ECE531 Homework Assignment Number 7 Solution

Due by 8:50pm on Wednesday 30-Mar-2011

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

1. 4 points. Kay I: 2.6.

   Solution: I will use the notation $y_n = x[n]$ for my observations. We are given the parameter $\theta = A$ and

$$\hat{A} = \sum_{n=0}^{N-1} a_n y_n.$$

The unbiased constraint requires $\mathrm{E}(\hat{A}) = A$ for all A, so we can write

$$\begin{aligned} & \mathrm{E}[\hat{A}] = A \\ \Leftrightarrow\ & \sum_{n=0}^{N-1} a_n \mathrm{E}[Y_n] = A \\ \Leftrightarrow\ & \sum_{n=0}^{N-1} a_n A = A \\ \Leftrightarrow\ & \sum_{n=0}^{N-1} a_n = 1 \\ \Leftrightarrow\ & a^\top 1 = 1 \end{aligned}$$

where the 1 on the lefthand side is a vector of all ones and $a = [a_0, \dots, a_{N-1}]^\top$. This establishes a linear equality constraint on the coefficients $a_0, \dots, a_{N-1}$.

We can also calculate the variance of the unbiased estimator as

$$\begin{aligned} \mathrm{var}[\hat{A}] &= \sum_{n=0}^{N-1} a_n^2 \mathrm{var}[Y_n] \qquad \text{(independent observations)} \\ &= \sum_{n=0}^{N-1} a_n^2 \sigma^2 \\ &= \sigma^2 a^\top a. \end{aligned}$$

So we want to find the coefficient vector a that minimize the variance subject to the linear equality constraint above. We can do this with Lagrange multipliers as follows.

Let $f(a) = \sigma^2 a^\top a$ and $g(a) = a^\top 1 - 1$ (where the second 1 in this equation is a scalar). To find the extreme values of f subject to the constraint $g(a) = 0$, we have to solve the system of equations

$$\begin{aligned} \nabla f(a) &= \lambda \nabla g(a) \\ g(a) &= 0 \end{aligned}$$

Doing the calculus, we get

$$\nabla f(a) = 2\sigma^2 a$$

and

$$\lambda \nabla g(a) = \lambda 1$$

where this 1 is a vector. Putting these results together, we have

$$2\sigma^2 a = \lambda 1$$

which implies

$$a = \frac{\lambda}{2\sigma^2} 1. \tag{1}$$

All that remains is to find λ. The linear equality constraint says

$$\begin{aligned} & g(a) = 0 \\ \Leftrightarrow \quad & a^\top 1 - 1 = 0 \\ \Leftrightarrow \quad & \frac{\lambda}{2\sigma^2} 1^\top 1 - 1 = 0 \\ \Leftrightarrow \quad & \frac{\lambda N}{2\sigma^2} - 1 = 0 \\ \Leftrightarrow \quad & \lambda = \frac{2\sigma^2}{N}. \end{aligned}$$

Plugging this result back into (1), we have

$$a = \frac{1}{N} 1.$$

Which is the expected "sample-mean" result, i.e. $a_0 = a_1 = \cdots = a_{N-1} = \frac{1}{N}$. It is easy confirm the estimator is unbiased and the variance can also be calculated as $\text{var}[\hat{A}] = \sigma^2/N$.

2. 5 points total.

(a) 3 points. Kay I: 5.3. Please find a *complete* sufficient statistic and prove that it is complete.

Solution: The easiest way to the solution on this problem is to use the completeness theorem for exponential families (slide 26 of the lecture notes). Since the observations are i.i.d., the joint pdf of the observations can be written as

$$\begin{aligned} p_Y(y\,;\theta) &= \begin{cases} \prod_{n=0}^{N-1} \lambda e^{-\lambda y_n} & \min y_n > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= \prod_{n=0}^{N-1} \lambda e^{-\lambda y_n} u(\min y_n) \\ &= \underbrace{\lambda^N}_{a(\lambda)} \exp\left\{-\lambda \sum_{n=0}^{N-1} y_n\right\} \underbrace{u(\min y_n)}_{h(y)} \end{aligned}$$

where $u(x)$ is the usual unit step function. This is exactly in the form of the theorem with $\theta = \lambda$ and $T(y) = -\sum_{n=0}^{N-1} y_n$. One-to-one transformations don't change sufficiency or completeness, so we can go with

$$T(y) = \sum_{n=0}^{N-1} y_n = N\bar{y}$$

as our complete sufficient statistic.

(b) 2 points. Use the RBLS theorem to find a MVU estimator of the non-random parameter λ.

