ENEE 621 Spring 2016 Homework 8 Solutions

1. See problem 1 of homework 7.

2.a) CR1 is equivalent to that Θ is an open set in R^p

CR2a is equivalent to that $c(\theta)q(y)e^{Q(\theta)'k(y)}$ is absolutely continuous in R^k

CR2b is equivalent to $f_\theta(y)>0$ for $y \in R^k$, which is also equivalent to $q(y)>0$ for $y \in R^k$. It makes no difference for $\theta \in \Theta$, thus CR2b holds

CR3 is equivalent to the condition that the mappings c and Q are differentiable on Θ.

To identify CR4, we notice that

$$\frac{\partial}{\partial \theta_i} \ln f_\theta(y) = \frac{\partial}{\partial \theta_i} \ln c(\theta) + \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)^T k(y),$$

$$i=1 \cdots p, \quad \theta \in \Theta.$$

thus, CR4 holds if $E_\theta[|k_l(y)|^2] < \infty$, $l=1, \cdots, d$, $\theta \in \Theta$

CR5 is equivalent to the condition that

$$0 = C(\theta) \int_S \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)^T K(y) \cdot \exp[Q(\theta)^T K(y)]\, q(y)\, dF(y)$$

$$+ \frac{\partial}{\partial \theta_i} C(\theta) \cdot \int_S \exp[Q(\theta)^T K(y)]\, q(y)\, dF(y), \quad i = 1, \cdots, p,$$

which is farther equivalent to

$$0 = \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)^T E_\theta[K(y)] + \frac{1}{C(\theta)} \frac{\partial}{\partial \theta_i} C(\theta), \quad i = 1 \cdots p,$$

$$\Leftrightarrow \quad \frac{\partial}{\partial \theta_i} \ln C(\theta) = -\left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)^T E_\theta[K(y)]$$

2.b) Let $M_{ij}(\theta)$ denote the element at the position (i, j),

$$\cancel{M_{ij}(\theta) = E_\theta[(\frac{\partial}{\partial \theta_i} Q(\theta))^T}$$

using the conclusion of 2.a),

$$\frac{\partial}{\partial \theta_i} \log f_\theta(y) = \frac{\partial}{\partial \theta_i} \log C(\theta) + \frac{\partial}{\partial \theta_i} Q(\theta)' K(y)$$

$$\Rightarrow \frac{\partial}{\partial \theta_i} \log f_\theta(y) = \frac{\partial}{\partial \theta_i} Q(\theta)' [K(y) - E_\theta[K(y)]]$$

Thus, $M_{ij}(\theta) = E_\theta\left[\frac{\partial}{\partial\theta_i} Q(\theta)' [K(y) - E_\theta[K(y)]][K(y) - E_\theta[K(y)]]' \frac{\partial}{\partial\theta_i} Q(\theta)\right]$

$$= \frac{\partial}{\partial\theta_i} Q(\theta)' \, cov_\theta[K(y)] \cdot \frac{\partial}{\partial\theta_i} Q(\theta)$$

c) $E_\theta\left[g(y) \left(\frac{\partial}{\partial\theta_i} \log f_\theta(y)\right)\right]$

$$= \int_S g(y) \frac{\partial}{\partial\theta_i} \left(C(\theta) \exp[Q(\theta)^T K(y)]\right) q(y) dF(y)$$

$$= \int_S g(y) \left(\frac{1}{C(\theta)} \frac{\partial}{\partial\theta_i} C(\theta) + \frac{\partial}{\partial\theta_i} Q(\theta)' K(y) \right) f_\theta(y) dF(y)$$

$$= \frac{\partial}{\partial\theta_i} \log C(\theta) \, E_\theta[g(y)] + E_\theta\left[g(y) \frac{\partial}{\partial\theta_i} Q(\theta)' K(y)\right]$$

$$= E_\theta\left[(g(y) - E_\theta[g(y)]) \frac{\partial}{\partial\theta_i} Q(\theta)' K(y)\right], \quad i = 1 \cdots p, \quad \theta \in \Theta$$

