MA 413 - Statistics for Data Science

Solutions to Exercise 8

1. The parameter θ has been replaced by $1/\theta$. So, $f_Y(y;\theta) = \frac{1}{\theta}e^{-\frac{y}{\theta}}\mathbf{1}_{y\geq 0}$, we calculate

$$\begin{aligned} I(\theta) &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f_Y(Y;\theta)\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\left(-\log\theta - \frac{y}{\theta}\right)\right] \\ &= \frac{1}{\theta^2} \end{aligned}$$

So, the Cramer-Rao bound is as in: $\operatorname{Var}(\hat{\theta}) \geq \frac{1}{n}\theta^2$.

2. The question has been changed to consider $Y_1, \ldots, Y_n \sim$ Uniform $[0,\theta]$. Then the sample mean is $\bar{Y} = \frac{1}{n}\sum_{j=1}^n Y_j$. So

$$\begin{aligned} \mathbb{E}[\bar{Y}] &= \frac{1}{n}\sum_{j=1}^{n}\frac{\theta}{2} = \frac{\theta}{2} \\ \operatorname{Var}(\bar{Y}) &= \frac{1}{n^2}\cdot n\operatorname{Var}(Y_j) = \frac{\theta^2}{12n} \end{aligned}$$

For the median we shall assume that $n = 2m+1$. The density of the median $\tilde{X}$ is given by

$$f_{\tilde{X}}(x;\theta) = (m+1)\begin{pmatrix} 2m+1 \\ m \end{pmatrix}\frac{1}{\theta}\left[\frac{x}{\theta}\right]^m\left[1-\frac{x}{\theta}\right]^m\mathbf{1}_{\{0\leq x\leq\theta\}}$$

Calculating the expectation and variance,

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\theta x f_{\tilde{X}}(x;\theta)dx \\ &= (m+1)\begin{pmatrix} 2m+1 \\ m \end{pmatrix}\cdot\theta\cdot\int_0^1 u^{m+1}(1-u)^m du \\ &= (m+1)\begin{pmatrix} 2m+1 \\ m \end{pmatrix}\cdot\theta\cdot\frac{(m+1)!m!}{(2m+2)!} = \frac{\theta}{2} \\ \mathbb{E}\left[\tilde{X}^2\right] &= \int_0^\theta x^2 f_{\tilde{X}}(x;\theta)dx \\ &= (m+1)\begin{pmatrix} 2m+1 \\ m \end{pmatrix}\cdot\theta^2\cdot\int_0^1 u^{m+2}(1-u)^m du \\ &= (m+1)\begin{pmatrix} 2m+1 \\ m \end{pmatrix}\cdot\theta^2\cdot\frac{(m+2)!m!}{(2m+3)!} = \left[\frac{m+2}{2(2m+3)}\right]\theta^2 \\ \operatorname{Var}(\tilde{X}) &= \frac{1}{4(2m+3)}\theta^2 \end{aligned}$$

Remark. Interestingly, the median has less variance than the mean.

3. **Note.** We have been told to calculate the mean and variance of a rather silly estimator: X_j, a single independent realization of the given distribution. We calculate as follows,

$$\begin{aligned} \mathbb{E}[X_j] &= \int_0^\theta \frac{1}{\theta}x dx = \frac{\theta}{2} \\ \operatorname{Var}(X_j) &= \int_0^\theta \frac{1}{\theta}x^2 dx - \left[\int_0^\theta \frac{1}{\theta}x dx\right]^2 = \frac{\theta^2}{12} \end{aligned}$$

4. Let $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$. So, the Fisher information is given by

$$I(\mu) = -\mathbb{E}\left[\frac{\partial^2}{\partial \mu^2}\left(-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{(X-\mu)^2}{2\sigma^2}\right)\right] = \frac{1}{\sigma^2}$$

Therefore, for any unbiased estimator of μ, say $\hat{\theta}_n$, we have $\operatorname{Var}(\hat{\theta}) \geq \frac{\sigma^2}{n}$. For the sample mean $\bar{X} = \frac{1}{n}\sum_{j=1}^n X_j, \mathbb{E}[\bar{X}] = \mu$ so it is unbiased and the variance is $\operatorname{Var}(\bar{X}) = \frac{1}{n^2} \cdot n \operatorname{Var}\left(X_i\right) = \frac{\sigma^2}{n}$. It follows that $\operatorname{Var}(\hat{\theta}) \geq \operatorname{Var}(\bar{X})$ and the conclusion follows.

