UC Berkeley
Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set 3- Solutions
Fall 2014

Issued: Friday, September 19
Due: Thursday, September 25

Problem 3.1

(a) First we find a sufficient complete statistic of $g(\theta)$. The joint distribution of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n; \theta) = \frac{1}{(2\pi)^{-n/2}} e^{-\sum_{i=1}^n x_i^2/2} 1\{\theta > \max_{i=1,\ldots,n} x_i\}.$$

By factorization theorem, $T := \max_{i=1,\ldots,n} X_i$ is a sufficient statistic of θ. The density function of T satisfies

$$P(T \leq t) = (\Phi(t)/\Phi(\theta))^n \Rightarrow f_T(t) = n\phi(t)\Phi(t)^{n-1}/\Phi(\theta)^n.$$

To show completeness, let $h(t)$ be any function such that $\mathbb{E}_\theta h(T) = c$ for some constant c and all $\theta \in \mathbb{R}$, we have

$$\int_{-\infty}^{\theta} h(t) \cdot n\phi(t)\Phi(t)^{n-1} dt = c \cdot \Phi(\theta)^n.$$

Taking derivative with respect to θ on both side, we have

$$h(\theta) \cdot n\phi(\theta)\Phi(\theta)^n = c \cdot n\phi(\theta)\Phi(\theta)^n \Rightarrow h(\theta) = c, \forall \theta.$$

Therefore T is a sufficient complete statistic of θ.

Now assume $\delta(T)$ is an unbiased estimator of $g(\theta)$, we have $\mathbb{E}_\theta(\delta(T)) = g(\theta)$, therefore

$$\int_{-\infty}^{\theta} \delta(t) \cdot n\phi(t)\Phi(t)^{n-1} dt = g(\theta)\Phi(\theta)^n.$$

Again taking derivative on both sides and moving terms, we have

$$\delta(\theta) = g(\theta) + g'(\theta)\frac{\Phi(\theta)}{n\phi(\theta)}, \forall \theta.$$

Finally since $\mathbb{E}(\delta(T)|T) = \delta(T)$, the UMVU we are looking for is

$$\delta(T) = g(T) + g'(T)\frac{\Phi(T)}{n\phi(T)}, \text{ where } T = \max_{i=1,\ldots,n} X_i.$$

(b) Take $g(\theta) = \theta^2$, the UMVU we are looking is $\delta(T) = T^2 + 2T\Phi(T)/3\phi(T) = 0$.

Problem 3.2

(a) Letting $g(x;\theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right)$, we know that:

$$\log(g(x;\theta)) = -\log(\theta) + \log f\left(\frac{x}{\theta}\right)$$

Differentiating with respect to θ:

$$\frac{\partial}{\partial\theta}\log(g(x;\theta)) = -\frac{1}{\theta}\left[1 + \frac{x}{\theta}\frac{f'\left(\frac{x}{\theta}\right)}{f\left(\frac{x}{\theta}\right)}\right]$$

So:

$$\begin{aligned}
I(\theta) &= \int\left[\frac{\partial}{\partial\theta}\log(g(x;\theta))\right]^2 g(x,\theta)dx \\
&= \frac{1}{\theta^2}\int\left[1 + \frac{x}{\theta}\frac{f'\left(\frac{x}{\theta}\right)}{f\left(\frac{x}{\theta}\right)}\right]^2 \frac{1}{\theta} f\left(\frac{x}{\theta}\right) dx \\
&\overset{\theta dy = dx}{=} \frac{1}{\theta^2}\int\left[1 + \frac{yf'(y)}{f(y)}\right]^2 f(y)dy
\end{aligned}$$

(b) For the parameter $\eta(\theta) = \log(\theta)$, we have from Keener 8.1.8 that:

$$\begin{aligned}
I(\eta) &= \left[\frac{\partial\eta^{-1}(\theta)}{\partial\theta}\right]^2 I(\theta) \\
&= \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]^{-2} I(\theta) \\
&= \int\left[1 + y\frac{f'(y)}{f(y)}\right]^2 f(y)dy
\end{aligned}$$