The regularity condition is equivalent to

$$\frac{\partial}{\partial\theta_i} E[g(y)] = E_\theta\left[(g(y) - E_\theta[g(y)]) \frac{\partial}{\partial\theta_i} Q(\theta)' K(y)\right],$$

$$i = 1 \cdots p, \quad \theta \in \Theta$$

3. See problem 7 in homework 7.

4. The Poisson distribution is

$$p(Y=y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

With y_1 and y_2, we have

$$p_\theta(y) = \frac{\lambda^{y_1} e^{-\lambda}}{y_1!} \cdot \frac{\lambda^{y_2} e^{-\lambda}}{y_2!} = \frac{\lambda^{(y_1+y_2)} e^{-2\lambda}}{y_1! \, y_2!}$$

Let $\theta = e^{-\lambda}$, then $\lambda = -\log\theta$

$$p_\theta(y) = \frac{(-\log\theta)^{y_1+y_2} \, \theta^2}{y_1! \, y_2!}$$

$$= \theta^2 \exp\{\log(-\log\theta) \cdot (y_1+y_2)\} \cdot \frac{1}{y_1! \, y_2!}$$

By proposition IV.C-3 in the textbook,

y_1+y_2 is a complete sufficient statistic for θ.

5. a) From $F_\theta(y) = [F(y)]^{\frac{1}{\theta}}$, we can have

$$p_\theta(y) = \exp\{\sum_{k=1}^{n} \log F(y_k)/\theta\},$$

which means that $\sum_{k=1}^{n} \log F(y_k)$ is a complete sufficient statistic for θ by the completeness theorem for exponential families.

$$E[\sum_{k=1}^{n} \log F(y_k)] = \frac{n}{\theta}\int_{-\infty}^{0} x e^{\frac{x}{\theta}}\,dx = -n\theta$$

Thus we have $E[\hat{\theta}_{MV}(y)] = \theta$

b) $$p(\theta|y) = c^m \exp(-((c - \sum_{k=1}^{n} \log F(y_k))/\theta)/(\Gamma(m)\theta^{m+1}), \quad \theta>0.$$

$$E[\theta|y] = \frac{(c - \sum_{k=1}^{n} \log F(y_k))}{m+n-1} = \hat{\theta}_{MMSE}(y)$$

c) $\hat{\theta}_{MV}(y)$ does not depend on the prior distribution of θ, in other words, no prior information is used.

For $\hat{\theta}_{MMSE}(y)$, if n is small, $E[\theta|y] \approx \frac{c}{m-1} = E[\theta]$, which is the estimate from prior information. With more observations available, n increases, and the weight of prior information decreases.

6. Since $E_\theta[\nabla_\theta \log f_\theta(Y)]=0$, it is easy to verify that

$$E_\theta[U(\theta,Y)]=0.$$

$$\begin{aligned}
\mathrm{cov}_\theta[U(\theta,Y)] &= E[U(\theta,Y)'U(\theta,Y)] \\
&= E[(g(Y)-\theta)'(g(Y)-\theta)] - b_\theta(g)b_\theta(g)' \\
&\quad - (I_p+\nabla_\theta b_\theta(g))M(\theta)M(\theta)^{-2}(I_p+\nabla_\theta b_\theta(g))' \\
&= \Sigma_\theta(g) - b_\theta(g)b_\theta(g)' - (I_p+\nabla_\theta b_\theta(g))M(\theta)^{-1}(I_p+\nabla_\theta b_\theta(g))'.
\end{aligned}$$

It is easy to see that $\mathrm{cov}_\theta(U(\theta,Y))\geq 0$

$$\Rightarrow \Sigma_\theta(g) \geq b_\theta(g)b_\theta(g)' + (I_p+\nabla_\theta b_\theta(g))M(\theta)^{-1}(I_p+\nabla_\theta b_\theta(g))',$$

which is the claim of Theorem 5.1 in the lecture notes.

In Theorem 5.1, ~~if~~ the equality holds iff

$$g(Y)-\theta = b_\theta(g) + K(\theta)\nabla_\theta \log f_\theta(Y) \qquad F\text{-a.e.}$$

where $K(\theta)=(I_p+\nabla_\theta b_\theta(g))M(\theta)^{-1}$, which means $U(\theta,Y)=0$.

