14.381 Solutions Problem Set 9
Statistics Fall, 2004

TA: José Tessada
(tessada@mit.edu)

1. C&B 7.38

   In both cases you can use Corollary 7.3.15. It is easy to check that the conditions for the Cramer-Rao inequality hold for both pdfs. In order to apply the Corollary we need to be able to write $\frac{\partial}{\partial\theta}\log L(\theta|\mathtt{x}) = a(\theta)[W(\mathtt{x}) - \tau(\theta)]$, where $a(\theta)$ is a function of θ only, it CANNOT be a function of the data as some of you initially thought.

   (a)

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(\theta|\mathtt{x}) &= \frac{\partial}{\partial\theta}\log\prod_i \theta x_i^{\theta-1} = \frac{\partial}{\partial\theta}\sum_i\left[\log\theta + (\theta-1)\log x_i\right] \\ &= \sum_i\left[\frac{1}{\theta} + \log x_i\right] = -n\left[-\sum_i\frac{\log x_i}{n} - \frac{1}{\theta}\right]. \qquad (1)\end{aligned}$$

   According to (1), the sample average of $\log x_i$ is the UMVU estimator of $g(\theta) = \frac{1}{\theta}$.

   (b)

$$\begin{aligned}\frac{\partial}{\partial\theta}\log L(\theta|\mathtt{x}) &= \frac{\partial}{\partial\theta}\log\prod_i\frac{\log\theta}{\theta-1}\theta^{x_i} = \frac{\partial}{\partial\theta}\sum_i\left[\log\log\theta - \log(\theta-1) + x_i\log\theta\right] \\ &= \sum_i\left(\frac{1}{\theta\log\theta} - \frac{1}{\theta-1}\right) + \frac{1}{\theta}\sum_i x_i = \frac{n}{\theta\log\theta} - \frac{n}{\theta-1} + \frac{n\overline{x}}{\theta} \\ &= \frac{n}{\theta}\left[\overline{x} - \left(\frac{\theta}{\theta-1} - \frac{1}{\log\theta}\right)\right]. \qquad (2)\end{aligned}$$

   Using the same argument as in part (a), we can say that the sample average of x is the UMVU estiamtor of $g(\theta) = \left(\frac{\theta}{\theta-1} - \frac{1}{\log\theta}\right)$.

   One additional comment, as all the conditions for the Cramer-Rao inequality hold we can safely assume that the estimators we just found are unbiased, a key element in the statement UMVU. For that we should check that the conditions stated in Theorem 7.3.9 hold. That is a homework.

2. C&B 7.48: This is not a hard question, you just need to be careful with the derivatives. The expectation and the variance of $\widehat{p}_{ML}$ are easy to compute.

   Recall that $\widehat{p}_{ML} = \overline{X}$, so $E(\widehat{p}_{ML}) = p$ and $Var(\widehat{p}_{ML}) = \frac{p(1-p)}{n}$. Now, compute the Cramer-Rao lower bound. As we have a random sample we can use the following formula (Corollary 7.3.10):

$$\begin{aligned}\frac{\left(\frac{d}{d\theta}E_\theta(W(\mathtt{X}))\right)^2}{nE_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)} &= \frac{\left(\frac{d}{d\theta}E_\theta(W(\mathtt{X}))\right)^2}{-nE_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)} = \frac{\left(\frac{d}{dp}p\right)^2}{-nE\left(\frac{d^2}{dp^2}\log L(p|X)\right)} \\ &= \frac{1}{-nE\left[\frac{d^2}{dp^2}\log\left[p^x(1-p)^{(1-x)}\right]\right]} = \frac{1}{-nE\left[-\frac{x}{p^2} - \frac{(1-x)}{(1-p)^2}\right]} \\ &= \frac{p(1-p)}{n}\end{aligned}$$

   Thus, $\widehat{p}_{ML}$ achieves the Cramer-Rao lower bound and it is the UMVU estimator.

   There is an alternative way to do this problem. You can proceed as in the previous question and, assuming the conditions stated in Corollary 7.3.15 hold, show that effectively $\widehat{p}_{ML}$ achieves the C-R lower bound. In that way you do not need to deal with all the derivatives and can claim that $Var(\widehat{p}_{ML})$ is the C-R lower bound.

