UC Berkeley

Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set 2- Solutions

Fall 2014

Issued: Thursday, September 11

Due: Thursday, September 18

Problem 2.1

(a) Let $f_{X,Y}$ be the density of (XY). Then

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left\{-\frac{1}{2(1-\theta^2)}(x^2+y^2) + \frac{\theta}{1-\theta^2}xy\right\}$$

$$\therefore \prod_{i=1}^{n} f_{X_i,Y_i}(x_i,y_i) \quad = \quad \left(\frac{1}{2\pi\sqrt{1-\theta^2}}\right)^n \exp\left\{-\frac{1}{2(1-\theta^2)}\sum_{i=1}^{n}(x_i^2+y_i^2) + \frac{\theta}{1-\theta^2}\sum_{i=1}^{n}x_i y_i\right\}$$

Thus, $\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2, \sum_{i=1}^{n} X_i Y_i\right)$ is a sufficient statistics.
Also minimal statistics since it is 2-dimensional curved exponential family.

(b) $\mathbb{E}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2\right) = n + n = 2n$. Thus, let $g(x,y) = x - 2n \neq 0$.
Then $\mathbb{E}\left[g\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2, \sum_{i=1}^{n} X_i Y_i\right)\right] = 2n - 2n = 0$.
Therefore, $(T_1, T_2) = \left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2, \sum_{i=1}^{n} X_i Y_i\right)$ is not complete.

(c) $Z_1 \sim \chi^2(n)$ and $Z_2 \sim \chi^2(n)$: not depend on θ. Thus, Z_1 and Z_2 are both ancillary. But $Cov(Z_1, Z_2) = \mathbb{E}(Z_1 Z_2) - \mathbb{E}(Z_1)\mathbb{E}(Z_2) = n^2 + 2n\theta^2 - n^2 = 2n\theta^2$: depends on θ. Thus, the distribution of (Z_1, Z_2) depends on θ. Not ancillary.

Problem 2.2

(a) The joint distribution function is:

$$\begin{aligned}
f_{X_1,...,X_n}(x_1,...,x_n) &= \frac{1}{(2\theta)^n}\prod_i \mathbf{1}(-\theta \le x_i \le \theta) \\
&= \frac{1}{(2\theta)^n}\mathbf{1}(-\theta \le \min_i x_i)\mathbf{1}(\max_i x_i \le \theta) \\
&= \frac{1}{(2\theta)^n}\mathbf{1}(\max\{-x_{(1)}, x_{(n)}\} \le \theta) \\
&= \exp\Big\{\ln \mathbf{1}(\max\{-x_{(1)}, x_{(n)}\} \le \theta) - n\ln 2\theta\Big\}.
\end{aligned}$$

By factorization theorem $T = \max\{-X_{(1)}, X_{(n)}\}$ is a sufficient statistics. Consider the ratio $f_{Y_1,...,Y_n}(y_1,...,y_n)/f_{X_1,...,X_n}(x_1,...,x_n)$. It does not depend on θ implies:

$$\frac{\mathbf{1}(\max\{-x_{(1)}, x_{(n)}\} \le \theta)}{\mathbf{1}(\max\{-y_{(1)}, y_{(n)})\} \le \theta)}$$

does not depend on θ, hence $T(X) = T(Y)$. If $T(X) = T(Y)$ then $f_{Y_1,...,Y_n}(y_1,...,y_n)/f_{X_1,...,X_n}(x_1,...,x_n)$ does not depend on θ. Therefore $T(X) = \max\{-X_{(1)}, X_{(n)}\}$ is a minimal sufficient statistics.

(b) For minimal sufficient statistic $T(X)$, we have:

$$P(T(X) \le x) = P(\max\{-X_{(1)}, X_{(n)}\} \le x) = P(x_1,...,x_n \in [-x,x]) = \big(\frac{x}{\theta}\big)^n \mathbf{1}(x \ge 0).$$

Then $f_T(x) = n\frac{x^{n-1}}{\theta^n}\mathbf{1}(x \ge 0)$. If for any function f, $E[f(T)] = c$(constant) then:

$$E[f(T)] = \int_0^\theta f(x)n\frac{x^{n-1}}{\theta^n} = c$$

which is to say: $\int_0^\theta f(x)nx^{n-1} = c\theta^n$, take derivative wrt θ yields $f(\theta) = c, \forall\theta$, then $f(T) \equiv c$. Therefore we prove the completeness. So T is both sufficient and complete.