Solution: Now that we have our complete sufficient statistic, we just need to follow the steps in the RBLS theorem. First, we need an unbiased estimator of λ. Let's try

$$\hat{\lambda}(y) = \frac{1}{\frac{1}{N}\sum_{n=0}^{N-1} y_n} = \frac{1}{\bar{y}}$$

It turns out that this is actually a biased estimator. To see this, let $Z = \frac{1}{N}\sum_{n=0}^{N-1} Y_n$ be the sample mean (a random variable). The pdf of Z can be derived (or found in a textbook[1]) as

$$p_Z(z\,;\,\theta) = \frac{\lambda N e^{-\lambda N z}(\lambda N z)^{N-1}}{(N-1)!}$$

The mean of our estimator can then be calculated as

$$\begin{aligned} E[\hat{\lambda}(Y)] &= E[1/Z] \\ &= \int_0^\infty \frac{1}{z} \cdot \frac{\lambda N e^{-\lambda N z}(\lambda N z)^{N-1}}{(N-1)!}\, dz \\ &= \frac{N\lambda(N-2)!}{(N-1)!} \\ &= \frac{N}{N-1}\lambda \end{aligned}$$

which is clearly biased. But we can correct this by scaling by $\frac{N-1}{N}$. Hence, the estimator

$$\hat{\lambda}(y) = \frac{N-1}{N} \cdot \frac{1}{\bar{y}} = \frac{N-1}{N\bar{y}}$$

is unbiased.

So we set

$$\hat{g}(Y) = \frac{N-1}{N\bar{Y}} = \frac{N-1}{T(Y)}$$

[1]See page 300 of http://www.dartmouth.edu/ chance/teaching_aids/books_articles/probability_book/Chapter7.pdf.

and the last step is then to compute the conditional expectation where we condition on $T(Y) = T(y) = N\bar{y}$. Specifically,

$$\begin{aligned}
\tilde{g}[T(y)] &= \mathrm{E}_\lambda[\hat{g}(Y) \,|\, T(Y) = T(y)] \\
&= \mathrm{E}_\lambda\left[\frac{N-1}{T(Y)} \,|\, T(Y) = N\bar{y}\right] \\
&= \mathrm{E}_\lambda\left[\frac{N-1}{N\bar{y}}\right] \\
&= \frac{N-1}{N\bar{y}}
\end{aligned}$$

which is the same as our earlier unbiased estimator $\hat{\lambda}(y)$. According to the RBLS theorem, this is then the MVU estimator of the non-random parameter λ.

3. 4 points. Kay I: 5.7.

Solution: Let's first write the joint pdf on the observations, parameterized by f_0 as

$$p_Y(y\,;\,f_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} (y_n - \cos(2\pi f_0 n))^2 \right\}$$

since the noise is i.i.d. The Neyman-Fisher factorization theorem is an "if and only if" theorem, so we can use it to show that the only possible sufficient statistic is the trivial $T(y) = y$ statistic, where $y = [y_0, \ldots, y_{N-1}]^\top$. To confirm this, we need to do a little algebra

$$\begin{aligned} p_Y(y\,;\,f_0) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} (y_n - \cos(2\pi f_0 n))^2 \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ \frac{-y^\top y}{2\sigma^2} \right\} \exp\left\{ \frac{2y^\top z(f_0)}{2\sigma^2} \right\} \exp\left\{ \frac{-z^\top(f_0) z(f_0)}{2\sigma^2} \right\} \end{aligned}$$

where $z(f_0) := [1, \cos(2\pi f_0), \ldots, \cos(2\pi f_0(N-1))]^\top$. All the components that are a function of f_0 must be put into the $g_\theta(T(y))$ part of the factorization. Hence, we can say

$$g_\theta(T(y)) = \exp\left\{ \frac{2y^\top z(f_0)}{2\sigma^2} \right\} \exp\left\{ \frac{-z^\top(f_0) z(f_0)}{2\sigma^2} \right\}$$

and the other components of $p_Y(y\,;\,f_0)$ go into $h(y)$.

Note that y only appears in one place in $g_\theta(T(y))$. So it is tempting to say $T(y) = y^\top z(f_0)$. But this is incorrect because $T(y)$ can't be a function of the unknown parameter. So the only statistic for f_0 that is sufficient is $T(y) = y$ itself (or some one-to-one transformation of this), which is the trivial sufficient statistic. In other words, in this problem, there is no more concise summary of the observations than the observations themselves.

4. 4 points. Kay I: 5.13. Do not assume the sufficient statistic is complete; prove it.

