Imperial College London

Problem Sheet 2 Solutions

MATH50011
Statistical Modelling 1

Week 2

Lecture 3 (CRLB)

1. In the lecture notes, we sketched the proof of the Cramér-Rao lower bound (CRLB) for continuous random variables. Prove the CRLB for discrete random variables with finite support. (Recall that the *support* of X is the set of values where the pdf/pmf is greater than zero.)

Solution. Without loss of generality, assume X takes values $1, 2, \dots, K$ and let $f_\theta(k)$ denote its pmf. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
Var_\theta(T) I_f(\theta) &= E_\theta[(T - E_\theta T)^2] E_\theta[(\frac{\partial}{\partial\theta} \log f_\theta(X))^2] \\
&\geq \left(E_\theta \left[(T - E_\theta T) \frac{\partial}{\partial\theta} \log f_\theta(X) \right] \right)^2 .
\end{aligned}
$$

As in the lecture notes, the lower bound in the preceding display equals one

$$
\begin{aligned}
E_\theta \left[(T - E_\theta T) \frac{\partial}{\partial\theta} \log f_\theta(X) \right] &= E_\theta \left[(T - E_\theta T) \frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)} \right] \\
&= \sum_{x=1}^{K} (T(x) - E_\theta T)) \frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)} f_\theta(x) \\
&= \sum_{x=1}^{K} T(x) \frac{\partial}{\partial\theta} f_\theta(x) - \sum_{x=1}^{K} E_\theta(T) \frac{\partial}{\partial\theta} f_\theta(x) \\
&= \frac{\partial}{\partial\theta} \sum_{x=1}^{K} T(x) f_\theta(x) - E_\theta(T) \frac{\partial}{\partial\theta} \sum_{x=1}^{K} f_\theta(x) \\
&= \frac{\partial}{\partial\theta} E_\theta(T) - 0 \\
&= \frac{\partial}{\partial\theta} \theta = 1.
\end{aligned}
$$

Note that we do not need to worry about the validity of interchanging a sum with $K < \infty$ terms and differentiation. Thus, $Var_\theta(T) \geq \frac{1}{I_f(\theta)}$. □

2. Find the CRLB for estimating θ based on a random sample of size n from the following distributions

   (a) Exponential(θ);

   (b) Normal(θ, σ^2) with known $\sigma^2 > 0$;

   (c) Bernoulli(θ); (see Example 8)

   (d) Poisson(θ).

Solution. We let f_θ be the pdf for $n = 1$ and $I_n(\theta)$ be the information for general $n \geq 1$.

(a) For the exponential distribution we have

$$\begin{aligned}
f_\theta(x) &= \theta e^{-\theta x} \\
\log f_\theta(x) &= \log \theta - \theta x \\
\frac{\partial}{\partial \theta} \log f_\theta(x) &= \frac{1}{\theta} - x \\
\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) &= -\frac{1}{\theta^2} \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right\} = \frac{n}{\theta^2} \\
CRLB &= \frac{\theta^2}{n}
\end{aligned}$$

(b) For the normal distribution we have

$$\begin{aligned}
f_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\theta)^2)}{2\sigma^2}\right\} \\
\frac{\partial}{\partial \theta} \log f_\theta(x) &= \frac{x-\theta}{\sigma^2} \\
\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) &= \frac{-1}{\sigma^2} \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right\} = \frac{n}{\sigma^2} \\
CRLB &= \frac{\sigma^2}{n}
\end{aligned}$$

(c) For the Bernoulli distribution we have

$$\begin{aligned}
f_\theta(x) &= \theta^x (1-\theta)^{1-x} \\
\frac{\partial}{\partial \theta} \log f_\theta(x) &= \frac{x}{\theta} - \frac{1-x}{1-\theta} \\
\frac{\partial^2}{\partial \theta^2} \log f_\theta(x) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2} \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right\} = \frac{n}{\theta(1-\theta)} \\
CRLB &= \frac{\theta(1-\theta)}{n}
\end{aligned}$$