By

$$\frac{\bar{X}}{X_{(n)} - X_{(1)}} = \frac{\bar{X}/2\theta}{\frac{X_{(n)}}{2\theta} - \frac{X_{(1)}}{2\theta}} = \frac{Z_1}{Z_2 - Z_3} \tag{1}$$

where Z_1, Z_2, Z_3 does not depend on θ thus the ratio statistics is ancillary. The conclusion follows by Basu's theorem.

Problem 2.3

(a) Let $f_i(x)$ be the density of X_i, $X_{(1)} = \min\{X_i : i = 1, \dots, n\}$ and $X_{(n)} = \max\{X_i : i = 1, \dots, n\}$. Then,

$$\begin{aligned}
\prod_{i=1}^{n} f_i(x_i) &= \prod_{i=1}^{n} \mathbf{1}(\theta - 1/2 \le x_i \le \theta + 1/2) \\
&= \mathbf{1}(\theta - 1/2 \le x_{(1)} \le \theta + 1/2)\mathbf{1}(\theta - 1/2 \le x_{(n)} \le \theta + 1/2) \cdot 1 \\
&= g(x_{(1)}, x_{(n)}, \theta) \cdot h(\mathbf{x})
\end{aligned}$$

where $h(\mathbf{x}) = 1$.

Thus, by factorization theorem, $T = \left(X_{(1)}, X_{(n)}\right)$ is a sufficient statistic.

(b) $\mathbb{E}(\bar{X}) = \theta$. *i.e.* $\bar{X}$ in unbiased estimator of θ. And $L(\theta, a) = (\theta - a)^2$ in strictly convex. Thus, by Rao-Blackwell theorem, $\delta(X_1, \dots, X_n) = \mathbb{E}\left[\bar{X}|X_{(1)}, X_{(n)}\right]$ has a strictly better MSE than the MSE of $\bar{X}$. ($\delta(X_1, \dots, X_n) \neq \bar{X}$).

Conditioned on $X_{(1)} = a, X_{(n)} = b$, There is a $\frac{1}{n}$ chance that $X_j = a$ and a $\frac{1}{n}$ chance that $X_j = b$. The remaining mass $\frac{n-2}{n}$ is uniformly spread between a and b. Thus,

$$\begin{aligned}
\mathbb{E}\left[X_i|X_{(1)} = a, X_{(n)} = b\right] &= \frac{1}{n}a + \frac{1}{n}b + \frac{n-2}{n}\int_a^b \frac{(x-a)}{(b-a)}dx \\
&= \frac{1}{n}(a+b) + \frac{n-2}{2n}(a+b) = \frac{a+b}{2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\delta(X_1, \dots, X_n) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i|X_{(1)}, X_{(n)}\right] \\
&= \frac{X_{(1)} + X_{(n)}}{2}
\end{aligned}$$

Problem 2.4

(a) We want to show that the distribution of $\frac{X_1 - X_i}{b}$ for any $i = 2, \dots, n$ does not depend on a. To do that, first consider $Y_i = X_i - a$. We have that:

$$\mathbb{P}_a(Y_i \le y) = \mathbb{P}_a(X_i - a \le y) = \mathbb{P}_a(X_i \le a + y) = F\left(\frac{a + y - a}{b}\right) = F(\frac{y}{b})$$

so for each $i = 1, \dots, n$ the distribution of Y_i does not depend on a. The result follows by noticing that $\frac{X_1 - X_i}{b} = \frac{Y_i - Y_i}{b}$;

(b) Similarly to item a, we first define $Z_i = \frac{X_i - a}{b}$ and show that Z_i does not depend on b for each $i = 1, \ldots, n$:

$$\mathbb{P}_a(Z_i \leq z) \;=\; \mathbb{P}_a(\frac{X_i - a}{b} \leq z) = \mathbb{P}_a(X_i \leq bz + a) = F\left(\frac{a + bz - a}{b}\right) = F(z)$$

and, so, the distribution of $Z_i = \frac{X_i - a}{b}$ for each i does not depend on b. The result follows by noticing that $\frac{X_1 - a}{X_i - a} = \frac{Z_1}{Z_i}$. Since the right hand side of this identity does not depend on b, neither does the left hand side.

