MATH10282 Introduction to Statistics
Semester 2, 2019/2020
Example Sheet 8 - solutions

1. (i) Observe that

$$\begin{aligned}\mathrm{E}(\hat{\mu}_1) &= \mathrm{E}\left(\frac{X_1+X_2+X_3}{3}\right) = \frac{1}{3}\left[\mathrm{E}(X_1)+\mathrm{E}(X_2)+\mathrm{E}(X_3)\right] \\ &= (1/3)(3\mu) = \mu\,.\end{aligned}$$

Hence $\hat{\mu}_1$ is unbiased.

$$\begin{aligned}\mathrm{E}(\hat{\mu}_2) &= \mathrm{E}\left(\frac{X_1+2X_2+X_3}{4}\right) = \frac{1}{4}\left[\mathrm{E}(X_1)+2\,\mathrm{E}(X_2)+\mathrm{E}(X_3)\right] \\ &= (1/4)(4\mu) = \mu\,.\end{aligned}$$

Hence $\hat{\mu}_2$ is unbiased.

(ii) Note that

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_1) &= \mathrm{Var}\left(\frac{X_1+X_2+X_3}{3}\right) \\ &= \frac{1}{9}[\mathrm{Var}\,X_1+\mathrm{Var}\,X_2+\mathrm{Var}\,X_3] \qquad \text{by independence} \\ &= \frac{\sigma^2}{3}\,.\end{aligned}$$

Moreover

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_2) &= \mathrm{Var}\left(\frac{X_1+2X_2+X_3}{4}\right) \\ &= \frac{1}{16}[\mathrm{Var}\,X_1+4\,\mathrm{Var}\,X_2+\mathrm{Var}\,X_3] \qquad \text{by independence} \\ &= \frac{3\sigma^2}{8}\,.\end{aligned}$$

We prefer to use $\hat{\mu}_1$ in practice, as $\mathrm{Var}\,\hat{\mu}_1 < \mathrm{Var}\,\hat{\mu}_2$.

(iii) The sampling distributions are $\hat{\mu}_1 \sim N(\mu, \sigma^2/3)$, $\hat{\mu}_2 \sim N(\mu, 3\sigma^2/8)$.

2. Note that $\bar{X}_1 \sim N(\mu, \sigma^2/n_1)$ and $\bar{X}_2 \sim N(\mu, \sigma^2/n_2)$. For the estimator $\hat{\mu} = a\bar{X}_1 + (1-a)\bar{X}_2$, we have

$$\mathrm{E}(\hat{\mu}) = a\,\mathrm{E}(\bar{X}_1) + (1-a)\,\mathrm{E}(\bar{X}_2) = a\mu + (1-a)\mu = \mu\,.$$

Thus $\mathrm{bias}(\hat{\mu}) = \mathrm{E}(\hat{\mu}) - \mu = \mu - \mu = 0$. For the variance, note

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) &= a^2\,\mathrm{Var}(\bar{X}_1) + (1-a)^2\,\mathrm{Var}(\bar{X}_2) \\ &= \frac{a^2\sigma^2}{n_1} + \frac{(1-a)^2\sigma^2}{n_2}\,.\end{aligned}$$

Differentiating with respect to a and setting equal to zero, we find

$$\frac{d\,\mathrm{Var}(\hat{\mu})}{da} = \frac{2a\sigma^2}{n_1} - \frac{2(1-a)\sigma^2}{n_2} = 0\,.$$

Multiplying both sides by $n_1 n_2$,

$$\begin{aligned} &2a\sigma^2 n_2 - 2(1-a)\sigma^2 n_1 = 0 \\ &a(2n_2\sigma^2 + 2n_1\sigma^2) = 2\sigma^2 n_1 \\ &a = \frac{2\sigma^2 n_1}{2n_2\sigma^2 + 2n_1\sigma^2} = \frac{n_1}{n_1+n_2}\,. \end{aligned}$$

Checking the second derivatives,

$$\frac{d^2\,\mathrm{Var}(\hat{\mu})}{da^2} = \frac{2\sigma^2}{n_1} + \frac{2\sigma^2}{n_2} > 0$$

Therefore $a = n_1/(n_1+n_2)$ minimizes $\mathrm{Var}(\hat{\mu})$.

3. $X_1, \ldots, X_n$ is a random sample from the distribution with p.d.f.

$$f(x) = \begin{cases} e^{-(x-\delta)}\,, & x > \delta \\ 0\,, & \text{otherwise}\,. \end{cases}$$

(i)

$$\begin{aligned} \mathrm{E}(\bar{X}_n) = \mathrm{E}(X_1) &= \int_\delta^\infty x e^{-(x-\delta)} dx \\ &= \big[-xe^{-(x-\delta)}\big]_{x=\delta}^{x=\infty} + \int_\delta^\infty 1 \times e^{-(x-\delta)} dx \\ &\qquad \text{using integration by parts} \\ &= \delta + \big[-e^{-(x-\delta)}\big]_{x=\delta}^{x=\infty} = 1 + \delta \neq \delta \end{aligned}$$

Thus $\bar{X}_n$ is biased for δ.

(ii) $\mathrm{bias}(\bar{X}_n) = 1 + \delta - \delta = 1$. The bias remains constant as $n \to \infty$.