(d) For the Poisson distribution we have

$$\begin{aligned} f_\theta(x) &= \frac{\theta^x e^{-\theta}}{x!} \\ \frac{\partial}{\partial\theta}\log f_\theta(x) &= \frac{x}{\theta} - 1 \\ \frac{\partial^2}{\partial\theta^2}\log f_\theta(x) &= -\frac{x}{\theta^2} \\ I_n(\theta) &= -E\left\{\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right\} = \frac{n}{\theta} \\ CRLB &= \frac{\theta}{n} \end{aligned}$$

3. For which of the distributions in 2(a-d) can the sample mean be used to construct an unbiased estimator T with variance equal to the CRLB for estimating θ?

Solution. Note that $\bar{X}$ is unbiased for $E_\theta(X) = \theta$ for each distribution in 2(b-d). Moreover, for each distribution in 2(b-d), the CRLB equals $Var(\bar{X})$ where $\bar{X}$ is the sample mean based on a random sample of size n from the given distribution. Hence, $\bar{X}$ itself meets both requirements.

For the Exponential(θ) distribution, $E_\theta(X) = 1/\theta$. However, by Jensen's inequality, we know that $E_\theta(1/\bar{X}) \neq \theta$. We will find a constant a_n to correct for the bias. First, note that $\sum_{i=1}^n X_i \sim Gamma(\theta, n)$.

Noting that $\Gamma(n) = (n-1)\Gamma(n-1)$, we have

$$\begin{aligned} E\left(1/\sum_{i=1}^n X_i\right) &= \int_0^\infty \frac{1}{x}\frac{\theta^n}{\Gamma(n)} x^{n-1}e^{-\theta x}dx \\ &= \int_0^\infty \frac{1}{x}\frac{\theta^{n-1}\theta}{(n-1)\Gamma(n-1)} x^{n-1}e^{-\theta x}dx \\ &= \theta\frac{1}{n-1}\int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} x^{(n-1)-1}e^{-\theta x}dx \\ &= \theta\frac{1}{n-1} \end{aligned}$$

so that $T = (n-1)/\sum_{i=1}^n X_i = (n-1)/(n\bar{X})$ is unbiased. A similar calculation shows that the second moment of $1/\sum_{i=1}^n X_i$ is

$$\frac{\theta^2}{(n-1)(n-2)}$$

so that

$$Var(T) = (n-1)^2\frac{\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{n-2} > CRLB.$$

Hence we cannot use $\bar{X}$ to construct an unbiased estimator that attains the CRLB in this case.

4. Suppose that we wish to estimate θ based on a random sample $X_1, \ldots, X_n$ of Bernoulli(θ) random variables. However, we are only able to obtain a random sample $(Y_i, R_i), \ldots, (Y_n, R_n)$ where the R_i's are iid Bernoulli(p_0) for known p_0, independent of the X_i (updated 21 Jan) and $Y_i = R_i X_i$ for $i = 1, \ldots, n$. Compare the CRLBs for estimating θ based on

(a) The full data distribution of the X_i's;

(b) The marginal distribution of the Y_i's;

(c) The joint distribution of the (Y_i, R_i)'s.

Solution.

(a) The CRLB_X is $\theta(1-\theta)/n$ from the either the notes or 2(c)

(b) Here, $P(Y_i = 1) = P(X_i = 1, R_i = 1) = \theta p_0$ so $Y_i \sim Bernoulli(\theta p_0)$ with p_0 known. We have

$$\begin{aligned}
f_\theta(y) &= [\theta p_0]^y (1-\theta p_0)^{1-y} \\
\frac{\partial}{\partial\theta} \log f_\theta(y) &= \frac{y}{\theta} - p_0 \frac{1-y}{1-\theta p_0} \\
\frac{\partial^2}{\partial\theta^2} \log f_\theta(y) &= -\frac{y}{\theta^2} - p_0^2 \frac{1-y}{(1-\theta p_0)^2} \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial\theta^2} \log f_\theta(Y)\right\} = \frac{np_0}{\theta(1-\theta p_0)} \\
CRLB_Y &= \frac{\theta(1-\theta p_0)}{np_0}
\end{aligned}$$

