Department of Mathematics and Statistics

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

# CSE Exercises - Week 7

(1) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with $E(X_1) < \infty$ and $V(X_1) < \infty$. Let $\overline{X}_n$ denote the sample mean. We know, by the WLLN, that $\overline{X}_n$ is asymptotically consistent for $E(X_1)$. Verify this using Proposition 95 (page 199) by showing that

(a) $\overline{X}_n$ is an unbiased estimator for $E(X_1)$; and

(b) $V(\overline{X}_n) \longrightarrow 0$ as $n \longrightarrow \infty$.

(2) Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli$(\theta^*)$ and let $S_n = X_1 + \cdots + X_n$. Consider the following estimator for $\theta^*$:

$$\hat{\theta}_n := (S_n + 1)/(n+2).$$

(a) Is $\hat{\theta}_n$ a biased or unbiased estimator for $\theta^*$? If biased, find the bias.

(b) Show clearly whether $\hat{\theta}_n$ is asymptotically consistent.

(3) Recall from Exercise 1 in week 2's exercises that for $X_1, \ldots, X_n$ IID with $E(X_1) < \infty$ and $V(X_1) < \infty$, the usual sample variance,

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 ,$$

is an unbiased estimator for $V(X_1)$.

In this exercise, we will discover by simulation that the property of unbiasedness is generally not preserved under transformations. Specifically, we will see that $S_n$ is <u>not</u> an unbiased estimator of the population standard deviation, $\sqrt{V(X_1)}$. Do the following:

(i) Generate $n = 10$ Uniform$(0,1)$ random values.

(ii) Compute $S_n^2$ and $S_n$ and store the answers.

(iii) Repeat steps (i) and (ii) 10,000 times.

(iv) Use the 10,000 $S_n^2$ values to obtain a 95% confidence interval for $E(S_n^2)$. Compare this with the Uniform $(0,1)$ population variance and comment on your result.

(v) Use the 10,000 $S_n$ values to obtain a 95% confidence interval for $E(S_n)$. Compare this with the Uniform $(0,1)$ population standard deviation and comment on your result.

4. Let $X$ be a continuous random variable with PDF,

$$f(x;c) = \begin{cases} c\,x^{c-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

where $c > 0$ is a parameter.

(a) Suppose that $X = 0.25$ is observed. Write down the likelihood function and log-likelihood function and plot them.

(b) Suppose that $X_1, X_2 \overset{IID}{\sim} f(x;c)$, and $X_1 = 0.7$ and $X_2 = 0.9$ are observed. Write down the likelihood function and log-likelihood function and plot them.

(c) If $X_1, \dots, X_n \overset{IID}{\sim} f(x;c)$, write down the likelihood function and log-likelihood function.

Solutions

① (a) $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$

$$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$= \frac{1}{n} \cdot n E(X_1) = E(X_1).$$

Hence, $\overline{X}_n$ is unbiased for $E(X_1)$.

(b) $$V(\overline{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$

$$= \frac{1}{n^2} \cdot n V(X_1)$$

$$= \frac{V(X_1)}{n} \longrightarrow 0 \quad \text{as } n \to \infty.$$

② $X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(\theta^*)$, $S_n = X_1 + \cdots + X_n$,

$$\hat{\theta}_n := \frac{S_n + 1}{n+2}$$

(a) Since $\overline{X}_n = S_n / n$,

$$\hat{\theta}_n = \frac{n\overline{X}_n + 1}{n+2}$$

$$E(\hat{\theta}_n) = E\left(\frac{n\overline{X}_n + 1}{n+2}\right)$$

$$= \frac{n\,E(\bar{X}_n) + 1}{n+2}$$

$$= \frac{n\,\theta^* + 1}{n+2} \qquad \left(\text{since } \bar{X}_n \text{ is unbiased for } \theta^*\right)$$

$$\neq \theta^*$$

and so $\hat{\theta}_n$ is a biased estimator for $\theta^*$.

