STAT 713 hw 4

Rao-Blackwell, Fisher information, Cramér-Rao lower bound

Do problems 7.40, 7.41 from CB. In addition:

1. Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Poisson}(\theta)$, $\theta > 0$, and consider estimating $\tau(\theta) = \theta(1+\theta)$.

   (a) Verify that $T = \sum_{i=1}^n X_i$ is a complete sufficient statistic for θ.

   You can show this using the exponential families result.

   (b) Propose an unbiased estimator $\tilde{\tau}$ of $\tau(\theta)$ based on X_1.

   Since $\mathbb{E}X_1 = \theta$ and $\text{Var}\,X_1 = \theta$, we have $\mathbb{E}X_1^2 = \theta + \theta^2 = \theta(1+\theta)$. So we can set $\tilde{\tau} = X_1^2$.

   (c) Now construct another unbiased estimator $\hat{\tau}$ for $\tau(\theta)$ by Rao-Blackwellization, that is, as $\hat{\tau} = \mathbb{E}[\tilde{\tau}|T]$. *Hint: You will need to find the conditional pmf of X_1 given T.*

   In order to take the expectation $\hat{\tau}(t) = \mathbb{E}[X_1^2|T = t]$, we need the conditional pmf of X_1 given $T = t$. For each $t = 0, 1, \dots$ We have, for $x_1 = 0, 1, \dots, t$,

$$\begin{aligned} P(X_1 = x_1|T = t) &= \frac{P_\theta(X_1 = x_1 \cap T = t)}{P_\theta(T = t)} \\ &= \frac{P_\theta(X_1 = x_1 \cap \sum_{i=2}^n X_i = t - x_1)}{P_\theta(T = t)} \\ &= \frac{(e^{-\theta}\theta^{x_1}/x_1!) \cdot (e^{-(n-1)\theta}[(n-1)\theta]^{t-x_1}/(t-x_1)!)}{e^{-n\theta}(n\theta)^t/t!} \\ &= \binom{n}{x_1}(1/n)^{x_1}(1-1/n)^{t-x_1}, \end{aligned}$$

   where we have used the fact that $T \sim \text{Poisson}(n\theta)$ and $\sum_{i=2}^n X_i \sim \text{Poisson}((n-1)\theta)$. Now we have

$$\hat{\tau}(t) = \mathbb{E}[X_1^2|T = t] = \text{Var}[X_1|T = t] + (\mathbb{E}[X_1|T = t])^2 = t(1/n)(1-1/n) + (t/n)^2.$$

   Simplifying, we have

$$\hat{\tau} = \frac{T}{n} + \frac{T(T-1)}{n^2}.$$

   (d) Is the estimator from part (c) the UMVUE for $\tau(\theta)$?

   Yes, $\hat{\tau}$ is the UMVUE for $\tau(\theta)$, since it is a function of a complete sufficient statistic and it is unbiased.

2. Let $Y_1, \ldots, Y_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\theta)$, $\theta > 0$, and consider estimating $\tau(\theta) = P_\theta(Y_1 \leq a) = 1 - e^{-a/\theta}$ for some $a > 0$.

   (a) Propose an unbiased estimator $\tilde{\tau}$ for τ.

   An unbiased estimator for τ is $\tilde{\tau} = \mathbf{1}(Y_1 \leq a)$.

   (b) Identify a complete sufficient statistic T for θ.

   By writing the Exponential(θ) pdf in exponential family form, we can see that $T = \sum_{i=1}^n Y_i$ is a complete sufficient statistic for θ.

   (c) Obtain the UMVUE for τ by finding $\hat{\tau} = \mathbb{E}[\tilde{\tau}|T]$.