(c) For the Cauchy distribution $C(0,\theta)$:

$$p(x;\theta) = \frac{1}{\pi}\frac{\theta}{x^2+\theta^2} = \frac{1}{\theta} f\left(\frac{x}{\theta}\right)$$

where

$$f(y) = \frac{1}{\pi}\frac{1}{1+y^2}$$

so, we can apply the result from (a):

$$f'(y) = -\frac{1}{\pi}\frac{2y}{(1+y^2)^2}$$

and as a result:

$$1 - \frac{yf'(y)}{f(y)} = 1 - \frac{2y^2}{1+y^2} \quad = \quad \frac{1-y^2}{1+y^2}$$

Finally, from (a):

$$I(\theta) \quad = \quad \frac{1}{\pi\theta^2}\int \frac{1-y^2}{(1+y^2)^2}dy$$

and transforming $y = \tan(t)$:

$$\begin{aligned} I(\theta) &= \frac{1}{\pi\theta^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2(2t)dt \\ &= \frac{1}{2\theta^2} \end{aligned}$$

Problem 3.3

(a)

$$\begin{aligned} p(Y_0 = y_o, \ldots, Y_n = y_n) &= p(Y_0 = y_0) \cdot p(Y_1 = y_1|Y_0 = y_0) \cdots p(Y_n = y_n|Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) \\ &= p(Y_0 = y_0) \cdot p(Y_1 = y_1|Y_0 = y_0) \cdots p(Y_n = y_n|Y_{n-1} = y_{n-1}) \\ &= \frac{e^{-\theta}\theta^{y_0}}{y_0!} \cdot \frac{e^{-y_0\theta}(y_0\theta)^{y_1}}{y_1!} \cdots \frac{e^{-y_{n-1}\theta}(y_{n-1}\theta)^{y_n}}{y_n!} \end{aligned}$$

$$\begin{aligned} l(\theta; y_0, \ldots, y_n) &= \log(p(Y_0 = y_o, \ldots, Y_n = y_n)) \\ &= -\theta\sum_{i=0}^{n} y_{i-1} + (\log\theta)\sum_{i=0}^{n} y_i + g(y_0, \ldots, y_n) \ \ \textit{where} \ \ y_{-1} = 1 \\ \frac{\partial l(\theta)}{\partial\theta} &= -\sum_{i=0}^{n} y_{i-1} + \frac{1}{\theta}\sum_{i=0}^{n} y_i \\ \frac{\partial^2 l(\theta)}{\partial\theta^2} &= -\frac{1}{\theta^2}\sum_{i=0}^{n} y_i \leq 0 \end{aligned}$$

Therefore, $\widehat{\theta} = \frac{\sum_{j=0}^{n} Y_j}{\sum_{j=0}^{n} Y_{j-1}} = \frac{\sum_{j=0}^{n} Y_j}{1 + \sum_{j=0}^{n-1} Y_j}$.

(b) Note that $\mathbb{E}(Y_i) = \mathbb{E}(\mathbb{E}(Y_i|Y_{i-1})) = \mathbb{E}(\theta Y_{i-1}) = \theta\mathbb{E}(Y_{i-1})$ for $i = 1, \ldots, n$ and $\mathbb{E}(Y_0) = \theta$. Thus, $\mathbb{E}(Y_i) = \theta^{i+1}$.

$$\therefore I(\theta) \quad = \quad \mathbb{E}\left[-\frac{\partial^2 l(\theta)}{\partial\theta^2}\right] = \mathbb{E}\left[\frac{1}{\theta^2}\sum_{i=0}^{n} Y_i\right] = \frac{1}{\theta^2}(\theta + \theta^2 + \ldots + \theta^{n+1}) = \frac{1-\theta^{n+1}}{\theta(1-\theta)}$$

If $\theta < 1$, $I(\theta) \to \dfrac{1}{\theta(1-\theta)} \geq 4$

Intuitively, as time goes on, the information we can get is decreasing because Y_i depends on the previous Y_{i-1} and $\theta < 1$. The information they carry depends on each other and $\theta < 1$ means, the information we can get is decreasing.