Solution: The joint pdf of the i.i.d. observations parameterized by θ is

$$\begin{aligned} p_Y(y\,;\,\theta) &= \exp\{-N(\bar{y}-\theta)\}\,u(\min y_n - \theta) \\ &= \underbrace{\exp\{N\theta\}\,u(\min y_n - \theta)}_{g_\theta(T(y))}\,\underbrace{\exp\{-N\bar{y}\}}_{h(y)} \end{aligned}$$

where $u(\cdot)$ is the usual unit step function and $\bar{y} := \frac{1}{N}\sum_{n=0}^{N-1} y_n$ is the usual sample mean. Hence $T(y) = \min y_n$ is a sufficient statistic. Unfortunately, this does not appear to fall into the form necessary to be an exponential family, so we are going to have to prove this sufficient statistic is complete by using the definition. First, we need a pdf for the sufficient statistic

$$Z = T(Y) = \min Y_n.$$

You can derive (or find in a textbook) that the minimum of N i.i.d. exponential random variables with rate parameter λ is also an exponentially distributed random variable with rate parameter $N\lambda$. Hence, since we are dealing here with shifted exponential distributions, we can say that

$$p_Z(z\,;\,\theta) = N\exp\{-N(z-\theta)\}\,u(z-\theta).$$

For this family of pdfs to be incomplete, we need to find a non-zero function f such that

$$s(\theta) = \int_{-\infty}^{\infty} f(z) N\exp\{-N(z-\theta)\}\,u(z-\theta)\,dz = 0$$

for all $\theta \in \mathbb{R}$. We recognize that $s(\theta)$ is the convolution of the function $f(z)$ with the function $g(z) = \exp(Nz)u(-z)$. Convolution in "time" domain is multiplication in frequency domain. You can derive (or find in your undergraduate signals textbook) the Fourier transform pair

$$g(z) = e^{az}u(-z) \qquad \leftrightarrow \qquad G(\omega) = \frac{1}{a - j\omega}$$

for $a > 0$. Hence, we can transform this problem into frequency domain and write

$$S(\omega) = \mathcal{F}(s(\theta)) = F(\omega)\cdot\frac{1}{N - j\omega}.$$

Clearly, there is no non-zero $F(\omega)$ that causes this to be equal to zero for all ω since $\frac{1}{N-j\omega}$ is non-zero for all ω. Hence the sufficient statistic $Z = T(Y) = \min Y_n$ is complete.

Now we need to find an unbiased estimator to proceed with the RBLS theorem. Is $\hat{\theta}(T(y))$ an unbiased estimator? The mean of $Z = T(Y) = \min Y_n$ can be computed as

$$\mathrm{E}_\theta(Z) = \theta + \frac{1}{N}$$

where we used the fact that Z is a shifted exponentially distributed random variable. So, even though $\hat{\theta}(T(y))$ is biased, we can make an unbiased estimator by writing

$$\hat{g}(y) = T(y) - \frac{1}{N} = \min y_n - \frac{1}{N}.$$

This is a perfectly valid unbiased estimator because you know $\{y_0, \ldots, y_{N-1}\}$ and you know N.

The last step is then to compute the conditional expectation where we condition on $T(Y) = T(y) = \min y_n$. Specifically,

$$\begin{aligned} \tilde{g}[T(y)] &= \mathrm{E}_\theta[\hat{g}(Y)\,|\,T(Y) = T(y)] \\ &= \mathrm{E}_\theta\left[T(Y) - \frac{1}{N}\,|\,T(Y) = \min y_n\right] \\ &= \min y_n - \frac{1}{N} \end{aligned}$$

which is the same as our earlier unbiased estimator $\hat{g}(y)$. According to the RBLS theorem, this is then the MVU estimator of the non-random parameter θ.

5. 4 points. Kay I: 5.17. Do not assume the sufficient statistic is complete; prove it.