5. The estimator for variance is $\hat{\nu} = n \cdot \frac{Y}{n}\left(1 - \frac{Y}{n}\right) = \frac{1}{n}Y(n-Y)$. Of course, one can evaluate the expectation and variance by calculating moments using the moment generating function. But here I am calculating directly.

$$\begin{aligned}
\frac{j(n-j)}{n}\binom{n}{j} &= \frac{(n-1)!}{(j-1)!(n-j-1)!} = (n-1)\cdot\binom{n-2}{j-1} \\
\left[\frac{j(n-j)}{n}\right]^2\binom{n}{j} &= \frac{j(n-j)}{n}\cdot(n-1)\cdot\binom{n-2}{j-1} \\
&= \frac{(j-1)(n-j-1)+(n-1)}{n}\cdot(n-1)\cdot\binom{n-2}{j-1} \\
&= \frac{(n-1)(n-2)(n-3)}{n}\binom{n-4}{j-2} + \frac{(n-1)^2}{n}\binom{n-2}{j-1}
\end{aligned}$$

Now, the calculation is straightforward:

$$\begin{aligned}
\mathbb{E}[\hat{\nu}] &= \sum_{j=0}^{n}\frac{j(n-j)}{n}\binom{n}{j}\theta^j(1-\theta)^{n-j} \\
&= (n-1)\theta(1-\theta)\sum_{j=1}^{n-1}\binom{n-2}{j-1}\theta^{j-1}(1-\theta)^{n-j-1} \\
&= (n-1)\theta(1-\theta) \\
\mathbb{E}\left[\hat{\nu}^2\right] &= \sum_{j=0}^{n}\left[\frac{j(n-j)}{n}\right]^2\binom{n}{j}\theta^j(1-\theta)^{n-j} \\
&= \sum_{j=2}^{n-4}\frac{(n-1)(n-2)(n-3)}{n}\binom{n-4}{j-2}\theta^j(1-\theta)^{n-j} \\
&+ \sum_{j=1}^{n-2}\frac{(n-1)^2}{n}\binom{n-2}{j-1}\theta^j(1-\theta)^{n-j} \\
&= \frac{(n-1)(n-2)(n-3)}{n}\theta^2(1-\theta)^2\sum_{j=2}^{n-4}\binom{n-4}{j-2}\theta^{j-2}(1-\theta)^{n-j-2} \\
&= \frac{(n-1)^2}{n}\theta(1-\theta)\sum_{j=1}^{n-2}\binom{n-2}{j-1}\theta^{j-1}(1-\theta)^{n-j-1} \\
\operatorname{Var}(\hat{\nu}) &= \mathbb{E}\left[\hat{\nu}^2\right] - (n-1)^2\theta^2(1-\theta)^2 \\
&= -6\frac{(n-1)^2}{n^2}\theta^2(1-\theta)^2 + \frac{(n-1)^2}{n}\theta(1-\theta) \\
&= \frac{(n-1)^2}{n}\theta(1-\theta)\left[1 - \frac{6}{n}\theta(1-\theta)\right]
\end{aligned}$$

6. Let $X_1, \ldots, X_n \sim$ Uniform $[0, \theta]$. since, $f_X(x;\theta) = \frac{1}{\theta}\cdot\mathbf{1}_{\{0\leq x\leq\theta\}}$, the joint density is given by

$$f_{\mathbf{X}}\left(x_1, \ldots, x_n; \theta\right) = \prod_{j=1}^{n}\frac{1}{\theta}\cdot\mathbf{1}_{\{0\leq x_j\leq\theta\}} = \frac{1}{\theta^n}\cdot\mathbf{1}_{\{0\leq\min_j x_j\}}\cdot\mathbf{1}_{\{\max_j x_j\leq\theta\}}$$

The maximum is achieved at $\theta = \max_j x_j$. Therefore, the maximum likelihood estimator is $\hat{\theta} = \max_j X_j$

Remark. If it is not clear to you how the maximum was derived, just draw a graph!