The bias is $E(\hat{\theta}_n) - \theta^*$

$$= \frac{n\,\theta^* + 1}{n+2} - \theta^*$$

$$= \frac{1 - 2\theta^*}{n+2}.$$

(b) Bias of $\hat{\theta}_n = \dfrac{1-2\theta^*}{n+2} \to 0$ as $n \to \infty$.

$$V(\hat{\theta}_n) = V\left(\frac{n\bar{X}_n + 1}{n+2}\right)$$

$$= \left(\frac{n}{n+2}\right)^2 V(\bar{X}_n)$$

$$= \left(\frac{n}{n+2}\right)^2 \frac{\theta^*(1-\theta^*)}{n}$$

$$= \frac{n\,\theta^*(1-\theta^*)}{(n+2)^2} \longrightarrow 0 \quad \text{as } n \to \infty.$$

$\therefore$ $\hat{\theta}_n$ is asymptotically consistent.

3. Results for my generated values are as follows:

(iv) Let $\overline{S_N^2}$ denote the sample mean of my $N = 10{,}000$ $S_n^2$ values (here $n = 10$). Then

$$\overline{S_N^2} = 0.0836$$

$$se_N(\overline{S_N^2}) = \sqrt{\frac{\text{sample variance of } 10{,}000\ S_n^2 \text{ values}}{N}}$$

$$= 0.000264$$

95% confidence interval for $E(\overline{S_n^2})$ is

$$\overline{S_N^2} \pm 1.96\, se_N(\overline{S_N^2})$$

$$= 0.0836 \pm 1.96 \times 0.000264$$

$$= (0.0830,\ 0.0841).$$

The Uniform $(0,1)$ variance is $\frac{1}{12} = 0.0833$, which lies within the confidence interval. Thus, we are 95% confident that $E(S_n^2)$ is equal to the population variance, i.e. that $S_n^2$ is an unbiased estimator for the population variance.

(v) Let $\overline{S_N}$ denote the sample mean of my $N = 10{,}000$ $S_n$ values (here $n = 10$). Then

$$\overline{S_N} = 0.2852$$

$$se_N(\overline{S_N}) = \sqrt{\frac{\text{sample variance of } 10{,}000 \ S_n \text{ values}}{N}}$$

$$= 0.000469$$

95% confidence interval for $E(\overline{S_n})$ is

$$\overline{S_N} \pm 1.96 \, se_N(\overline{S_N})$$

$$= 0.2852 \pm 1.96 \times 0.000469$$

$$= (0.2843 \, , \; 0.2862).$$

The Uniform (0,1) standard deviation is $\frac{1}{\sqrt{12}} = 0.2887$, which lies outside the confidence interval. Thus, we are 95% confident that $E(S_n)$ is <u>not</u> equal to the population standard deviation, i.e. that $S_n$ is a biased estimator for the population standard deviation.

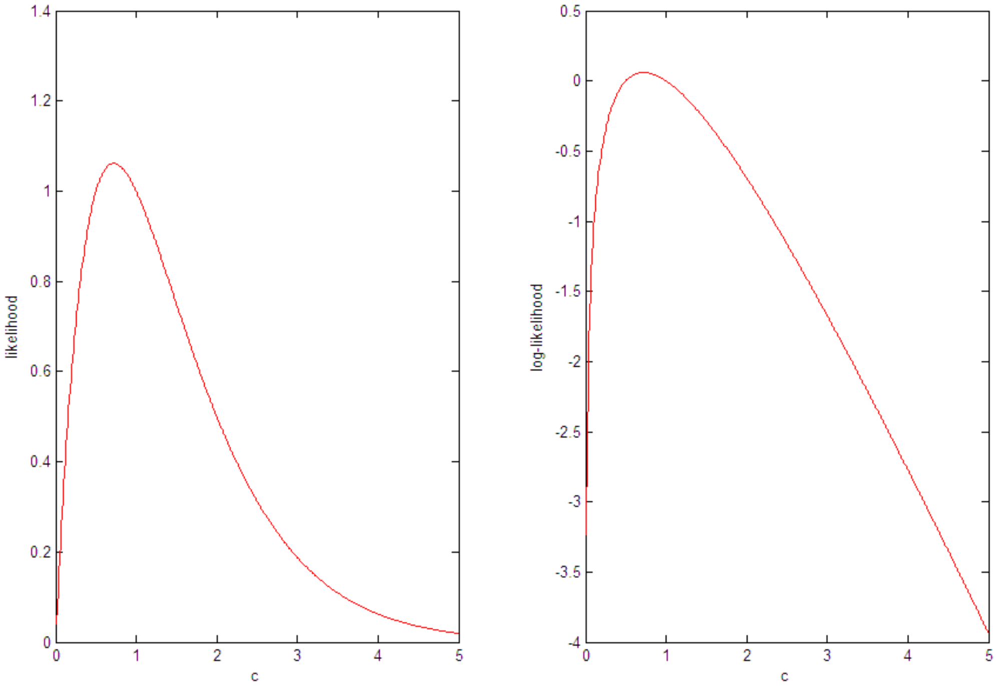
④ $$f(x;c) = \begin{cases} c\,x^{c-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise}. \end{cases}$$

