(n)) = \sqrt{\frac{1}{10}\left(-\hat{\theta}_3^{\,2} + \hat{\theta}_3 + \frac{1}{18}\right)} \approx 0.173.$$

- For $\hat{\theta}_4(n)$: The estimated standard error is

$$\widehat{SE}(\hat{\theta}_4(n)) = \sqrt{\frac{1}{10}\left(-\hat{\theta}_4^{\,2} + \frac{13}{16}\hat{\theta}_4 + \frac{25}{128}\right)} \approx 0.19.$$

$\square$

Problem 2 (Bias and Standard Error)

Let X_1, X_2, ..., X_n be i.i.d (random sample) from the exponential distribution whose PDF is $f(x|\theta) = \frac{1}{\theta}e^{-x/\theta}$, where $x > 0$, $\theta > 0$.

(a) Show that $\overline{X}$ is an unbiased estimator of θ.

(b) Show that the variance of $\overline{X}$ is $\dfrac{\theta^2}{n}$.

$\boxed{\textbf{Solution}}$ Recall that if $X \sim Exp(\theta)$, then $\mathbb{E}[X] = \theta$, $\mathrm{Var}[X] = \theta^2$.

(a)

$$\begin{aligned} \mathrm{Bias}(\overline{X}) &= \mathbb{E}[\overline{X}] - \theta = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \theta = \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n} X_i] - \theta \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] - \theta = \frac{1}{n}\sum_{i=1}^{n}\theta - \theta = \frac{1}{n}n\theta - \theta = 0. \end{aligned}$$

(b)

$$\begin{aligned} \mathrm{Var}[\overline{X}] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\theta^2 = \frac{1}{n^2}n\theta^2 = \frac{\theta^2}{n}. \end{aligned}$$

$\square$

Problem 3 (Maximum Likelihood Estimation)

Let $X_1, \ldots, X_n$ be an i.i.d with PDF

$$f(x|\theta) = \frac{\theta}{\sqrt{2\pi}}e^{-\frac{\theta^2 x^2}{2}} \text{ for all } x \in \mathbb{R},$$

where $\theta \in (0, \infty)$ is an unknown parameter. Compute the MLE for θ based on the oberservations

$$x_1 = 1.5, x_2 = 2.2, x_3 = 1.3, x_4 = 3.5, x_5 = 3.3.$$

$\boxed{\textbf{Solution}}$ The likelihood function of θ is

$$L = L(x_1, \ldots, x_5|\theta) = \prod_{i=1}^{5} f(x_i|\theta) = \left(\frac{\theta}{\sqrt{2\pi}}\right)^5 e^{-\frac{\theta^2 \sum_{i=1}^{5} x_i^2}{2}}.$$

Write $C = \left(\frac{1}{\sqrt{2\pi}}\right)^5$. We compute

$$\ln L = \ln L(x_1, \ldots, x_5|\theta) = \ln C + 5\ln\theta - \frac{\theta^2 \sum_{i=1}^5 x_i^2}{2},$$
$$\frac{d}{d\theta}\ln L = \frac{5}{\theta} - \theta\sum_{i=1}^5 x_i^2$$

Hence we need to solve $\frac{5}{\theta} - \theta\sum_{i=1}^5 x_i^2 = 0$. This gives $\frac{5}{\theta} = \theta\sum_{i=1}^5 x_i^2$, $5 = \theta^2\sum_{i=1}^5 x_i^2$, and thus

$$\theta = \sqrt{\frac{5}{\sum_{i=1}^5 x_i^2}}.$$

From the oberservations, we have

$$\sum_{i=1}^5 x_i^2 = 1.5^2 + 2.2^2 + 1.3^2 + 3.5^2 + 3.3^2 = 31.92.$$

Hence the MLE for θ is

$$\theta = \sqrt{\frac{5}{31.92}} \approx 0.396.$$

Problem 4 (Maximum Likelihood Estimation)