(iii) The alternative estimator $\hat{\delta} = \bar{X}_n - 1$ is unbiased for δ, since

$$\mathrm{E}(\hat{\delta}) = \mathrm{E}(\bar{X}_n - 1) = (1+\delta) - 1 = \delta\,.$$

4. (i)

$$\begin{aligned} L(p) &= \prod_{i=1}^{5} \binom{3}{x_i} p^{x_i} (1-p)^{3-x_i} \\ &= p^{\sum_{i=1}^5 x_i} (1-p)^{15-\sum_{i=1}^5 x_i} \prod_{i=1}^{5} \binom{3}{x_i} \end{aligned}$$

(ii) The given data is $x_1 = 1$, $x_2 = 3$, $x_3 = 2$, $x_4 = 2$, $x_5 = 3$. Hence $\sum_{i=1}^5 x_i = 11$ and $\prod_{i=1}^5 \binom{3}{x_i} = 3 \times 1 \times 3 \times 3 \times 1 = 27$.

p	$L(p)$
0.5	$27(0.5)^{11}(0.5)^4 = 8.24 \times 10^{-4}$
0.6	$27(0.6)^{11}(0.4)^4 = 2.51 \times 10^{-3}$
0.7	$27(0.7)^{11}(0.3)^4 = 4.32 \times 10^{-3}$
0.8	$27(0.8)^{11}(0.2)^4 = 3.71 \times 10^{-3}$

The value of p which maximizes $L(p)$ out of the set $\{0.5, 0.6, 0.7, 0.8\}$ is $p = 0.7$. We would choose this value to be the maximum likelihood estimate of p in this case.

5. (i) The data $x_1, \ldots, x_n$ were obtained by random sampling from a Geom(p) distribution. The probability mass function for this distribution is

$$p(x) = (1-p)^{x-1} p\,, \qquad x = 1, 2, 3, \ldots.$$

Hence, by independence of the observations, the likelihood function is

$$\begin{aligned} L(p) &= \prod_{i=1}^n p(x_i) = \prod_{i=1}^n (1-p)^{x_i - 1} p \\ &= (1-p)^{\sum_{i=1}^n x_i - n} p^n\,. \end{aligned}$$

The log-likelihood is

$$\ell(p) = \left(\sum_{i=1}^n x_i - n \right) \log(1-p) + n \log p\,.$$

(ii) Differentiating the log-likelihood,

$$\begin{aligned} \left.\frac{d\ell}{dp}\right|_{p=\hat{p}} &= -\frac{\sum_{i=1}^n x_i - n}{1-\hat{p}} + \frac{n}{\hat{p}} = 0 \\ \hat{p}\left(\sum_{i=1}^n x_i - n \right) &= n(1-\hat{p}) \\ \hat{p} \sum_{i=1}^n x_i &= n \end{aligned}$$

Hence $\hat{p} = n / \sum_{i=1}^n x_i = 1/\bar{x}$. Checking the second derivatives,

$$\left.\frac{d^2\ell}{dp^2}\right|_{p=\hat{p}} = -\frac{\sum_{i=1}^n x_i - n}{(1-\hat{p})^2} - \frac{n}{\hat{p}^2}\,.$$

As $x_i \geq 1$, for $i = 1, 2, \ldots$, we have that $\sum_{i=1}^n x_i \geq n$ and so the second derivative is negative. Hence $\hat{p} = 1/\bar{x}$ does indeed maximize the likelihood.

6. From the sample, we calculate $\bar{x} = 11.48$. Also, $n = 10$, $\sigma = 0.8$ is known, and the measurements are normally distributed.

(i) For a 95% confidence interval, set $\alpha = 0.05$. In this case, to find $z_{1-\alpha/2} = z_{0.975}$ note that $\Phi(z_{0.975}) = 0.975$, i.e. $z_{0.975}$ is the 0.975 quantile of a $N(0,1)$ distribution. From tables/R, $z_{0.975} = 1.96$. The end points of the 95% confidence interval are

$$\bar{x} \pm \frac{z_{1-\alpha/2}\,\sigma}{\sqrt{n}} = 11.48 \pm 1.96 \times 0.8/\sqrt{10}\,.$$

Hence the 95% confidence interval for the mean PCB level of the fish is $(10.99, 11.98)$.

(ii) For a 99% confidence interval, set $\alpha = 0.01$. Here $z_{1-\alpha/2} = z_{0.995}$ satisfies $\Phi(z_{0.995}) = 0.995$, i.e. $z_{0.995}$ is the 0.995 quantile of a $N(0,1)$ distribution. From tables/R, $z_{0.995} = 2.576$. The end points of the 99% CI are

$$\bar{x} \pm \frac{z_{1-\alpha/2}\,\sigma}{\sqrt{n}} = 11.48 \pm 2.576 \times 0.8/\sqrt{10}\,,$$

i.e. the 99% confidence interval for the mean PCB level of the fish is $(10.83, 12.13)$.

7. (i) For a 95% confidence interval, we have $z_{1-\alpha/2} = z_{0.975} = \Phi^{-1}(0.975) = 1.96$. The end points of the 95% confidence interval are

$$\bar{x} \pm \frac{z_{1-\alpha/2}\,\sigma}{\sqrt{n}} = 100.8 \pm 1.96 \times 3.4/\sqrt{9}\,.$$

Hence a 95% confidence interval for the mean lifetime is (98.58, 103.02).

(ii) For a 98% confidence interval, $z_{1-\alpha/2} = z_{0.99} = \Phi^{-1}(0.99) = 2.3263$. Thus the end points of the 98% CI are

$$100.8 \pm 2.3263 \times 3.4/\sqrt{9}\,.$$

Hence the 98% confidence interval for the mean lifetime is $(98.16, 103.44)$.

Note that as we increase the confidence level, the width of the resulting confidence interval increases.