   Begin by writing

$$\hat{\tau}(t) = \mathbb{E}[\tilde{\tau}|T = t] = P(Y_1 \leq a | \textstyle\sum_{i=1}^n Y_i = t)$$

   In order to obtain this probability, we need to find the conditional density of Y_1 given the sum $\sum_{i=1}^n Y_i$. To find this, set $Y_{-1} = \sum_{i=2}^n Y_i$ and note that $Y_{-1} \sim \text{Gamma}(n-1, \theta)$, which can be shown using moment generating functions. Letting $U = Y_1$ and $V = Y_1 + Y_{-1}$, we see that we must find the conditional density of $U|V = v$. The joint pdf of (Y_1, Y_{-1}) is given by

$$f_{Y_1,Y_{-1}}(y_1, y_{-1}) = f_{Y_1}(y_1; \theta) f_{Y_{-1}}(y_{-1}; \theta) = \frac{1}{\theta} e^{-y_1/\theta} \frac{1}{\Gamma(n-1)\theta^{(n-1)}} y_{-1}^{(n-1)-1} e^{-y_{-1}/\theta},$$

   owing to the independence of Y_1 and Y_{-1}. By the bivariate transformation method, the joint pdf of U and V is given by

$$f_{u,v}(u, v) = \frac{1}{\theta} e^{-u/\theta} \frac{1}{\Gamma(n-1)\theta^{(n-1)}} (v - u)^{(n-1)-1} e^{-(v-u)/\theta} \mathbf{1}(0 < u < v).$$

   The marginal distribution of $V = Y_1 + Y_{-1}$ is the Gamma(n, θ) distribution, so the conditional pdf of U given $V = v$ is given by

$$f_{U|V}(u|v) = \frac{f_{U,V}(u, v)}{f_V(v)} = (n-1)\frac{1}{v}\left(1 - \frac{u}{v}\right)^{(n-1)-1} \mathbf{1}(0 < u < v).$$

   Lastly, we obtain $\hat{\tau}(t)$ as

$$\hat{\tau}(t) = \int_0^a (n-1)\frac{1}{t}\left(1 - \frac{u}{t}\right)^{(n-1)-1} du = \left[1 - \left(1 - \frac{a}{t}\right)^{n-1}\right],$$

   so that we may write

$$\hat{\tau} = \left[1 - \left(1 - \frac{a}{T}\right)^{n-1}\right],$$

   where $T = \sum_{i=1}^n Y_i$.

3. Let $Y_1, \dots, Y_n$ be independent rvs and $x_1, \dots, x_n$ known constants such that $Y_i = \beta x_i + \varepsilon_i$, for $i = 1, \dots, n$, where $\varepsilon_1, \dots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$, for some $\beta \in \mathbb{R}$ and $\sigma > 0$.

   (a) Find the Fisher information $I(\beta, \sigma^2)$. *You may want to put* $\gamma = \sigma^2$ *until your calculations are done.*

   We obtain

   $$I(\beta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n x_i^2 & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

   (b) Give the CRLB for unbiased estimators of β.

   For any unbiased estimator $\hat{\tau}$ of $\tau(\beta, \sigma^2) = \beta$ we have

   $$\text{Var}\,\hat{\tau} \geq \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

   (c) Give the CRLB for unbiased estimators of σ^2.

   For any unbiased estimator $\hat{\tau}$ of $\tau(\beta, \sigma^2) = \sigma^2$ we have

   $$\text{Var}\,\hat{\tau} \geq \frac{2\sigma^4}{n}.$$

   (d) Give the MLE for β and check whether it is unbiased.

   The MLE for β is

   $$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

   We have

   $$\mathbb{E}\hat{\beta} = \frac{\sum_{i=1}^n x_i \mathbb{E} Y_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i \cdot x_i \beta}{\sum_{i=1}^n x_i^2} = \beta.$$

   (e) Check whether the MLE for β achieves the CRLB.

   We have

   $$\text{Var}\,\hat{\beta} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2},$$

   so $\hat{\beta}$ achieves the CRLB.

4. Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$.

(a) Find the Fisher information $I(\alpha, \beta)$.
Hint: Use $I(\theta) = -\mathbb{E}[\frac{\partial^2}{\partial \theta^2} \log(f(\mathbf{X}; \theta))]$ *and let* $\psi'(z) = \frac{d^2}{dz^2} \log \Gamma(z)$.