Let $X_1, \ldots, X_n$ be an i.i.d from the geometric distribution $Geom(p)$, where $0 < p < 1$ is an unknown parameter. Compute the MLE for p based on the oberservations

$$x_1 = 2, x_2 = 3, x = 4$$

Solution The maximum likelihood function is

$$L = L(x_1, x_2, x_3|p) = (1-p)p(1-p)^2p(1-p)^3p = p^3(1-p)^{1+2+3} = p^3(1-p)^6.$$
$$\ln L = 3\ln p + 6\ln(1-p)$$

Solving $\frac{d}{dp}(\ln L) = 0$, we have

$$\frac{3}{p} - \frac{6}{1-p} = 0$$
$$3(1-p) - 6p = 0$$
$$3 - 9p = 0$$
$$p = \frac{3}{9} = \frac{1}{3}.$$

So the MLE of p is $\frac{1}{3}$.

$\square$

Problem 5 (Maximum Likelihood Estimation and Bias)

Let X_1, ..., X_n be i.i.d from the distribution with PDF $f(x|\theta) = \frac{1}{\theta}x^{(1-\theta)/\theta}$, $0 < x < 1$, $\theta > 0$. Show that the maximum likelihood estimator of θ is

$$-\frac{1}{n}\sum_{i=1}^{n} \ln X_i.$$

$\boxed{\textbf{Solution}}$ The maximum likelihood function is

$$L = L(X_1, \ldots, X_n|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} X_i^{(1-\theta)/\theta} = \frac{1}{\theta^n} \prod_{i=1}^{n} X_i^{(1-\theta)/\theta}.$$

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

Setting $\frac{d}{d\theta} \ln L = 0$, we have

$$\frac{-n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^{n} \ln X_i = 0$$

$$\implies \quad \theta = -\frac{1}{n} \sum_{i=1}^{n} \ln X_i.$$

□

Problem 6 (Confidence Intervals for Normal Distribution)

Suppose the fat content of certain steaks follows a $N(\mu, \sigma^2)$ distribution. The following observations $x_1, \ldots, x_{16}$ for the fat content are given.

$$5.33, 4.25, 3.15, 3.70, 1.61, 6.39, 3.12, 6.59, 3.53, 4.74, 0.11, 1.60, 5.49, 1.72, 4.15, 2.28$$

(i) Suppose $\sigma^2 = 3.2$. Find 90%, 95%, and 99% confidence intervals for μ based on the observations above.

(ii) In part (i), if we want cut down the length of the confidence intervals to half their length, how much would we need to increase sample size?

$\boxed{\textbf{Solution}}$ (i) We have

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

We have

$$\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.$$

Next, we express $-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$ as a condition on the true value μ

$$-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}. \tag{1}$$

For the given sample, we have $n = 16$, $\overline{X} = 3.61$, $\sigma = \sqrt{3.2}$.

1. For 90% confidence interval, $\alpha = 0.1$ and $z_{\alpha/2} = z_{0.05} \approx 1.645$. By (1), a 90% confidence interval for μ is
$$[2.874, 4.346].$$

2. For 95% confidence interval, $\alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} \approx 1.96$. By (1), a 95% confidence interval for μ is
$$[2.733, 4.487].$$

3. For 99% confidence interval, $\alpha = 0.01$ and $z_{\alpha/2} = z_{0.005} \approx 2.576$. By (1), a 99% confidence interval for μ is
$$[2.458, 4.762].$$

(ii) In the part (i), by (1), the length of the confidence interval is $2z_{\alpha/2}\sigma/\sqrt{n}$. Hence, if we want to cut down this length by half, we need a sample which is 4 times as large.

Answer Keys. **Q1(a)** $\widehat{\theta}_2(n)$, $\widehat{\theta}_3(n)$, $\widehat{\theta}_4(n)$ **Q1(c)** $\widehat{\theta}_2 = \frac{1}{3}$, $\widehat{\theta}_3 = \frac{5}{12}$, $\widehat{\theta}_4 = \frac{71}{160}$ **Q1(d)** 0.304, 0.173, 0.19 **Q3** 0.396 **Q4** $\frac{1}{3}$ **Q6** 90% CI: $[2.874, 4.346]$, 95% CI: $[2.733, 4.487]$, 99% CI: $[2.458, 4.762]$