Be careful, because in order to apply Corollary 7.3.15, or any other property related to the C-R lower bound, you must be sure that the conditions for the C-R lower bound must hold, or else you will compute a useless bound (which is not a bound actually), see example with a uniform distribution in the book.

3. C&B 7.50

   (a) This is easy, just take expectation of the "new" estimator and you get that it is unbiased. $E\left(a\overline{X} + (1-a)\, cS\right) = \theta$.

   (b) Do not try to deal with c using the formula the book states, just treat it as a constant and keep in mind the $E(cS) = \theta$. Define $\widetilde{\theta} = a\overline{X} + (1-a)\, cS$, we can compute $Var\left(\widetilde{\theta}\right)$ using $Var\left(\overline{X}\right)$ and $Var(cS)$.

   Given that $\overline{X}$ and S are independent, then $Var\left(\widetilde{\theta}\right) = a^2 Var\left(\overline{X}\right) + (1-a)^2 Var(cS)$. So we need to find a^*, where

$$\begin{aligned} a^* &= \arg\min\left\{Var\left(\widetilde{\theta}\right)\right\} \\ &= \frac{Var(cS)}{Var\left(\overline{X}\right) + Var(cS)} \end{aligned} \tag{3}$$

   For those of you who have worked a bit more in econometrics or decision theory, this formula is basically the standard solution of a signal extraction problem.
   We know that $Var\left(\overline{X}\right) = \frac{\theta^2}{n}$ while $Var(cS)$ can be computed as follows:

$$Var(cS) = E\left[(cS)^2\right] - E[cS]^2 = \left(c^2 - 1\right)\theta^2$$

   Substituting the previous results in (3) we can rewrite a^* as:

$$a^* = \frac{\left(c^2-1\right)}{\left(\frac{1}{n} + c^2 - 1\right)}$$

   (c) Here you just need to do build a function of $\lambda = \left(\overline{X}, S\right)$ such that it has expectation equal to 0, but it is not equal to 0 with probability 1. The simplest case: $g(\lambda) = \left[cS - \overline{X}\right]$.

4. C&B 7.52

   (a) You do not need to prove anything in this case. Claim that the Poisson is an exponential family, use the corresponding theorem and you have proved that $\sum_i X_i$ is a complete sufficient statistic. Now, $\overline{X}$ is a function of $\sum_i X_i$ and it is an unbiased estimator of λ, then it must be the best unbiased estimator of λ.

   (b) First of all, we know that $\sum_i X_i$ is a complete sufficient statistic, so $\overline{X}$ also is a complete sufficient statistic (this is easy to see if you write the pdf of a Poisson in the "exponential family form" and see that you can rewrite it in a way such that $t(x) = x/n$).
   From part (a) we know that $\overline{X}$ is the best unbiased estimator. Now, S^2 is an unbiased estimator, we want to use the following theorem from the book:

   Theorem 1 *Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.*

   Before that, by *Rao-Blackwell* we know that $\psi\left(\overline{X}\right) = E\left(S^2 | \overline{X}\right) = \lambda$, as we are conditioning an unbiased estimator on a sufficient statistic (which happens to be complete too). So, by Theorem

1, we have that $\psi\left(\overline{X}\right)$ is the unique best unbiased estimator of λ, so then given our result in (a), it must be that $\psi\left(\overline{X}\right) = E\left(S^2|\overline{X}\right) = \overline{X}$.

With this result in hand the next part is trivial. Use the variance decomposition formula to write

$$\begin{aligned} Var\left(S^2\right) &= Var\left(E\left(S^2|\overline{X}\right)\right) + E\left(Var\left(S^2|\overline{X}\right)\right) \\ &= Var\left(\overline{X}\right) + E\left(Var\left(S^2|\overline{X}\right)\right) \\ &> Var\left(\overline{X}\right). \end{aligned}$$

So we stablished that $\overline{X}$ is more efficient than S^2.

(c) *Let $T(\mathtt{X})$ be a complete sufficient statistic, and let $\widetilde{T}(\mathtt{X})$ be any statistic other than $T(\mathtt{X})$ such that $E(T(\mathtt{X})) = E\left(\widetilde{T}(\mathtt{X})\right)$. Then $E\left(\widetilde{T}(\mathtt{X})|T(\mathtt{X})\right) = T(\mathtt{X})$ and $Var\left(\widetilde{T}(\mathtt{X})\right) > Var(T(\mathtt{X}))$.*