The likelihood and log-likelihood are given by

$$\mathcal{L}(\alpha, \beta; \mathbf{X}) = [\Gamma(\alpha)\beta^\alpha] - n \left(\prod_{i=1}^n X_i\right)^{\alpha-1} \exp\left[-\sum_{i=1}^n X_i/\beta\right]$$
$$\ell(\alpha, \beta; \mathbf{X}) = -n \log \Gamma(\alpha) - n\alpha \log \beta + (\alpha - 1) \sum_{i=1}^n \log X_i - \frac{1}{\beta} \sum_{i=1}^n X_i.$$

From here we obtain the entries of the score vector

$$\frac{\partial}{\partial \alpha} \mathcal{L}(\alpha, \beta; \mathbf{X}) = -n\psi(\alpha) - n \log \beta + \sum_{i=1}^n \log X_i$$
$$\frac{\partial}{\partial \beta} \mathcal{L}(\alpha, \beta; \mathbf{X}) = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n X_i,$$

where $\psi(\alpha) = \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)$ is the digamma function. Then the entries of the Hessian (the second-derivative matrix) are

$$\frac{\partial^2}{\partial \alpha^2} \mathcal{L}(\alpha, \beta; \mathbf{X}) = -n\psi'(\alpha)$$
$$\frac{\partial^2}{\partial \beta \partial \alpha} \mathcal{L}(\alpha, \beta; \mathbf{X}) = -\frac{n}{\beta}$$
$$\frac{\partial^2}{\partial \beta^2} \mathcal{L}(\alpha, \beta; \mathbf{X}) = \frac{n\alpha}{\beta^2} - \frac{2}{\beta^3} \sum_{i=1}^n X_i,$$

where $\psi'(\alpha) = \frac{\partial^2}{\partial \alpha^2} \log \Gamma(\alpha)$ is the trigamma function. The Fisher information is thus

$$I(\alpha, \beta) = \begin{bmatrix} n\psi'(\alpha) & n/\beta \\ n/\beta & n\alpha/\beta^2 \end{bmatrix}.$$

(b) Find the CRLB for unbiased estimators of $\tau(\alpha, \beta) = \alpha\beta$.

We have $\frac{\partial}{\partial \alpha} \tau(\alpha, \beta) = \beta$ and $\frac{\partial}{\partial \beta} \tau(\alpha, \beta) = \alpha$, so the the CRLB for unbiased estimators of $\tau(\alpha, \beta) = \alpha\beta$ is given by

$$\begin{bmatrix} \beta & \alpha \end{bmatrix} \begin{bmatrix} n\psi'(\alpha) & n/\beta \\ n/\beta & n\alpha/\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \beta & \alpha \end{bmatrix} \frac{\beta^2}{n(\alpha\psi'(\alpha) - 1)} \begin{bmatrix} n\alpha/\beta^2 & -n/\beta \\ -n/\beta & n\psi'(\alpha) \end{bmatrix}^{-1} \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$$
$$= \alpha\beta^2/n.$$

(c) Check whether $\hat{\tau} = \bar{X}_n$ achieves the CRLB for estimating $\tau(\alpha, \beta) = \alpha\beta$.

We know that $\mathbb{E}\bar{X}_n = \alpha\beta$ and $\operatorname{Var}\bar{X}_n = \alpha\beta^2/n$, so it achieves the CRLB.

5. (Optional) For $t = 1, \ldots, T$, let $X_t \sim \text{Normal}(\cos(2\pi t/T + \phi), 1)$, with $X_1, \ldots, X_T$ indep., for some $\phi \in [0, 2\pi)$.

(a) Show that the MLE for ϕ is the minimizer of the least-squares criterion

$$Q(\phi) = \sum_{t=1}^{T} (X_t - \cos(2\pi t/T + \phi))^2 .$$

The likelihood function is given by

$$\mathcal{L}(\phi; \mathbf{X}) = (2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{T} (X_t - \cos(2\pi t/T + \phi))^2\right].$$

We can see that this is maximized when the least-squares criterion $Q(\phi)$ is minimized.

(b) Show that the Fisher information is $I(\phi) = \sum_{t=1}^{T} \sin^2(2\pi t/T + \phi) = T/2$.
Hint: Use $\sin(2x) = 2\cos x \sin x$ *and* $\cos 2x = \cos^2 x - \sin^2 x$ *and use* $I(\phi) = -\mathbb{E}\frac{\partial^2}{\partial \phi^2} \log f(\mathbf{X}; \phi)$.

