

# ECE531 Homework Assignment Number 8 Solution

Due by 8:50pm on Wednesday 6-Apr-2011

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

1. 5 points. Suppose  $Y$  is a scalar observation drawn from a parameterized Poisson distribution

$$p_Y(y; \theta) = \text{Prob}(Y = y) = \frac{\theta^y e^{-\theta}}{y!}$$

for  $y = 0, 1, 2, \dots$ . Find the Fisher information  $I(\theta)$  and the CRLB for estimating the scalar parameter  $\theta$ . Confirm all of the regularity conditions are satisfied. Can you find an MVU estimator that achieves the CRLB in this case?

**Solution:** First, let's check the regularity conditions:

- $\Lambda = \mathbb{R}$ , which is indeed an open interval.
- The support of the density does  $p_Y(y; \theta)$  does not change as a function of  $\theta$  since  $p_Y(y; \theta) > 0$  for all  $y = 0, 1, 2, \dots$  irrespective of  $\theta$ .
- We can compute

$$\frac{\partial}{\partial \theta} p_Y(y; \theta) = \frac{y\theta^{y-1}e^{-\theta} - \theta^y e^{-\theta}}{y!}$$

which implies that  $\frac{\partial}{\partial \theta} p_Y(y; \theta)$  exists and is finite for all  $\theta \in \Lambda$  and all  $y = 0, 1, \dots$

- Finally, we just need to check if the integral and derivative can be exchanged. We can compute

$$\begin{aligned} \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} p_Y(y; \theta) dy &= \sum_{y=0}^{\infty} \frac{y\theta^{y-1}e^{-\theta} - \theta^y e^{-\theta}}{y!} \\ &= e^{-\theta} \sum_{y=0}^{\infty} \frac{y\theta^{y-1} - \theta^y}{y!} \\ &= e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^{y-1}}{(y-1)!} - e^{-\theta} \sum_{y=0}^{\infty} \frac{\theta^y}{y!} \\ &= e^{-\theta} e^{\theta} - e^{-\theta} e^{\theta} \\ &= 0 \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} p_Y(y; \theta) dy. \end{aligned}$$

So this verifies the first four regularity conditions of the theorem, except for the part about exchanging the derivative and integral when the estimator is inside the integral.

- The fifth regularity condition can also be confirmed by taking another derivative of  $p_Y(y; \theta)$ , confirming that it exists, and integrating it to get zero.

With the regularity conditions all confirmed, we can compute

$$\frac{\partial}{\partial \theta} \ln p_Y(y; \theta) = \frac{\partial}{\partial \theta} (-\theta + y \ln \theta) = -1 + \frac{y}{\theta},$$

and

$$\frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) = -\frac{y}{\theta^2} < 0.$$

The Fisher information is given by

$$I_\theta = -E_\theta \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_Y(Y; \theta) \right\} = \frac{E_\theta \{Y\}}{\theta^2} = \frac{1}{\theta}.$$

So the CRLB in this case is simply

$$\text{var}[\hat{\theta}(Y)] \geq \theta.$$

As for an MVU estimator, let's try

$$\hat{\theta}(y) = y.$$

Since  $Y$  is Poisson, we have  $E_\theta\{\hat{\theta}(Y)\} = \text{var}_\theta\{\hat{\theta}(Y)\} = \theta$ . So  $\hat{\theta}(y)$  is an unbiased estimator of  $\theta$  and achieves the CRLB. This estimator must then be MVU. As a final step, you should confirm that the integral and derivative can be exchanged when this estimator is inside the integral (the second part of regularity condition 4). We can compute