Problem 3.4

(a) Let $V = \delta(X)$ and $W = \dfrac{p(X;\theta')}{p(X;\theta)}$ for $\theta' \geq \theta$. Then

$$\begin{aligned}
\mathbb{E}_\theta(W) &= \int_{x_1,\ldots,x_n \geq \theta}^{\infty} p(x;\theta')dx_1\cdots dx_n = \int_{x_1,\ldots,x_n \geq \theta'}^{\infty} p(x;\theta')dx_1\cdots dx_n = 1 \\
\mathbb{E}_\theta(VW) &= \int_{x_1,\ldots,x_n \geq \theta}^{\infty} \delta(x)p(x;\theta')dx_1\cdots dx_n \\
&= \int_{x_1,\ldots,x_n \geq \theta'}^{\infty} \delta(x)p(x;\theta')dx_1\cdots dx_n \\
&= \mathbb{E}_{\theta'}(\delta(X)) = \theta'
\end{aligned}$$

$$\begin{aligned}
Var_\theta(W) &= \mathbb{E}_\theta\left[(W-1)^2\right] \\
Cov_\theta(V,W) &= \theta' - \theta
\end{aligned}$$

By Cauchy-Schwartz inequality, $Var_\theta(V) \geq \dfrac{Cov_\theta(V,W)^2}{Var_\theta(W)} = \dfrac{(\theta-\theta')^2}{\mathbb{E}_\theta\left[(W-1)^2\right]}$. Because $Var_\theta(V)$ does not depend on θ', the inequality holds.

(b) $\mathbb{E}_\theta\left[\dfrac{p(x;\theta')}{p(x;\theta)}\right]^2 = e^{2n\theta'-n\theta}\left[\int_{\theta'}^{\infty} e^{-x}dx\right]^n = e^{n(\theta'-\theta)}$. (RHS) in (a) is $\sup\limits_{\theta'\geq\theta} \dfrac{(\theta'-\theta)^2}{e^{n(\theta'-\theta)}-1}$

Let $f(x) = \dfrac{x^2}{e^{nx}-1}$ for $x \geq 0$. Then $f(x) \geq 0\ \forall x \geq 0$ and $\lim\limits_{x\to 0} f(x) = \lim\limits_{x\to\infty} f(x) = 0$

Let $h(x) = \log f(x) = 2\log x - \log(e^{nx}-1)$ for $x > 0$. $h'(x) = \dfrac{2}{x} - \dfrac{ne^{nx}}{e^{nx}-1}$ $h''(x) = -\dfrac{2}{x^2} + \dfrac{n^2e^{nx}}{(e^{nx}-1)^2}$. For the α such that $h'(\alpha) = 0$, $h''(\alpha) = -\dfrac{2}{\alpha^2} + \dfrac{2}{\alpha^2}(2-n\alpha) = \dfrac{2}{\alpha^2}(1-n\alpha) < 0$. (Note that $1 < n\alpha < 1.6$) Thus, such α is local maxima and one of α is global maxima, because $f(x) \geq 0\ \forall x \geq 0$ and $\lim\limits_{x\to 0} f(x) = \lim\limits_{x\to\infty} f(x) = 0$. (Actually, such α is unique, except $\alpha = 0$. It will be a good exercise to show why it has unique solution except 0, analytically, not numerically.) Then, $f(\alpha) = \dfrac{\alpha^2}{e^{n\alpha}-1} = \dfrac{\alpha^2(2-n\alpha)}{n\alpha} = \dfrac{\alpha}{n}(2-n\alpha) \geq \dfrac{\alpha}{n^2}$ for $n > 2$. So, $\alpha^* = \alpha$ attains the inequality.

(c) There exists t_0 such that $\dfrac{2}{t_0} - \dfrac{e^{t_0}}{e^{t_0}-1} = 0$. Then $a^* = \dfrac{t_0}{n}$. Then $Var_\theta(\delta(X)) \geq$

$\frac{t_0}{n^3}$. Actually Sharp bound for $Var_\theta(\delta(X))$ is $\frac{t_0^2}{n^2}(2-t_0) = O\left(\frac{1}{n^2}\right)$ which differ from this scaling. The distribution in this problem violate one of regularity conditions (the distribution share common support) to guarantee CR bound. That's why the lower bound is better than $O(1/n)$.