The log-likelihood is given by

$$\ell(\phi; \mathbf{X}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} (X_t - \cos(2\pi t/T + \phi))^2 ,$$

of which the first and second derivatives with respect to ϕ are given by

$$\frac{\partial}{\partial \phi}\ell(\phi; \mathbf{X}) = -\sum_{t=1}^{T} X_t \sin(2\pi t/T + \phi) + \frac{1}{2}\sum_{t=1}^{T} \sin(4\pi t/T + 2\phi)$$

$$\frac{\partial^2}{\partial \phi^2}\ell(\phi; \mathbf{X}) = -\sum_{t=1}^{T} X_t \cos(2\pi t/T + \phi) + \sum_{t=1}^{T} \cos(4\pi t/T + 2\phi),$$

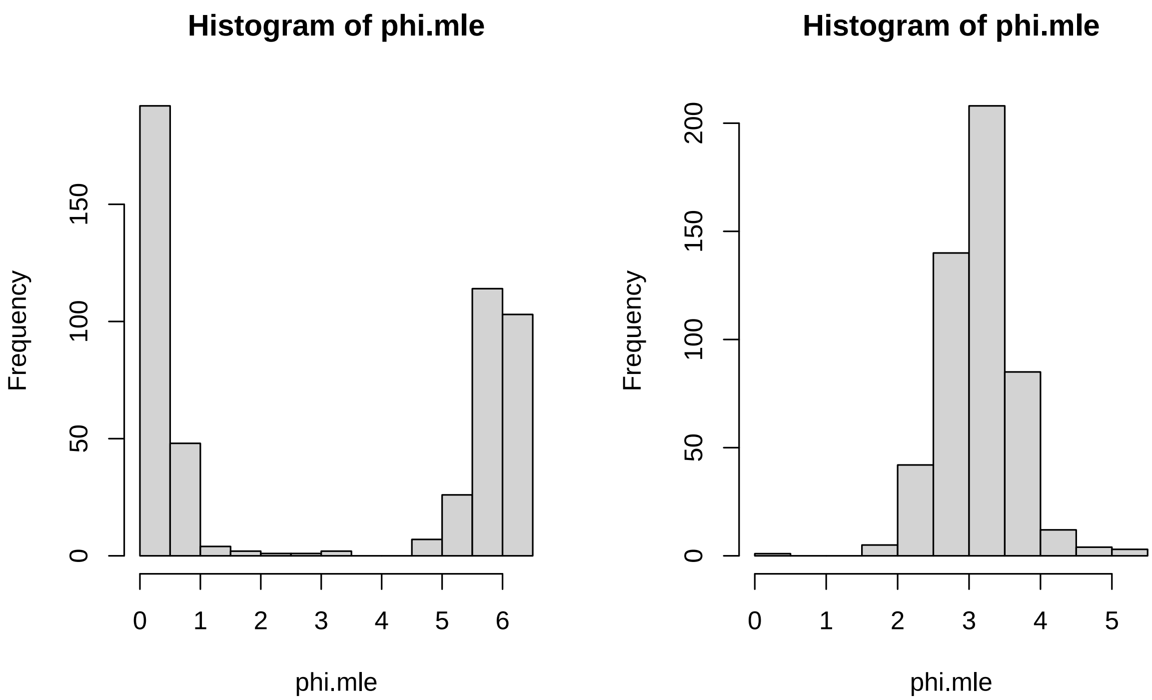
where we have made use of the trigonometric identity $\sin(2x) = 2\cos x \sin x$. Now

$$-\mathbb{E}\frac{\partial^2}{\partial \phi^2}\ell(\phi; \mathbf{X}) = \sum_{t=1}^{T} \cos^2(2\pi t/T + \phi) - \sum_{t=1}^{T} \cos(4\pi t/T + 2\phi) = \sum_{t=1}^{T} \sin^2(2\pi t/T + \phi),$$

where we have used $\cos 2x = \cos^2 x - \sin^2 x$. It turns out this sums to $T/2$.

(c) For $\phi = 0$ and $\phi = \pi$, generate 500 data sets with $T = 10$ and compute the MLE for ϕ on each of the 500 simulated data sets. Make histograms of the 500 values of the MLE and report the variance of the MLE values. Does the MLE appear to achieve the CRLB for unbiased estimators of ϕ? Are the cases of $\phi = 0$ and $\phi = \pi$ different?