7\. i) $f_\theta(y_k) = \begin{cases} \theta e^{-\theta y_k} & \text{if } y_k \geq 0 \\ 0 & \text{otherwise} \end{cases}$

$f_\theta(y) = \begin{cases} \theta^n e^{-\theta \sum_{k=1}^{n} y_k} & \text{if } \min(y_1 \cdots y_n) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

$$\log f_\theta(y) = n \log\theta - \theta \sum_{k=1}^{n} y_k$$

$$\frac{\partial}{\partial\theta} \log f_\theta(y) = \frac{n}{\theta} - \sum_{k=1}^{n} y_k$$

$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(y) = -\frac{n}{\theta^2}$$

The Fisher information matrix $M^{(1)}(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2} \log f_\theta(y)\right]$

$$= \cancel{\frac{d^2}{n}} \frac{n}{\theta^2}$$

ii) The MVUE is the estimator that achieves unbiased estimation with lowest variance. therefore we just investigate whether the variance is greater than the CRLB.

It's easy to see that $\sum_{i=1}^{n} y_k$ is a complete sufficient statistic for θ, and

$$E\left[\frac{n-1}{n} \cdot \frac{n}{\sum_{i=1}^{n} y_i}\right] = \theta \cancel{=\theta}$$

and $\dfrac{n-1}{\sum_{i=1}^{n} y_i}$ is the MVUE

$Var[\hat{\theta}_{MV}(y)] = \dfrac{\theta^2}{n-2}$, while the CRLB is $\dfrac{\theta^2}{n}$,

therefore, there is no efficient estimator for θ.

iii) Since $\dfrac{\partial}{\partial\theta} \log f_\theta(y) = \dfrac{n}{\theta} - \sum_{i=1}^{n} y_i$

$$\hat{\theta}_{ML}(y) = \frac{n}{\sum_{i=1}^{n} y_i}$$

iv) $E_\theta[\hat{\theta}_{ML}(y)] = \dfrac{n\theta}{n-1}$, it is biased

$\lim_{n\to\infty} E_\theta[\hat{\theta}_{ML}(y)] = \theta$, it is asymptotically unbiased.

Since $\frac{1}{\hat{\theta}_{ML}(y)} = \frac{\sum_{i=1}^{n} y_n}{n}$, by the law of large numbers,

with $n \to +\infty$, it converges to $\frac{1}{\theta}$ almost surely, thus $\hat{\theta}_{ML}(y)$ is consistent.

Since $\hat{\theta}_{ML}(y)$ can not be written in the form of sum of i.i.d random variables, it is not asymptotically normal.

8. i) $\log f_\theta(y) = -\log\pi - \log(1+(y-\theta)^2)$

$$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{2(y-\theta)}{1+(y-\theta)^2}$$

$$\frac{\partial^2}{\partial\theta^2}\log f_\theta(y) = -2\frac{1-(y-\theta)^2}{(1+(y-\theta)^2)^2}$$

thus, $I(\theta) = E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f_\theta(y)\right]$

$$= 2\int_{-\infty}^{\infty}\frac{1-(y-\theta)^2}{(1+(y-\theta^2))^2}\cdot\frac{1}{\pi}\cdot\frac{1}{1+(y-\theta)^2}\,dy$$

$$= \frac{2}{\pi}\int_{-\infty}^{\infty}\frac{1-x^2}{(1+x^2)^3}\,dx$$

$$= \frac{1}{2} = M^{(1)}(\theta)$$

ii) If there exists an effiecient estimator, it satisfies

$$g(y)-\theta = M(\theta)^{-1}\frac{\partial}{\partial\theta}\log f_\theta(y) \quad \text{when } n=1$$

In this problem, $M(\theta)=\frac{1}{2}$, $\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{2(y-\theta)}{1+(y-\theta)^2}$

Then we obtain

$$g(y) - \theta = \frac{2 \cdot 2(y-\theta)}{1+(y-\theta)^2}$$

$$g(y) = \theta + \frac{4(y-\theta)}{1+(y-\theta)^2}$$

which can not be reduced to the form independent of θ.
In this problem, there is no efficient estimator.

iii) While there is no closed form solution of the ML estimator with $n>1$, we obtain the ML estimator with $n=1$.