(c) Consider Z_i as defined in item (b) above. Notice that Z_i does not depend on a nor b. Now, notice that $\frac{X_1 - X_i}{X_2 - X_i} = \frac{Z_1 - Z_i}{Z_2 - Z_i}$. Since the right hand side of this identity does not depend on (a, b), neither does the left hand side and that proves ancillaity.

Problem 2.5

(a) $\mathbb{E}_\lambda(S_1) = \mathbb{P}_\lambda(X_1 = 0) = g(\lambda)$ and $\mathbb{E}_\lambda(S_2) = \frac{1}{n}\sum_{i=1}^n \mathbb{P}_\lambda(X_i = 0) = g(\lambda)$. Thus, S_1 and S_2 are both unbiased fot $g(\theta)$.

(b)

$$\begin{aligned}\prod_{i=1}^n f_i(x_i) &= e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i} \cdot \frac{\prod_{i=1}^n \mathbf{1}_{\{0,1,\ldots,\}}(x_i)}{\prod_{i=1}^n x_i!} \\ &= g\left(\sum_{i=1}^n x_i, \lambda\right) \cdot h(\mathbf{x})\end{aligned}$$

where $h(\mathbf{x}) = \dfrac{\prod_{i=1}^n \mathbf{1}_{\{0,1,\ldots,\}}(x_i)}{\prod_{i=1}^n x_i!}$

Therefore, by factorization theorem, $T = \displaystyle\sum_{i=1}^n X_i$ is sufficient.

(c) For all $i = 1, \ldots, n$,

$$\begin{aligned}\mathbb{E}[\mathbf{1}(X_i = 0)|T = t] &= \mathbb{P}(X_j = 0|T = t) = \frac{\mathbb{P}(X_j = 0, \sum_{i\neq j} X_i = t)}{\mathbb{P}(T = t)} \\ &= \frac{\frac{e^{-\lambda}\lambda^0}{1} \cdot \frac{e^{-(n-1)\lambda}((n-1)\lambda)^t}{t!}}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}} = \left(\frac{n-1}{n}\right)^t\end{aligned}$$

Therefore,

$$\begin{aligned}S_1^* &= \mathbb{P}(\mathbf{1}(X_1 = 0|T) = \left(\frac{n-1}{n}\right)^T \\ S_2^* &= \frac{1}{n}\sum_{i=1}^n \mathbb{P}(\mathbf{1}(X_i = 0|T) = \left(\frac{n-1}{n}\right)^T\end{aligned}$$

Now let's show that $\sum_{i=1}^{n} X_i$ is a complete statistic.

$$
\begin{aligned}
\mathbb{E}_\lambda\left[g\left(\sum_{i=1}^{n} X_i\right)\right] = 0, \forall\lambda \geq 0 \quad &\Longleftrightarrow \quad \sum_{k=0}^{\infty} g(k)\frac{e^{-n\lambda}(n\lambda)^k}{k!} = 0,\ \forall\lambda \geq 0 \\
&\Longleftrightarrow \quad \sum_{k=0}^{\infty} \frac{g(k)}{k!}\eta^k = 0,\ \forall\eta = n\lambda \geq 0 \\
&\Longleftrightarrow \quad \frac{g(k)}{k!} = 0,\ \forall k = 0, \ldots,\ \text{(By uniqueness of power series expansion)} \\
&\Longleftrightarrow \quad \mathbb{P}_X\left(g\left(\sum_{i=1}^{n} X_i = 0\right)\right) = 1\ \forall\lambda \geq 1
\end{aligned}
$$

Because $T = \displaystyle\sum_{i=1}^{n} X_i$ is sufficient and complete, for any unbiased estimator $\delta(X)$, $\mathbb{E}(\delta(X)|T)$ is always same.

Problem 2.6

Poisson distribution belongs to full-rank exponential family, X is a sufficient and complete statistics for parameter λ. To construct a UMVU estimator, the only thing to do is to find an unbiased estimator and Rao-Blackwellized it by X. Let $T = (-1)^X$ compute $E[T]$ as follows:

$$
\begin{aligned}
E[T] &= \sum_{x=0}^{\infty}(-1)^x\frac{\lambda^x}{x!}e^{-\lambda} \\
&= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(-\lambda)^x}{x!} \\
&= e^{-2\lambda}.
\end{aligned}
$$

Therefore T is unbiased estimator and $E[T|X] = T$, which indicates it is a UMVU estimator.