The histograms look like this:



The left one is under $\phi = 0$ and the right one is under $phi = \pi$. The variances of the MLE under $phi = 0$ and $\phi = \pi$ were 7.699 and 0.270, respectively, and the CRLB is $2/10 = 0.20$, so under $\phi = \pi$ the CRLB is almost achieved. It seems like it a phase shift close to zero is indistinguishable from a phase shift close to 2π, accounting for the bimodality of the first histogram. In order to meaningfully compute the variance of the MLE in this case, one should take a circular view of the estimates (estimates close to 2π are actually close to 0, rather than far away, as the histogram shows).

Problems 7.40 and 7.41 from CB

7.40 Let $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$.

We have $\text{Var}\, \bar{X}_n = \frac{p(1-p)}{n}$ and

$$L(p; \underset{\sim}{x}) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^{n\bar{x}_n} (1-p)^{n - n\bar{x}_n}$$

$$\ell(p; \underset{\sim}{x}) = n\bar{x}_n \log p + (n - n\bar{x}_n) \log(1-p)$$

$$S(p; \underset{\sim}{x}) = \frac{\partial}{\partial p} \ell(p; \underset{\sim}{x})$$

$$= \frac{n\bar{x}_n}{p} + \frac{n - n\bar{x}_n}{1-p}$$

$$= n\bar{x}_n \left[\frac{1}{p} + \frac{1}{1-p} \right] + \frac{n}{1-p}$$

$$= \frac{n\bar{x}_n}{p(1-p)} + \frac{n}{1-p}$$

$$I(p) = \text{Var}\, S(p; \underset{\sim}{X}) = \frac{n^2}{p^2(1-p)^2} \frac{p(1-p)}{n} = \frac{n}{p(1-p)}.$$

The CRLB for unbiased estimators of p is

$$\text{CRLB} = \frac{1}{I(p)} = \frac{p(1-p)}{n}.$$

Since $\text{Var}\, \bar{X}_n$ is equal to this it is the best unbiased estimator.

7.41 Let $X_1, \dots, X_n$ be iid with mean μ and variance σ^2.
Let $\hat{\mu}_{\underset{\sim}{a}} = \sum_{i=1}^n a_i X_i$.

(a) We have $\mathbb{E}\, \hat{\mu}_{\underset{\sim}{a}} = \mathbb{E} \sum_{i=1}^n a_i X_i = \sum_{i=1}^n a_i \mu = \mu$ if $\sum_{i=1}^n a_i = 1$.

(b) We have $\operatorname{Var} \hat{\mu}_{\underset{\sim}{a}} = \sigma^2 \sum_{i=1}^n a_i^2$.

To minimize $\operatorname{Var} \hat{\mu}_{\underset{\sim}{a}}$ subject to the constraint $\sum_{i=1}^n a_i = 1$ we set

$$\mathcal{L}(\underset{\sim}{a}, \lambda) = \sigma^2 \sum_{i=1}^n a_i^2 + \lambda \left(1 - \sum_{i=1}^n a_i\right).$$

Then we have

$$\frac{\partial}{\partial a_1} \mathcal{L}(\underset{\sim}{a}, \lambda) = 2\sigma^2 a_1 - \lambda = 0$$

$$\vdots$$

$$\frac{\partial}{\partial a_n} \mathcal{L}(\underset{\sim}{a}, \lambda) = 2\sigma^2 a_n - \lambda = 0$$

$$\frac{\partial}{\partial \lambda} \mathcal{L}(\underset{\sim}{a}, \lambda) = 1 - \sum_{i=1}^n a_i = 0$$

The first n equations give

$$a_1 = \dots = a_n = \frac{\lambda}{2\sigma^2} \implies \sum_{i=1}^n a_i = \frac{n\lambda}{2\sigma^2}.$$

Then the last equation gives

$$1 = \sum_{i=1}^n a_i = \frac{n\lambda}{2\sigma^2} \implies \lambda = \frac{2\sigma^2}{n}.$$

So we have

$$a_1 = \dots = a_n = \frac{1}{n}.$$

Under this choice of $a_1, \dots, a_n$, $\operatorname{Var} \hat{\mu}_{\underset{\sim}{a}} = \frac{\sigma^2}{n}$.