(c) Some students have difficulties in solving this question. So I wrote all the passages explicitely. Note that the joint distribution has support on the points $(0,0)$, $(0,1)$, and $(1,1)$ since Y_i cannot be 1 unless $R_i = 1$. In particular, we have that

$$\begin{aligned}
f_\theta(y,r) &= P((Y_i, R_i) = (y,r)) = P(Y_i = y, R_i = r) \\
&= P(Y_i = y | R_i = r) P(R_i = r) = P(rX_i = y | R_i = r) P(R_i = r)
\end{aligned}$$

We know that $P(R_i = r) = p_0^r (1-p_0)^{1-r}$. Further, notice that for $r = 1$ we have that

$$P(rX_i = y | R_i = 1) = P(X_i = y | R_i = 1) = P(X_i = y) = \theta^y (1-\theta)^{1-y}$$

and for $r = 0$ (which means that $y = 0$ because we can never have $y \neq 0$ if $r = 0$) we have that

$$P(rX_i = y | R_i = 0) = P(0 = 0 | R_i = 0) = 1.$$

Thus, $P(rX_i = y | R_i = r)$ is equal to $\theta^y (1-\theta)^{1-y}$ when $r = 1$, and it is equal to 1 when $r = 0$. This means that $P(rX_i = y | R_i = r) = \{\theta^y (1-\theta)^{1-y}\}^r$.
Hence, we have

$$\begin{aligned}
f_\theta(y,r) &= p_0^r (1-p_0)^{1-r} \{\theta^y (1-\theta)^{1-y}\}^r \\
\frac{\partial}{\partial\theta} \log f_\theta(y,r) &= r\left[\frac{y}{\theta} - \frac{1-y}{1-\theta}\right] \\
\frac{\partial^2}{\partial\theta^2} \log f_\theta(y,r) &= -r\left[\frac{y}{\theta^2} + \frac{1-y}{(1-\theta)^2}\right] \\
I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial\theta^2} \log f_\theta(Y,R)\right\} = \frac{np_0}{\theta(1-\theta)} \\
CRLB_{Y,R} &= \frac{\theta(1-\theta)}{np_0}
\end{aligned}$$

This is an example where the responses X_i are *missing completely at random*. We see that

$$\frac{\theta(1-\theta)}{n} \leq \frac{\theta(1-\theta)}{np_0} \leq \frac{\theta(1-\theta p_0)}{np_0}$$

so $CRLB_X \leq CRLB_{Y,R} \leq CRLB_Y$. In particular, the best (lowest) possible variance for an unbiased estimator of θ arises when we observe the X_i directly.

Unless $p_0 = 1$ (so the X_i are observed with probability 1), we lose information for estimating θ when they data are generated this way.

Fascinatingly, we can attain (in theory) better precision by using the joint distribution of the observable (Y_i, R_i) than we can by using the marginal distribution of the Y_i even though we already know the distribution of R_i exactly.

5. Show that an asymptotically unbiased estimator sequence need not be consistent. (Hint: consider estimating μ based on a sequence of independent rv's $X_i \sim N(\mu, 2i)$ for $i = 1, 2, 3, \dots$)

Solution. Since $E\bar{X} = \mu$, it is unbiased, hence asymptotically unbiased. $Var(\bar{X}) = 2\sum_i \frac{i}{n^2} = \frac{n+1}{n}$. Hence,

$$\bar{X} \sim N(\mu, \frac{n+1}{n})$$

Fix $\delta > 0$. Notice that if $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(|X-\theta| \geq \delta) &= P(X-\theta > \delta) + P(X-\theta \leq -\delta) \\ &= 1 - \Phi\left(\frac{\delta+\theta-\mu}{\sigma}\right) + \Phi\left(\frac{-\delta+\theta-\mu}{\sigma}\right) \end{aligned} \tag{1}$$