7. The joint density of $(\mathbf{X}, \mathbf{Y})$ is given by

$$f_{\mathbf{X},\mathbf{Y}}(x_1, \ldots, x_n, y_1, \ldots, y_n; \theta_1, \theta_2) = \frac{1}{\theta_1^n \theta_2^n} \exp\left[-\frac{1}{\theta_1}\sum_{j=1}^n x_j - \frac{1}{\theta_2}\sum_{j=1}^n y_j\right]$$

The maximum is achieved when,

$$\frac{\partial}{\partial \theta_1}\left[\log f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}; \theta_1, \theta_2)\right] = \frac{\partial}{\partial \theta_2}\left[\log f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}; \theta_1, \theta_2)\right] = 0$$

which implies, $\theta_1 = \frac{1}{n}\sum_{j=1}^n x_j$ and $\theta_2 = \frac{1}{n}\sum_{j=1}^n y_j$. So, the maximum likelihood estimators are

$$\hat{\theta}_1 = \frac{1}{n}\sum_{j=1}^n X_j \quad \hat{\theta}_2 = \frac{1}{n}\sum_{j=1}^n Y_j$$

Now, keeping in mind that X and Y are independent, we evaluate the Fisher information for θ_1 and θ_2 :

$$I(\theta_1) = -\mathbb{E}\left[\tfrac{\partial^2}{\partial \theta_1^2}\left(-\log\theta_1 - \tfrac{X}{\theta_1}\right)\right] = \tfrac{1}{\theta_1^2}$$
$$I(\theta_2) = -\mathbb{E}\left[\tfrac{\partial^2}{\partial \theta_2^2}\left(-\log\theta_2 - \tfrac{Y}{\theta_2}\right)\right] = \tfrac{1}{\theta_2^2}$$

Now the covariance matrix of a vector is a matrix, so to understand the behaviour of $\widehat{\boldsymbol{\theta}} = \begin{pmatrix}\widehat{\theta}_1 \\ \widehat{\theta}_2\end{pmatrix}$, we need to figure out how to compare the variance of linear combinations of the vector, with the variance of linear combinations of another estimator, $\widetilde{\boldsymbol{\theta}} = \begin{pmatrix}\widetilde{\theta}_1 \\ \widetilde{\theta}_2\end{pmatrix}$. Now as the covariance matrix of $\widehat{\boldsymbol{\theta}} = \begin{pmatrix}\widehat{\theta}_1 \\ \widehat{\theta}_2\end{pmatrix}$ is diagonal, it is appropriate to look at the trace of the covariance matrix and compare it to the trace of any other; say based on the Cramer lower bound of the estimators derived from the two independent samples, as X and Y are independent.

8. (a) Clearly, $\mathbb{E}[\hat{\theta}(C)] = \frac{n}{n+1}C\theta$ and $\operatorname{Var}(\hat{\theta}(C)) = \frac{n}{(n+2)(n+1)^2}C^2\theta^2$

(b) We calculate:

$$\begin{aligned}\operatorname{MSE}[\hat{\theta}(C)] &= [\mathbb{E}[\hat{\theta}(C)] - \theta]^2 + \operatorname{Var}(\hat{\theta}(C)) \\ &= \left[\frac{n}{n+1}C\theta - \theta\right]^2 + \frac{n}{(n+2)(n+1)^2}C^2\theta^2\end{aligned}$$

(c) $\mathbb{E}[\hat{\theta}(C)] = \theta$ implies that $C = 1 + \frac{1}{n}$

(d) $\operatorname{MSE}[\hat{\theta}(C)]$ is minimum when

$$\frac{d}{dC}\operatorname{MSE}[\hat{\theta}(C)] = 2\left[\frac{n}{n+1}C\theta - \theta\right] \cdot \frac{n}{n+1} \cdot \theta + \frac{n}{(n+2)(n+1)^2}\theta^2 \cdot 2C = 0$$

which gives $C = 1 + \frac{1}{n+1}$. We conclude that the best estimator in the mean square sense doesn't have to be unbiased. Or that allowing for a nonzero bias can improve the mean square performance of an estimator.