Solution to part (a): Let's first write the joint pdf on the observations, parameterized by $\theta = A$ as

$$\begin{aligned} p_Y(y\,;\,A) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{\frac{-1}{2\sigma^2}\sum_{n=0}^{N-1}(y_n - A\cos(2\pi f_0 n))^2\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{\frac{-1}{2\sigma^2}y^\top y\right\}\exp\left\{\frac{1}{2\sigma^2}2y^\top z(A)\right\}\exp\left\{\frac{-1}{2\sigma^2}z(A)^\top z(A)\right\} \\ &= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{\frac{-1}{2\sigma^2}z(A)^\top z(A)\right\}}_{a(\theta)}\exp\left\{\frac{1}{2\sigma^2}2y^\top z(A)\right\}\underbrace{\exp\left\{\frac{-1}{2\sigma^2}y^\top y\right\}}_{h(y)} \end{aligned}$$

where $z(A) := [A, A\cos(2\pi f_0), \ldots, A\cos(2\pi f_0(N-1))]^\top$. To confirm that this is in the "exponential family" form, we need to look at the term inside the middle exponential. We can write

$$\begin{aligned} \frac{1}{2\sigma^2}2y^\top z(A) &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1} y_n A\cos(2\pi f_0 n) \\ &= A\frac{1}{\sigma^2}\sum_{n=0}^{N-1} y_n \cos(2\pi f_0 n) \\ &= AT(y). \end{aligned}$$

In other words, the sufficient statistic $T(y)$ for the unknown parameter A is an inner product of the observations with a cosine waveform at the known frequency f_0. Moreover, we have confirmed that this is an exponential family, hence $T(y)$ is sufficient and complete.

Now we need to find an unbiased estimator to proceed with the RBLS theorem. Is $\hat{\theta}(T(y))$ an unbiased estimator? Let's check:

$$\begin{aligned} \mathrm{E}_\theta(T(Y)) &= \mathrm{E}_\theta\left[A\frac{1}{\sigma^2}\sum_{n=0}^{N-1} Y_n \cos(2\pi f_0 n)\right] \\ &= A\frac{1}{\sigma^2}\sum_{n=0}^{N-1} \mathrm{E}_\theta[Y_n]\cos(2\pi f_0 n) \\ &= A\frac{1}{\sigma^2}\sum_{n=0}^{N-1} \cos^2(2\pi f_0 n) \end{aligned}$$

which is clearly biased. But, we can make an unbiased estimator by writing

$$\hat{g}(y) = \frac{T(y)}{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n)} = \frac{\sum_{n=0}^{N-1} y_n\cos(2\pi f_0 n)}{\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n)}.$$

This is a perfectly valid unbiased estimator because you know $\{y_0, \ldots, y_{N-1}\}$, you know σ^2, you know f_0, and you know N.

The last step is then to compute the conditional expectation where we condition on $T(Y) = T(y) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} y_n\cos(2\pi f_0 n)$. Specifically,

$$\begin{aligned} \tilde{g}[T(y)] &= \mathrm{E}_\theta[\hat{g}(Y)\,|\,T(Y) = T(y)] \\ &= \mathrm{E}_\theta\left[\frac{T(Y)}{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n)}\,|\,T(Y) = T(y)\right] \\ &= \frac{T(y)}{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n)} \\ &= \frac{\sum_{n=0}^{N-1} y_n\cos(2\pi f_0 n)}{\sum_{n=0}^{N-1}\cos^2(2\pi f_0 n)} \end{aligned}$$

which is the same as our earlier unbiased estimator $\hat{g}(y)$. According to the RBLS theorem, this is then the MVU estimator of the non-random parameter A.

6. 4 points. Kay I: 5.18. Try to find a minimal sufficient statistic. Is your sufficient statistic complete?

Solution: The marginal parameterized pdf of a single observation is

$$p_{Y_n}(y_n\,;\,\theta) = \begin{cases} \frac{1}{\theta_2-\theta_1} & \theta_1 < y_n < \theta_2 \\ 0 & \text{otherwise} \end{cases}$$
$$= \frac{1}{\theta_2-\theta_1} u(y_n-\theta_1)u(\theta_2-y_n)$$

where $u(\cdot)$ is the usual unit step function. Since the observations are i.i.d., the joint pdf is

$$p_Y(y\,;\,\theta) = \frac{1}{(\theta_2-\theta_1)^N}\prod_n u(y_n-\theta_1)u(\theta_2-y_n)$$
$$= \underbrace{\frac{1}{(\theta_2-\theta_1)^N} u(\min y_n-\theta_1)u(\theta_2-\max y_n)}_{g_\theta(T(y))}$$

where $h(y)=1$ in the Neyman-Fisher factorization theorem. This problem has two parameters, hence $T(y) = [\min y_n, \max y_n]^\top$ is the minimal sufficient statistic. You can confirm this is sufficient by computing the pdf conditioned on $T(Y)=t=[t_1,t_2]^top$

$$p_Y(y\,|\,T(Y)=t\,;\,\theta) = \delta(\min y_n - t_1)\delta(\max y_n - t_2)$$

since all the randomness is removed from the joint pdf when we condition on $T(Y)=t$.