$$\begin{aligned} \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \hat{\theta}(y) p_Y(y; \theta) dy &= \sum_{y=0}^{\infty} \frac{y^2 \theta^{y-1} e^{-\theta} - y \theta^y e^{-\theta}}{y!} \\ &= e^{-\theta} \sum_{y=0}^{\infty} \frac{y^2 \theta^{y-1} - y \theta^y}{y!} \\ &= e^{-\theta} \sum_{y=1}^{\infty} \frac{y \theta^{y-1}}{(y-1)!} - e^{-\theta} \sum_{y=1}^{\infty} \frac{y \theta^y}{y!} \\ &= e^{-\theta} \sum_{y=1}^{\infty} \frac{(y-1) \theta^{y-1}}{(y-1)!} - e^{-\theta} \sum_{y=1}^{\infty} \frac{y \theta^y}{y!} + e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^{y-1}}{(y-1)!} \\ &= 0 + e^{-\theta} e^\theta \\ &= 1 \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} \hat{\theta}(y) p_Y(y; \theta) dy \end{aligned}$$

since the estimator is unbiased. Hence, we confirmed the second part of the fourth regularity condition for this estimator.

2. 4 points. Kay I: 3.3. Confirm all of the regularity conditions are satisfied.

**Solution:** I will use  $y_n = x[n]$  for my observations and  $\theta = A$  for the unknown parameter. The joint distribution of the observations can be written as

$$p_Y(y; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} (y - \theta z)^\top (y - \theta z) \right\}$$

where  $z = [1, r, r^2, \dots, r^{N-1}]^\top$ . Let's check the regularity conditions...

- $\Lambda = \mathbb{R}$ , which is indeed an open interval.
- All densities in the family  $\{p_Y(y; \theta) : \theta \in \Lambda\}$  share common support, which is  $\mathbb{R}^N$ .
- The partial derivative

$$\frac{\partial}{\partial \theta} p_Y(y; \theta) = \frac{z^\top (y - \theta z)}{\sigma^2 (2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-(y - \theta z)^\top (y - \theta z)}{2\sigma^2} \right\}$$

clearly exists and is finite for all  $\theta \in \Lambda$  and all  $y$  in the common support.

- Since  $\frac{\partial}{\partial \theta} p_Y(y; \theta)$  is odd around  $y = \theta z$ , it is easy to see that

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}} p_Y(y; \theta) dy = \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} p_Y(y; \theta) dy = 0.$$

This confirms the first part of the fourth regularity condition.

- The fifth regularity condition can also be confirmed by taking another derivative of  $p_Y(y; \theta)$ , confirming that it exists, and integrating it to get zero.

With the regularity conditions confirmed (except for the part about exchanging the derivative and integral when the estimator is inside the integral in the fourth regularity condition), we can proceed.

We can compute

$$-E \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) \right\} = \frac{1}{\sigma^2} z^\top z$$

which implies the CRLB in this case is simply

$$\text{var}[\hat{\theta}(Y)] \geq \frac{\sigma^2}{z^\top z}.$$

To show an efficient estimator exists, we can write

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln p_Y(y; \theta) &= \frac{1}{\sigma^2} z^\top (y - \theta z) \\ &= \underbrace{\frac{z^\top z}{\sigma^2}}_{k(\theta) \equiv k} \left( \underbrace{\frac{z^\top y}{z^\top z}}_{\hat{\theta}(y)} - \theta \right) \end{aligned}$$

which is the form required to attain the information bound. Hence

$$\hat{\theta}(y) = \frac{z^\top y}{z^\top z}$$

is an efficient unbiased estimator with variance  $\frac{\sigma^2}{z^\top z}$ . This makes intuitive sense since the estimates should be worse for larger  $\sigma^2$  and better for larger  $r$ . For  $0 < r < 1$ , when  $N \rightarrow \infty$

we get  $z^\top z \rightarrow \frac{1}{1-r^2}$ . For  $r \geq 1$ , when  $N \rightarrow \infty$  we get  $z^\top z$  grows unbounded. Hence, we can say

$$\lim_{N \rightarrow \infty} \text{var}[\hat{\theta}(Y)] \geq \begin{cases} \sigma^2(1-r^2) & 0 < r < 1 \\ 0 & r \geq 1. \end{cases}$$