In general if $\theta = \mu$,

$$P(|X-\theta| \geq \delta) = 2\left(1 - \Phi\left(\frac{\delta}{\sigma}\right)\right) \tag{2}$$

Using (2) we get,

$$P(|\bar{X}-\mu| > \delta) = 2P(\bar{X}-\mu > \delta) = 2\left(1 - \Phi\left(\frac{\delta}{\sqrt{(n+1)/n}}\right)\right) \to 2(1-\Phi(\delta)) \neq 0.$$

Therefore $\bar{X}$ is not a consistent estimator of μ.

6. Show that a consistent estimator sequence T_n need not be asymptotically unbiased. (Hint: consider a sequence (T_n, Y_n) with $Y_n \sim$ Bernoulli(1/n) and $T_n|Y_n = 0 \sim N(\theta, \sigma^2/n)$ and $T_n|Y_n = 1 \sim N(n^2, 1)$.)

Solution. We will use the notation

$$\begin{aligned} &\textbf{if } Y_n = 0,\ T_n = Z_n \sim N(\theta, \frac{\sigma^2}{n}) \\ &\textbf{if } Y_n = 1,\ T_n = R_n \sim N(n^2, 1) \end{aligned}$$

Where we have $Y_n \sim Bernoulli(\frac{1}{n})$. We will show consistency using the definition of convergence in probability. For any $\delta > 0$,

$$\begin{aligned} P(|T_n-\theta| \geq \delta) &= P(|T_n-\theta| \geq \delta, Y_n = 0) + P(|T_n-\theta| \geq \delta, Y_n = 1) \\ =& P(|T_n-\theta| \geq \delta | Y_n = 0)P(Y_n = 0) + P(|T_n-\theta| \geq \delta | Y_n = 1)P(Y_n = 1) \\ =& P(|Z_n-\theta| \geq \delta)\left(1-\frac{1}{n}\right) + P(|R_n-\theta| \geq \delta)\frac{1}{n} \\ =& 2\left(1-\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)\right)\left(1-\frac{1}{n}\right) + \left(1-\Phi(\delta+\theta-n^2) + \Phi(-\delta+\theta-n^2)\right)\frac{1}{n} \\ & (\text{ Using } (2) \text{ and } (1)) \\ & \to 2(0)(1) + (1+0+0)0 = 0 \end{aligned}$$

Hence, T_n is consistent for θ. However, we see that

$$E(T_n) = E\{E(T_n|Y_n = 0)P(Y_n = 0) + E(T_n|Y_n = 1)P(Y_n = 1)\} = \theta(1 - \frac{1}{n}) + \frac{n^2}{n} \to \infty \qquad (3)$$

Hence, T_n is not asymptotically unbiased.

7. Let $X_1, X_2, \ldots$ be iid Uniform$(0, \theta)$ random variables and define $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$.

(a) Show that $\hat{\theta}_n$ is asymptotically unbiased and consistent.

(b) Find a sequence of constants a_n such that $a_n\hat{\theta}_n$ is unbiased and consistent.

(c) Compare the MSE of $\hat{\theta}_n$ and $a_n\hat{\theta}_n$.

Solution.