(d) **Direct Calculation**

$\mathbb{P}(X_{(1)} \geq t) = \prod_{i=1}^{n} \mathbb{P}(X_i \geq t) = e^{n(\theta-t)}$. Thus, $f_{X_{(1)}}(t) = ne^{n(\theta-t)}$.

$$
\begin{aligned}
\mathbb{E}(X_{(1)}) &= \int_\theta^\infty ne^{n\theta}te^{-nt}dt = \theta + \frac{1}{n} \\
\mathbb{E}(X_{(1)}^2) &= \theta^2 + \frac{2}{n}\theta + \frac{2}{n^2} \\
\mathbb{E}\left[\left(X_{(1)} - \frac{1}{n}\right)^2\right] &= \theta^2 + \frac{1}{n^2}
\end{aligned}
$$

Thus, $Var(\delta_a(X)) = \frac{1}{n^2}$.

Using exponential spacing

<u>**Fact**</u>

If $Y_1, \ldots, Y_n \sim$ *i.i.d.* $Exp(\lambda)$, then $(n-r+1)(Y_{(r)} - Y_{(r-1)}) \sim$ *i.i.d.* $Exp(\lambda)$ where $X_{(0)} = 0$.

Let $Y_i = X_i - \theta$. Then $Y_1, \ldots, Y_n \sim$ *i.i.d.* $Exp(1)$. By Fact, $nY_{(1)} \sim Exp(1)$.

$$
\begin{aligned}
\mathbb{E}(X_{(1)}) &= \mathbb{E}(Y_{(1)}) + \theta = \frac{1}{n} + \theta \\
Var\left(X_{(1)} - \frac{1}{n}\right) &= Var(X_{(1)}) = Var(Y_{(1)}) = \frac{1}{n^2}
\end{aligned}
$$

Problem 3.5

For the Poisson, we have $p(y,\lambda) = (y!)^{-1}\exp\left[-\lambda + k\log(\lambda)\right]$. Conditional on $Y \geq 0$, we notice that $P(y \geq 1) = 1 - P(Y=0) = 1 - \exp(-\lambda)$ the the probability mass function becomes:

$$
q(y,\lambda) = \mathbb{I}(y \geq 1)(y!)^{-1}\exp\left[-\lambda - \log(1-\exp(\lambda)) + y\log(\lambda)\right]
$$

$$
\begin{aligned}
\frac{\partial}{\partial\lambda}\log(q(y,\lambda)) &= -1 - \frac{e^{-\lambda}}{1-e^{-\lambda}} + \frac{y}{\lambda} \\
\frac{\partial^2}{\partial\lambda^2}\log(q(y,\lambda)) &= -\frac{y}{\lambda^2} + \left[\frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\right]
\end{aligned}
$$

Now, for a single observation (because $q(y, \lambda)$ is full rank exponential family) :

$$\begin{aligned} \mathcal{I}(\lambda) &= -\mathbb{E}\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log(q(y,\lambda))\right] \\ &= \frac{\mathbb{E}_\lambda Y}{\lambda^2} - \left[\frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\right] \end{aligned}$$

We can compute $\mathbb{E}_\lambda Y$ from the cumulant generating function as $\frac{\lambda}{1-e^{-\lambda}}$, so:

$$\begin{aligned} \mathcal{I}(\lambda) &= \frac{1}{\lambda(1-e^{-\lambda})} - \left[\frac{e^{-\lambda}}{(1-e^{-\lambda})^2}\right] \\ &= \frac{1-\exp(-\lambda)-\lambda\exp(-\lambda)}{\lambda(1-e^{-\lambda})^2} \end{aligned}$$

Additionally for n independent samples, we have:

$$\begin{aligned} I(\lambda) &= -\mathbb{E}\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log(\prod_{i=1}^{n} q(y_i,\lambda))\right] \\ &= n\mathcal{I}(\theta) \\ &= \frac{n\,(1-\exp(-\lambda)-\lambda\exp(-\lambda))}{\lambda(1-e^{-\lambda})^2} \end{aligned}$$

For an unbiased estimate of λ, $g(\lambda) = \lambda$, so $g'(\lambda) = 1$ and the information lower bound is then given by

$$\frac{1}{I(\lambda)} = \frac{\lambda(1-e^{-\lambda})^2}{n\,(1-\exp(-\lambda)-\lambda\exp(-\lambda))}$$