As a final step, you should confirm that the integral and derivative can be exchanged when this MVU estimator is inside the integral (the second part of regularity condition 4). We can compute xxx

$$\begin{aligned} \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \hat{\theta}(y) p_Y(y; \theta) dy &= \int_{\mathcal{Y}} \frac{z^\top y}{z^\top z} \frac{z^\top (y - \theta z)}{\sigma^2 (2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-(y - \theta z)^\top (y - \theta z)}{2\sigma^2} \right\} dy \\ &= \int_{\mathcal{Y}} \frac{z^\top (x + \theta z)}{z^\top z} \frac{z^\top x}{(2\pi)^{N/2}} \exp \left\{ \frac{-x^\top x}{2} \right\} dx \\ &= \int_{\mathcal{Y}} \frac{z^\top x x^\top z}{z^\top z} \frac{1}{(2\pi)^{N/2}} \exp \left\{ \frac{-x^\top x}{2} \right\} dx + \int_{\mathcal{Y}} \theta \frac{z^\top x}{(2\pi)^{N/2}} \exp \left\{ \frac{-x^\top x}{2} \right\} dx \\ &= \frac{z^\top I z}{z^\top z} + 0 \\ &= 1 \\ &= \frac{\partial}{\partial \theta} \int_{\mathcal{Y}} \hat{\theta}(y) p_Y(y; \theta) dy \end{aligned}$$

since the estimator is unbiased. Hence, we confirmed the second part of the fourth regularity condition for this estimator.

3. 4 points. Kay I: 3.4. You can assume all of the regularity conditions are satisfied.

**Solution:** I will use  $y_n = x[n]$  for my observations and  $\theta = r$  for the unknown parameter. The joint distribution of the observations can be written as

$$p_Y(y; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ \frac{-1}{2\sigma^2} (y - z(\theta))^\top (y - z(\theta)) \right\}$$

where  $z(\theta) = [1, \theta, \theta^2, \dots, \theta^{N-1}]^\top$ . We will proceed under the assumption that the regularity conditions are satisfied.

We can compute the Fisher information as

$$\begin{aligned} I(\theta) &= -\mathbb{E} \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) \right\} \\ &= -\mathbb{E} \left\{ \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [(n-1)n\theta^{n-2}y_n - n(2n-1)\theta^{2n-2}] \right\} \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^2 \theta^{2n-2} \\ &= \frac{1}{\sigma^2} v^\top v \end{aligned}$$

where  $v(\theta) := [0, 1, 2\theta, 3\theta^2, \dots, (N-1)\theta^{N-2}]^\top$ . Hence, the CRLB follows as

$$\text{var}[\hat{\theta}(Y)] \geq \frac{\sigma^2}{v^\top v}.$$

To check whether an efficient estimator exists, we can compute

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln p_Y(y; \theta) &= \frac{1}{\sigma^2} (y - z(\theta))^\top v(\theta) \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (y_n - \theta^n) n \theta^{n-1} \end{aligned}$$

where  $v(\theta) = [0, 1, 2\theta, 3\theta^2, \dots, (N-1)\theta^{N-2}]^\top$ . This can't be put into the required form to attain the information bound, hence there will not be an efficient estimator.

4. 4 points. Kay I: 3.9. You can assume all of the regularity conditions are satisfied.

**Solution:** I will use  $y_n = x[n]$  for my observations and  $\theta = A$  for the unknown parameter. The joint distribution of the *two* observations can be written as a bivariate Gaussian distribution

$$p_Y(y; \theta) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{(y_0 - A)^2 - 2\rho(y_0 - A)(y_1 - A) + (y_1 - A)^2}{2(1-\rho^2)} \right\}$$

We will proceed under the assumption that the regularity conditions are satisfied.

