ECE531 Homework Assignment Number 8

Solution

Make sure your reasoning and work are clear to receive full credit for each problem.

1. 4 points. The equality condition on the Cauchy-Schwarz inequality implies that the information bound is attained if and only if

$$\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) \quad = \quad k(\theta)\left[\hat{\theta}(y) - \mathrm{E}_\theta\left\{\hat{\theta}(Y)\right\}\right]$$

almost surely for some $k(\theta)$. Assuming all of the regularity conditions hold, show that

$$k(\theta) \quad = \quad \frac{I(\theta)}{\frac{\partial}{\partial\theta}\mathrm{E}_\theta\left\{\hat{\theta}(Y)\right\}}.$$

What is $k(\theta)$ when we have an unbiased estimator?

Solution: From slide 12 of the lecture notes, when Cauchy-Schwarz inequality becomes equality we have

$$\begin{aligned}
\left[\frac{\partial}{\partial\theta}\mathrm{E}_\theta\{\hat{\theta}(Y)\}\right]^2 &= \mathrm{E}_\theta\left\{\left[\hat{\theta}(Y) - \mathrm{E}_\theta\{\hat{\theta}(Y)\right]^2\right\}\cdot I(\theta) \\
&= \mathrm{E}_\theta\left\{\left[\frac{\frac{\partial}{\partial\theta}\ln p_Y(Y;\theta)}{k(\theta)}\right]^2\right\}\cdot I(\theta) \\
&= \frac{I(\theta)}{k^2(\theta)}\cdot I(\theta)
\end{aligned}$$

Hence

$$k(\theta) = \pm\frac{I(\theta)}{\frac{\partial}{\partial\theta}\mathrm{E}_\theta\{\hat{\theta}(Y)\}}.$$

is a constant (in y) that achieves the information bound. To show that the negative option can be eliminated, consider a scalar observation drawn from the family of densities

$$p_Y(y\,;\,\theta) \quad = \quad \frac{1}{\sqrt{2\pi}}\exp\left\{\frac{(y-\theta)^2}{2}\right\}.$$

We know that the MVU estimator in this case is simply $\hat{\theta}(y) = y$ and that this MVU estimator achieves the CRLB of $1/I(\theta) = 1$. It is not difficult to compute $\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) = y - \theta$. Hence, we have

$$\begin{aligned}
\frac{\partial}{\partial\theta}\ln p_Y(y\,;\,\theta) &= k(\theta)\left[\hat{\theta}(y) - \mathrm{E}_\theta\left\{\hat{\theta}(Y)\right\}\right] \\
y - \theta &= k(\theta)(y-\theta)
\end{aligned}$$

which has only one solution: $k(\theta) = 1$. Hence, in general, the unique $k(\theta)$ that achieves the bound is

$$k(\theta) = \frac{I(\theta)}{\frac{\partial}{\partial\theta}\mathrm{E}_\theta\{\hat{\theta}(Y)\}}.$$

When we have an unbiased estimator, $\mathrm{E}_\theta\{\hat{\theta}(Y)\} = \theta$ and $k(\theta) = I(\theta)$.

2. 4 points. Poor textbook Chapter IV, Problem 13 (c).

   Solution:

$$\frac{\partial^2}{\partial\theta^2}\log p_Y(Y\,;\,\theta) = -\frac{T(Y)}{\theta^2} - \frac{n - T(Y)}{(1-\theta)^2},$$

   from which

$$I(\theta) = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}.$$

   The CRLB for an unbiased estimator is thus

$$\mathrm{CRLB} = \frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{n}$$

3. 4 points. Poor textbook Chapter IV, Problem 20 (b). Hint: For a Gaussian random variable Y, we know $\frac{\mathrm{E}\{Y^4\}}{\mathrm{E}^2\{Y^2\}} = 3$.

   Solution: First compute the Fisher information:

$$I(\theta) = -\mathrm{E}_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log p_Y(y\,;\,\theta)\right\} = \sum_{k=1}^{n}\left\{\frac{s_k^4\mathrm{E}_\theta\{Y_k^2\}}{(1+\theta s_k^2)^3} - \frac{s_k^4}{2(1+\theta_k^2)^2}\right\}$$

$$= \frac{1}{2}\sum_{k=1}^{n}\frac{s_k^4}{(1+\theta s_k^2)^2}.$$

   So the CRLB is

$$\frac{2}{\sum_{k=1}^{n}\frac{s_k^4}{(1+\theta s_k^2)^2}}.$$

4. 4 points. Poor textbook Chapter IV, Problem 21 (a), (b), (c), and (e).