(a) Let $0 < \epsilon < 1$. We can show convergence in probability directly. First, note that

$$P(\hat{\theta}_n \leq \theta - \epsilon) = P(X_1, \ldots, X_n \leq \theta - \epsilon) = \left(\frac{\theta - \epsilon}{\theta}\right)^n = \left(1 - \frac{\epsilon}{\theta}\right)^n \to 0$$

Moreover, $P(\hat{\theta}_n \geq \theta + \epsilon) = 0$. Therefore,

$$P(|\hat{\theta}_n - \theta| \geq \epsilon) \to 0.$$

hence $\hat{\theta}_n$ is consistent for θ.
To show asymptotic unbiasedness, consider $0 \leq x \leq \theta$,

$$P(\hat{\theta}_n \leq x) = \left(\frac{x}{\theta}\right)^n$$

So that the pdf of $\hat{\theta}_n$ is $f_{\hat{\theta}_n}(x) = n\frac{x^{n-1}}{\theta^n}$. Then, we see

$$\begin{aligned} E\hat{\theta}_n &= \frac{n}{\theta^n}\int_0^\theta x x^{n-1} dx \\ &= \frac{n}{n+1}\theta. \end{aligned}$$

Clearly, $E\hat{\theta}_n \to \theta$ as $n \to \infty$. Hence, $\hat{\theta}_n$ is asymptotically unbiased.
As an alternative to direct proof of convergence in probability, we can show that $Var(\hat{\theta}_n) \to 0$. We have

$$E(\hat{\theta}_n^2) = \frac{n}{\theta^n}\int_0^\theta x^{n+1} dx = \frac{n}{n+2}\theta^2$$

and thus $Var(\hat{\theta}_n) = \frac{n}{n+2}\theta^2 - \left[\frac{n}{n+1}\theta\right]^2 = \theta^2\frac{n}{(n+2)(n+1)^2} \to 0$. Hence, $\hat{\theta}_n$ is consistent.

(b) From above, we see immediately that

$$E\frac{n+1}{n}\hat{\theta}_n = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta.$$

Hence, $a_n = \dfrac{n+1}{n}$. $a_n\hat{\theta}_n$ is asymptotically unbiased since it is an unbiased estimator of θ.

$$Var(a_n\hat{\theta}_n) = \frac{(n+1)^2}{n^2}\theta^2\frac{n}{(n+2)(n+1)^2} = \frac{1}{n(n+2)}\theta^2 \to 0$$

Therefore $\hat{\theta}_n$ and $a_n\hat{\theta}_n$ are consistent estimators of θ.

(c) Recall that $MSE(T) = Var(T) + bias(T)^2$. Using this, we find

$$\begin{aligned} MSE(\hat{\theta}_n) &= Var(\hat{\theta}_n) + bias(\hat{\theta}_n)^2 \\ &= \theta^2\frac{n}{(n+2)(n+1)^2} + \left(\frac{n}{n+1}\theta - \theta\right)^2 \\ &= \theta^2\frac{n}{(n+2)(n+1)^2} + \theta^2\left(\frac{n}{n+1} - 1\right)^2 \\ &= \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

and for the unbiased estimator $a_n\hat{\theta}_n$, we have

$$\begin{aligned} MSE(a_n\hat{\theta}_n) &= a_n^2 Var(\hat{\theta}) - 0 \\ &= \frac{1}{n(n+2)}\theta^2. \end{aligned}$$

We can compare the estimators using the ratio $MSE(\hat{\theta}_n)/MSE(a_n\hat{\theta}_n) = 2n/(n+1) > 1$ for $n > 1$. Hence the MSE for the unbiased $a_n\hat{\theta}_n$ is lower than for $\hat{\theta}_n$.

8. Let $X_1, X_2, \ldots$ be iid Bernoulli(θ) random variables and consider estimating $g(\theta) = \text{Var}(X_1) = \theta(1-\theta)$. Define the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$.

(a) Show that $T_n = \bar{X}_n(1-\bar{X}_n)$ is asymptotically unbiased and consistent.

(b) Find a sequence of constants a_n such that a_nT_n is unbiased and consistent.

(c) Compare the MSE of T_n and a_nT_n.

Note: This is intended to be a challenging problem and serves to highlight the utility of several asymptotic results we will see soon relative to the direct approach. You may simplify variance calculations significantly using the fact that

$$Var(S_n^2) = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$$

where $\sigma^2 = Var(X_i)$ and $\mu_4 = E\{(X_i - \mu)^4\}$. Direct calculations can also be avoided by appealing to Slutsky's lemma or the continuous mapping theorem (which may have been seen before this term).