We can compute the Fisher information as

$$\begin{aligned} I(\theta) &= -\mathbb{E} \left\{ \frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) \right\} \\ &= \frac{2}{\sigma^2(1+\rho)}. \end{aligned}$$

Hence, the CRLB follows as

$$\text{var}[\hat{\theta}(Y)] \geq \frac{\sigma^2(1+\rho)}{2}.$$

Let's look at some special cases:

- $\rho = 0$ : Recall that, when the observations are i.i.d., we have the information additivity property. In the case of two i.i.d. observations, the Fisher information for the parameter  $\theta = A$  would be

$$I(\theta) = \frac{2}{\sigma^2} \quad (\text{two i.i.d. observations})$$

which is consistent with our result above when  $\rho = 0$ . The CRLB in the case of two i.i.d. observations is, of course,

$$\text{var}[\hat{\theta}(Y)] \geq \frac{\sigma^2}{2} \quad (\text{two i.i.d. observations})$$

which is also consistent with our result above when  $\rho = 0$ .

- $\rho \rightarrow 1$ : In this case, the noise realization will be the same for both observations. So, intuitively, the second observation doesn't tell us anything new. We would expect in this case that the Fisher information should be the same as if we only had one observation. In fact, when  $\rho = 1$ , we can write

$$\lim_{\rho \rightarrow 1} I(\theta) = \lim_{\rho \rightarrow 1} \frac{2}{\sigma^2(1+\rho)} = \frac{1}{\sigma^2}$$

which agrees with our intuition. The CRLB in this case becomes

$$\text{var}[\hat{\theta}(Y)] \geq \sigma^2 \quad \rho \rightarrow 1.$$

- $\rho \rightarrow -1$ : In this case, the noise realization in the second observation will have opposite sign with the noise realization in the first observation. You could then just add the two observations to cancel out the noise and your estimate would be perfect. So, intuitively, the anti-correlated second observation gives us lots more information than an independent second observation. We should expect the Fisher information to be large in this case. In fact, when  $\rho \rightarrow -1$ , we can write

$$\lim_{\rho \rightarrow -1} I(\theta) = \lim_{\rho \rightarrow -1} \frac{2}{\sigma^2(1+\rho)}$$

which is clearly unbounded and agrees with our intuition. The CRLB in this case goes to zero.

Note that getting a second observation never reduces the amount of Fisher information received from the first observation (the Fisher information in each observation is non-negative). But the bottom line here is that the additivity property for Fisher information only holds for independent observations.

5. 4 points. Kay I: 3.12.

**Solution:** This problem is intended to show that the presence of additional unknown parameters never makes the lower bound on the variance of a parameter better (in fact, it usually makes it worse).

Since  $I(\theta)$  is a square invertible matrix, we can do an eigen-decomposition to write it as

$$I(\theta) = VDV^{-1}$$

where  $V$  is a square matrix of unit-norm orthogonal eigenvectors and  $D$  is diagonal matrix of strictly positive eigenvalues. We can also write

$$I(\theta) = VD^{-1}V^{-1}.$$

According to the hint, we define

$$\begin{aligned}\sqrt{I(\theta)} &:= VD^{1/2}V^{-1} \\ \sqrt{I^{-1}(\theta)} &:= VD^{-1/2}V^{-1}\end{aligned}$$

where the square root of a diagonal matrix is just another diagonal matrix with entries each equal to the square root of the entries of the original diagonal matrix. Note that

$$\sqrt{I(\theta)}\sqrt{I^{-1}(\theta)} = I$$

the identity matrix, hence

$$\left(e_i^\top \sqrt{I(\theta)} \sqrt{I^{-1}(\theta)} e_i\right)^2 = 1$$

since the effect of pre/post multiplying by  $e_i$  is to pick out the  $i^{\text{th}}$  diagonal element. Now we just follow the hint by applying the Cauchy-Schwarz inequality to write

$$\left(e_i^\top I(\theta) e_i\right) \left(e_i^\top I^{-1}(\theta) e_i\right) \geq 1.$$