From $\frac{\partial}{\partial \theta} \log f_\theta(y) = \frac{2(y-\theta)}{1+(y-\theta)^2}$,

it is easy to see that

$$\hat{\theta}_{ML}(y) = y$$

For $n>1$, it has to be solved by numerical methods.

iv) $E_\theta[\hat{\theta}_{ML}(y)] = E_\theta(y) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{1+(y-\theta)^2} dy$

$$= \frac{1}{\pi} \cdot \pi\theta = \theta,$$

it is unbiased ~~as well as asymptotically unbiased.~~

9. i) $f_\theta(y) = \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-\frac{y^2}{2\theta}\right\}$

$\log f_\theta(y) = -\frac{y^2}{2\theta} - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\theta$

$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{y^2}{2\theta^2} - \frac{1}{2\theta}$

$\frac{\partial}{\partial\theta^2}\log f_\theta(y) = -\frac{y^2}{\theta^3} + \frac{1}{2\theta^2}$

$M^{(1)}(\theta) = -E\left[\frac{\partial}{\partial\theta^2}\log f_\theta(y)\right] = -E\left[-\frac{y^2}{\theta^3} + \frac{1}{2\theta^2}\right] = \frac{1}{2\theta^2}$

With n observations, $M^{(1)}(\theta) = \frac{n}{2\theta^2}$

ii) With n independent observations,

$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{\Sigma y_i^2}{2\theta^2} - \frac{n}{2\theta}$

By lemma 5.2 in the lecture notes, we get

$g(y) - \theta = M(\theta)^{-1} \nabla_\theta \log f_\theta(y)$

$g(y) - \theta = \frac{\Sigma y_i^2}{n} - \theta$

$g(y) = \frac{\Sigma y_i^2}{n}$

iii) $\frac{\partial}{\partial\theta} \log f_\theta(y) = 0 \iff$

$$\frac{\sum y_i^2 - n\theta}{2\theta^2} = 0$$

$\hat{\theta}_{ML}(y) = \frac{\sum y_i^2}{n}$, which is the same as the efficient estimator.

iv) Since for any r.v X, $E[X^2] = Var(X) + E[X]^2$,

$E[y_i^2] = \theta$, and $E[\frac{\sum y_i^2}{n}] = \theta$.

Therefore, the estimator is unbiased, as well as asymptotically unbiased.

By the Law of large numbers,

$\lim_{n\to\infty} P_\theta[|\hat{\theta}_{ML}(y) - \theta| > \varepsilon] = 0$,

and $\lim_{n\to\infty} \hat{\theta}_{ML}(y) = \theta$ a.s.,

it is both strong and weak consistent.

By Theorem 9.3 in the lecture notes, it is also asymptotically normal.

10. a) The log-likelihood function is

$$\log p_\theta(y) = -\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left[y_k - A\sin\left(\frac{k\pi}{2}+\phi\right)\right]^2 - \frac{n}{2}\log(2\pi\sigma^2)$$

The likelihood equations are

$$\sum_{k=1}^{n}\left[y_k - \hat{A}\sin\left(\frac{k\pi}{2}+\hat{\phi}\right)\right]\sin\left(\frac{k\pi}{2}+\hat{\phi}\right)=0$$

and

$$\hat{A}\sum_{k=1}^{n}\left[y_k - \hat{A}\sin\left(\frac{k\pi}{2}+\hat{\phi}\right)\right]\cos\left(\frac{k\pi}{2}+\hat{\phi}\right)=0$$

The ML estimates can be expressed as

$$\hat{A}_{ML}(y) = \sqrt{y_c^2 + y_s^2}$$

$$\hat{\phi}_{ML}(y) = \tan^{-1}\left(\frac{y_c}{y_s}\right)$$

where $y_c = \frac{1}{n}\sum_{k=1}^{n} y_k \cos\left(\frac{k\pi}{2}\right)$

$y_s = \frac{1}{n}\sum_{k=1}^{n} y_k \sin\left(\frac{k\pi}{2}\right)$

b) Since the prior of ϕ does not contain any information,

$$\vec{\phi}_{MAP} = \vec{\phi}_{ML}.$$

Using the prior of A, the MAP estimate can be obtained as

$$\hat{A}_{MAP} = \frac{\hat{A}_{ML} + \sqrt{\left(\frac{\alpha}{n}\right)^2 + \frac{2(1+\alpha)\sigma^2}{n}}}{1+\alpha} \qquad (1)$$

where $\alpha = \frac{2\sigma^2}{n\beta^2}$

c) $\vec{\phi}_{MAP} = \vec{\phi}_{ML}$ since the prior contains no information.

Similarly, if $\beta \to \infty$, the prior of A is more and more flat. By (1), $\alpha \to 0$ and $\hat{A}_{MAP} = \hat{A}_{ML}$.