   Solution:

   (a) First we write out the pdf's of our two observations:

$$p_{Y_i}(y_i\,;\,\lambda) \quad = \quad \begin{cases} \frac{\lambda^{y_i}e^{-\lambda}}{y_i!} & y_i \in \{0,1,2,3,\dots\} \\ 0 & \text{else} \end{cases} \qquad \text{for } i = 1,2$$

   The two observations are independent and thus we can write the pdf of the vector observation y as

$$p_Y(y\,;\,\lambda) \quad = \quad p(y_1\,;\,\lambda)p(y_2\,;\,\lambda) \quad = \quad \frac{\lambda^{(y_1+y_2)}e^{-2\lambda}}{y_1!y_2!}\mathbb{I}_{\{y_1\geq 0,\ y_2\geq 0\}}$$

Using the factorization for an exponential family, let $\theta' = \log \lambda$

$$\begin{aligned}
p_Y(y\,;\,\lambda) &= e^{\theta'(y_1+y_2)}\frac{e^{-2e^{\theta'}}}{y_1!y_2!} \\
\Rightarrow T(y) &= y_1 + y_2 \\
h(y) &= \frac{1}{y_1!y_2!} I_{\{y_1 \geq 0,\ y_2 \geq 0\}} \\
C(\theta') &= e^{-2e^{\theta'}}
\end{aligned}$$

From the completeness theorem for exponential families, $T(y) = y_1 + y_2$ is a complete sufficient statistic for θ'. Since θ and θ' are in one-to-one correspondence, complete+sufficient for θ' implies complete+sufficient for θ.

(b) Recall that we want to estimate $\theta = e^{-\lambda}$.

$$\begin{aligned}
\mathbf{E}_\theta\left\{\hat{\theta}(Y)\right\} &= \frac{1}{2}\mathbf{E}_\theta\left\{f(Y_1) + f(Y_2)\right\} \\
&= \frac{1}{2}\sum_{y_1=0}^{\infty} f(y_1)\frac{\lambda^{y_1}e^{-\lambda}}{y_1!} + \frac{1}{2}\sum_{y_2=0}^{\infty} f(y_2)\frac{\lambda^{y_2}e^{-\lambda}}{y_2!} \\
&= \frac{1}{2}\frac{\lambda^0 e^{-\lambda}}{0!} + \frac{1}{2}\frac{\lambda^0 e^{-\lambda}}{0!} \\
&= \frac{1}{2}e^{-\lambda} + \frac{1}{2}e^{-\lambda} \;=\; \theta
\end{aligned}$$

Thus the estimator is **unbiased**.

(c) Using the RBLS theorem,

$$\tilde{g}(y) = \mathbf{E}_\theta\left\{\frac{1}{2}\left[f(Y_1) + f(Y_2)\right]\Big| T(Y) = y_1 + y_2\right\}$$

In order to determine this MVUE we need the conditional pdfs $p_{Y_i}(y_i\,;\,\lambda\,|\,T = t)$ for $i = 1, 2$. Using Bayes rule,

$$\begin{aligned}
p_{Y_1}(y_1\,;\,\lambda\,|\,T = t) &= \frac{q(t\,;\,\lambda\,|\,y_1)p_{Y_1}(y_1\,;\,\lambda)}{q(t\,;\,\lambda)} = \frac{\Pr_\lambda\{Y_2 = t - y_1\}p_{Y_1}(y_1\,;\,\lambda)}{q(t\,;\,\lambda)} \\
&= \frac{\lambda^{(t-y_1)}e^{-\lambda}}{(t-y_1)!} \times \frac{\lambda^{(y_1)}e^{-\lambda}}{(y_1)!} \times \frac{t!}{(2\lambda)^t e^{-2\lambda}} \\
&= \left(\frac{1}{2}\right)^t \frac{t!}{(t-y_1)!\,y_1!} \\
\Rightarrow\ \mathbf{E}_\lambda\left\{f(Y_1)|t\right\} &= \sum_{y_1=0}^{\infty} f(y_1)p(y_1\,;\,\lambda\,|\,t) = \left(\frac{1}{2}\right)^t \frac{t!}{(t-0)!\,0!} = \left(\frac{1}{2}\right)^t \\
\mathbf{E}_\lambda\left\{f(Y_2)|t\right\} &= \left(\frac{1}{2}\right)^t \text{ by same method.} \\
\Rightarrow\ \tilde{g}(T(y)) &= \left(\frac{1}{2}\right)^{(y_1+y_2)}
\end{aligned}$$