But  $e_i^\top I(\theta) e_i := [I(\theta)]_{ii}$  and  $e_i^\top I^{-1}(\theta) e_i := [I^{-1}(\theta)]_{ii}$ . Hence, we arrive at the desired result

$$[I^{-1}(\theta)]_{ii} \geq \frac{1}{[I(\theta)]_{ii}}.$$

This bound can only be achieved when the Cauchy-Schwarz inequality becomes an equality, which only occurs if the vectors

$$\begin{aligned}u_i &= \sqrt{I(\theta)} e_i \\ v_i &= \sqrt{I^{-1}(\theta)} e_i\end{aligned}$$

are proportional to each other for all  $i$ , i.e.  $u_i = a_i v_i$  for  $a_i \neq 0$  and all  $i$ . This can only occur if  $I(\theta)$  is a diagonal matrix (and the original attainability conditions are satisfied). This makes intuitive sense since a diagonal  $I(\theta)$  says that the parameters are uncoupled in the sense that knowing a parameter doesn't tell you anything about another parameter.

But the key point here is that additional unknown parameters never make the lower bound on the variance of a parameter better.



6. 4 points. Kay I: 3.17.

**Solution:** Example 3.14 is a very useful example since it derives the CRLB of the amplitude, phase, and frequency of a sinusoid in AWGN when all three parameters are unknown. Intuitively, changing the sampling indices should not change the amplitude estimate or the frequency estimate, since it doesn't matter when the samples are taken. Changing the sampling indices, however, will change the phase estimate since the phase  $\phi$  corresponds to the phase of the signal at time  $n = 0$ .

So looking at each of the  $[I(\theta)]_{ij}$  terms in Example 3.14, we see that  $[I(\theta)]_{23}$  will change because the sum  $\sum_{n=-M}^M n = 0$ . This causes both the  $[I(\theta)]_{23}$  and  $[I(\theta)]_{32}$  terms to become zero and the whole  $I(\theta)$  matrix to become *diagonal*. In fact, we have

$$I(\theta) = \frac{1}{\sigma^2} \begin{bmatrix} \frac{N}{2} & 0 & 0 \\ 0 & 2A^2\pi^2 \sum_{n=-M}^M n^2 & 0 \\ 0 & 0 & \frac{NA^2}{2} \end{bmatrix}$$

and, defining  $\eta := A^2/(2\sigma^2)$  as the SNR, the variances are then bounded as

$$\begin{aligned} \text{var}(\hat{A}) &\geq \frac{2\sigma^2}{N} && \text{(same as before)} \\ \text{var}(\hat{f}_0) &\geq \frac{\sigma^2}{2A^2\pi^2 \sum_{n=-M}^M n^2} = \frac{1}{(2\pi)^2\eta \sum_{n=-M}^M n^2} && \text{(need to check)} \\ \text{var}(\hat{A}) &\geq \frac{2\sigma^2}{NA^2} = \frac{1}{N\eta} && \text{(better than before).} \end{aligned}$$

To check whether the lower bound on the estimate for  $f_0$  is any better, we can use the fact that

$$\begin{aligned} \sum_{n=-M}^M n^2 &= 2 \sum_{n=1}^M n^2 \\ &= \frac{2M(M+1)(2M+1)}{6} \\ &= \frac{1}{3} \left(\frac{N-1}{2}\right) \left(\frac{N+1}{2}\right) N \\ &= \frac{N(N^2-1)}{12} \end{aligned}$$

where  $N = 2M + 1$ , hence

$$\text{var}(\hat{f}_0) \geq \frac{12}{(2\pi)^2\eta N(N^2-1)}$$

which is the same as before.

Our intuition was correct. Changing the sampling indices does not change the bounds on the variance of the amplitude estimate or the frequency estimate, since it doesn't matter when the samples are taken. But changing the sample indices does make the bounds on the variance of the *phase* estimate lower. This is because the phase and frequency estimates are no longer coupled when the sampling indices changed to  $-M, \dots, M$ .

Physically, this means that we can get better estimates of the phase at the *middle* of the observation than we can at the beginning of the observation (when the frequency is also not known). If we try to estimate the phase at the beginning of the observation and the

frequency is not known, frequency estimation error couples into the phase estimate. This makes the achievable performance worse. If, on the other hand, we try to estimate the phase in the middle of the observation, the unknown frequency has no effect on the achievable performance of the phase estimator.