(a) When $X = 0.25$,

$$L(c) = c\,(0.25)^{c-1}, \quad c > 0.$$

$$\ell(c) = \log L(c)$$

$$= \log c + (c-1)\log 0.25, \quad c > 0.$$

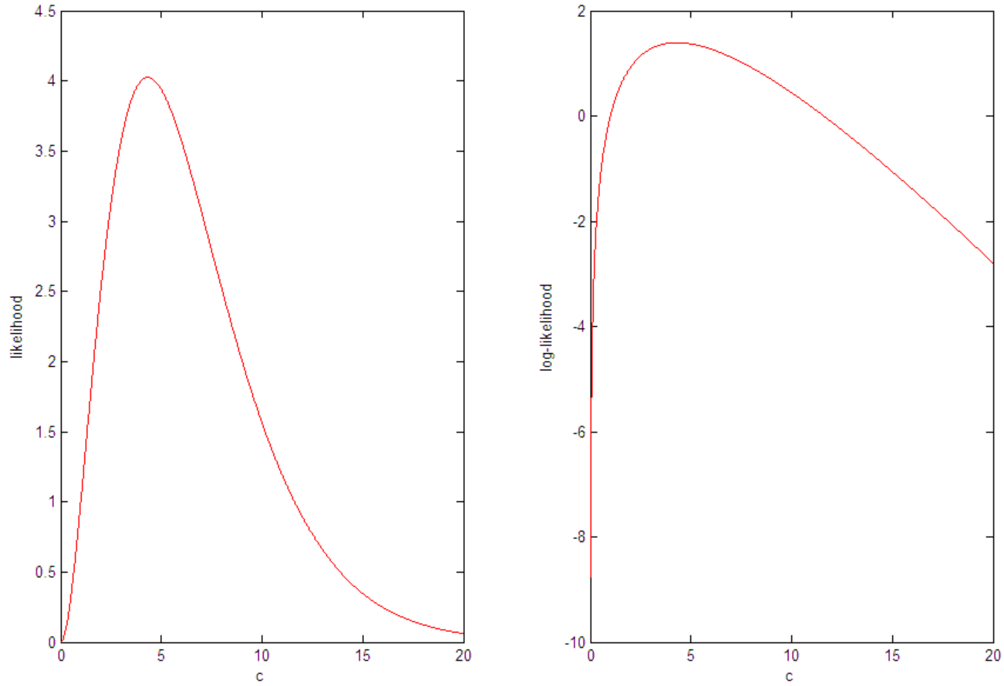


(b) Let $X_1, X_2 \overset{\text{iid}}{\sim} f(x;c)$. Then

$$
\begin{aligned}
L_2(c) &= f(x_1, x_2 ; c) \\
&= f(x_1;c)\, f(x_2;c) \\
&= (c\, x_1^{c-1})(c\, x_2^{c-1}) \\
&= c^2 (x_1 x_2)^{c-1}, \quad c>0. \\
\ell_2(c) &= \log L_2(c) \\
&= 2\log c + (c-1)\log(x_1 x_2), \; c>0.
\end{aligned}
$$

When $X_1 = 0\cdot 7$ and $X_2 = 0\cdot 9$,

$$
\begin{aligned}
L_2(c) &= c^2 (0\cdot 63)^{c-1}, \quad c>0. \\
\ell_2(c) &= 2\log c + (c-1)\log(0\cdot 63), \; c>0.
\end{aligned}
$$




(c) Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim} f(x; c)$. Then

$$
\begin{aligned}
L_n(c) &= f(x_1, \ldots, x_n; c) \\
&= \prod_{i=1}^{n} f(x_i; c) \\
&= c^n \left( \prod_{i=1}^{n} x_i \right)^{c-1}, \quad c > 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\ell_n(c) &= \log L_n(c) \\
&= n \log c + (c-1) \log \left( \prod_{i=1}^{n} x_i \right) \\
&= n \log c + (c-1) \sum_{i=1}^{n} \log x_i, \quad c > 0.
\end{aligned}
$$