By the RBLS theorem we know that $\tilde{g}$ is unbiased, and since $\tilde{g}$ is a function of a complete sufficient statistic we know that it must be MVUE. Verifying the unbiasedness (for fun):

$$\begin{aligned}\mathbf{E}_\lambda\left\{\tilde{g}(T)\right\} &= \sum_{t=0}^{\infty}\frac{(2\lambda)^t e^{-2\lambda}}{t!}\left(\frac{1}{2}\right)^t \\ &= e^{-2\lambda}\sum_{t=0}^{\infty}\frac{\lambda^t}{t!} = e^{-2\lambda}e^{\lambda} = e^{-\lambda} = \theta.\end{aligned}$$

(e) To determine the CRLB, we first calculate Fisher's information:

$$\begin{aligned}I(\theta) &= \mathbf{E}_\theta\left\{-\frac{\partial^2}{\partial\theta^2}\log p(Y\,;\,\theta)\right\} \\ &= \mathbf{E}_\theta\left\{\frac{1}{\theta^2}\left[\frac{Y_1+Y_2}{\log\theta}+2+\frac{Y_1+Y_2}{(\log\theta)^2}\right]\right\}\end{aligned}$$

$$\mathbf{E}_\theta\left\{Y_1+Y_2\right\} = 2\lambda = -2\log\theta$$

$$\Rightarrow\ I(\theta) = -\frac{2}{\theta^2\log\theta}$$

Finally, the CRLB for the variance of unbiased estimators of θ is simply $1/I_\theta$.

5. 4 points. Suppose we have a polynomial fitting problem where the observed samples are given as

$$Y_k = \sum_{\ell=0}^{L-1}\theta_\ell k^\ell + W_k \text{ for } k = 0, 1, \ldots, n-1$$

where $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$. Find the Fisher information matrix for estimating $\{\theta_0,\ldots,\theta_{L-1}\}$. If you have difficulty with the general case, try $L = 2$ (fitting a straight line) first.

Solution: First consider the $L = 2$ case. The 2×2 Fisher information matrix is

$$I(\theta) = \begin{bmatrix} -\mathbf{E}\left[\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_0^2}\right] & -\mathbf{E}\left[\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_0\partial\theta_1}\right] \\ -\mathbf{E}\left[\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_1\partial\theta_0}\right] & -\mathbf{E}\left[\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_1^2}\right]\end{bmatrix}$$

where $p(y;\ \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{1}{2\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta_0-\theta_1 k)^2}$. Hence

$$\frac{\partial\ln p(y;\ \theta)}{\partial\theta_0} = \frac{1}{\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta_0-\theta_1 k)$$

$$\frac{\partial\ln p(y;\ \theta)}{\partial\theta_1} = \frac{1}{\sigma^2}\sum_{k=0}^{n-1}(y_k-\theta_0-\theta_1 k)k$$

$$\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_0^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_0\partial\theta_1} = -\frac{1}{\sigma^2}\sum_{k=0}^{n-1}k$$

$$\frac{\partial^2\ln p(y;\ \theta)}{\partial\theta_1^2} = -\frac{1}{\sigma^2}\sum_{k=0}^{n-1}k^2$$

Accordingly, we have the Fisher information matrix as

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} n & \sum_{k=0}^{n-1} k \\ \sum_{k=0}^{n-1} k & \sum_{k=0}^{n-1} k^2 \end{bmatrix}$$

from which we can draw a few bits of insight:

- Estimating the slope of a line is more difficult than estimating its intercept, since $I_{22}(\theta) > I_{11}(\theta)$. One source of intuition for this is that it takes at least two samples to estimate the slope, whereas it takes only one sample to estimate the intercept.
- An unknown intercept leads to an increase in the minimum variance of the slope, and vice-versa. This is intuitive since it would be easier to estimate the intercept or the slope if the other were known.

The same approach can be applied to the case when $L > 2$. You can get the Fisher information matrix for general L as

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} n & \dots & \sum_{k=0}^{n-1} k^{L-1} \\ \vdots & \ddots & \vdots \\ \sum_{k=0}^{n-1} k^{L-1} & \dots & \sum_{k=0}^{n-1} k^{2(L-1)} \end{bmatrix}.$$

Once again, it is clear that the higher order coefficients are more difficult to estimate since their minimum variances are larger than the lower order coefficients.