Solution.

(a) Since the data are binary, the parametric estimator can be expressedly equivalently as

$$\begin{aligned} T_n(\vec{X}_n) &= \bar{X}_n(1-\bar{X}_n) \\ &= \bar{X}_n - (\bar{X}_n)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i - (\bar{X}_n)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 \\ &= \hat{\sigma}_n^2, \end{aligned}$$

which was defined previously. We know that the bias is

$$-\sigma^2/n = -g(\theta)/n \to 0$$

as $n \to \infty$, so T_n is asymptotically unbiased for $g(\theta) = \theta(1-\theta)$.

An application of Slutsky's lemma allows us to conclude that since $\bar{X}_n \to_p \theta$ and $1-\bar{X}_n \to_p 1-\theta$ that $T_n \to_p g(\theta)$.

A direct proof using the lemma on MSE and consistency requires showing $Var(T_n) \to 0$. A rather involved calculation leads to

$$Var(T_n) = \frac{n-1}{n^3}[(n-1)g(\theta)(1-3g(\theta)) - (n-3)g(\theta)^2] \to 0.$$

(b) Using part (a) of this problem, $a_n = n/(n-1)$. Hence,

$$a_n T_n = \frac{n}{n-1}\bar{X}_n(1-\bar{X}_n) = s_n^2.$$

Since $a_n \to 1$, a further application of Slutsky's lemma allows us to conclude that $a_n T_n$ is consistent for $g(\theta)$.

A direct proof using the lemma on MSE and consistency requires showing $Var(a_n T_n) \to 0$. A rather involved calculation leads to

$$Var(a_n T_n) = \frac{1}{n(n-1)}[(n-1)g(\theta)(1-3g(\theta)) - (n-3)g(\theta)^2] \to 0.$$

(c) This part requires direct comparison of the MSE, so no shortcut is readily available. We have that $MSE(a_n T_n) = Var(a_n T_n) + 0$ and

$$MSE(T_n) = \frac{(n-1)^2}{n^2}Var(a_n T_n) + \frac{g(\theta)^2}{n}$$

For sufficiently large n, the difference in MSE will be dominated by the squared bias term $g(\theta)^2/n$ for T_n. This term is maximized when $\theta = 1/2$ so that this term is $1/(16n)$.

9. In R, the code below implements the simulation study for $n = 10$ and $\epsilon = 0.1$.

```
set.seed(50011)
result <- logical(length = 1000)
n <- 10
epsilon <- .1
for(i in 1:1000){
   X <- rnorm(n, mean = 0)
   m <- median(X)
   result[i] <- abs(m - 0) < epsilon
}
mean(result)
```

Note that the command `set.seed(50011)` ensures that you obtain the same results each time you run this set of commands.

Type the above commands into your R console (or write a script) and then:

(a) Explore how the value of `mean(result)` changes by increasing the value of `n` in this code to, e.g. $n = 30, 50, 100, 200, 500, 1000$.

(b) Referring to the results of your experimentation, comment on whether the sample median appears to be consistent for μ in this setting.

Solution. The purpose of this problem is to obtain a better understanding of convergence in probability and consistency.

Running the above code (including setting the random seed) for each suggested value of n results in the following table

n	10	30	50	100	200	500	1000
$\hat{P}(\lvert m_n - 0\rvert < 0.1)$	0.238	0.340	0.467	0.613	0.764	0.930	0.985

where $\hat{P}(|m_n - 0| < 0.1))$ is the value of `mean(result)`.

Since $\hat{P}(|m_n - 0| < 0.1)) \to 1$ based on the results of this experiment, it suggests the sample median converges to μ in probability (is consistent). To make this argument more compelling, we could repeat the experiment with smaller values of ϵ and different values of μ in the normal distribution.