Solutions for Stat 512 — Take home exam V

1. Seasonal ranges (in hectares) for alligators were monitored on a lake outside Gainesville, Florida, by biologists from the Florida Game and Fish Commoission. Five alligators monitored in the spring showed ranges of 7.8, 12.3, 8.3, 18.4 and 31. Four different alligators monitored in the summer showed ranges of 102.3, 81, 55.2 and 51. Estimate the difference between mean spring summer ranges, with a 95% confidence interval. (10 pts) Hint: you can use the following code in R to get S_1^2 and S_2^2: (10 pts)

```
data1<-c(7.8, 12.3, 8.3, 18.4,31)
#######S1-square#########
var(data1)
data2<-c(102.3, 81, 55.2, 51)
#######S2-square#########
var(data2)
```

Solution:

Since $n_1 = 5$ and $n_2 = 4$, we can not use large sample CI formula here. Now,

$$S_1^2 = 92.503, \qquad \bar{Y}_1 = 15.56$$
$$S_2^2 = 573.9225, \qquad \bar{Y}_2 = 72.375$$
$$\Longrightarrow S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 298.8257$$
$$\Longrightarrow S_p = \sqrt{298.8257} = 17.287$$

Hence, the 95% CI for $\mu_1 - \mu_2$ is:

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{n_1+n_2-2,\alpha/2} \cdot S_P \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
$$= 15.59 - 72.375 \pm 2.364 * 17.287 * \sqrt{\frac{1}{5} + \frac{1}{4}}$$
$$= (-78.79, -34.84)$$

****

2. Suppose $Y_1, \ldots, Y_n \sim$ Poisson(λ).

a. Show that $\bar{Y}$ is an efficient estimator of λ using CRLB. (5 pts)

b. Find the MLE for λ. (5 pts)

Solution:

For part (a), let's first find out the Fisher information number:

$$
\begin{aligned}
& f(y|\theta) = \frac{e^{-\lambda}\lambda^y}{y!} \\
\Longrightarrow\ & logf(y|\theta) = -\lambda + ylog\lambda - log(y!) \\
\Longrightarrow\ & \frac{\partial}{\partial\lambda} logf(y|\theta) = -1 + \frac{y}{\lambda} \\
\Longrightarrow\ & \left(\frac{\partial}{\partial\lambda} logf(y|\theta)\right)^2 = \frac{y^2}{\lambda^2} - \frac{2y}{\lambda} + 1 \\
\Longrightarrow\ & E\left(\frac{\partial}{\partial\lambda} logf(y|\theta)\right)^2 = \frac{Ey^2}{\lambda^2} - \frac{2E(y)}{\lambda} + 1 = \frac{\lambda + \lambda^2}{\lambda^2} - \frac{2\lambda}{\lambda} + 1 = \frac{1}{\lambda} \\
\Longrightarrow\ & I(\lambda) = nE\left(\frac{\partial}{\partial\lambda} logf(y|\theta)\right)^2 = \frac{n}{\lambda}
\end{aligned}
$$

Hence $CRLB = \frac{1}{I(\lambda)} = \frac{1}{n/\lambda} = \frac{\lambda}{n}$. It is well known that $Var(\bar{Y}) = \frac{\lambda}{n}$, which means $\bar{Y}$ is an efficient estimator of λ.

For part (b),

$$
\begin{aligned}
& L(\lambda|\underline{y}) = \frac{e^{-n\lambda}\lambda^{\sum y_i}}{\prod y_i!} \\
\Longrightarrow\ & l(\lambda|\underline{y}) = logL(\lambda|\underline{y}) = -n\lambda + log\lambda \sum y_i - log \prod y_i! \\
\Longrightarrow\ & \frac{\partial}{\partial\lambda} l(\lambda|\underline{y}) = -n + \frac{\sum y_i}{\lambda} \overset{set}{=} 0 \\
\Longrightarrow\ & \hat{\lambda} = \frac{\sum y_i}{n} = \bar{Y} \qquad \left(\frac{\partial^2}{\partial\lambda^2} l(\lambda|\underline{y}) = -\frac{\sum y_i}{\lambda^2} < 0\right)
\end{aligned}
$$

Hence the MLE of λ is $\bar{Y}$.

****

3. Suppose $Y_1, \ldots, Y_n$ is a random sample from the pdf

$$f(y|\theta) = \theta y^{-2}, \qquad 0 < \theta \leq y < \infty$$

a. What is a sufficient statistic for θ? Hint: use Factorization Theorem. (5 pts)

b. Use Maximum Likelihood Method to obtain an estimator for θ, denoted is as $\hat{\theta}$. (5 pts)

c. Use moment method to obtain an estimator for θ, denoted is as $\widetilde{\theta}$. Hint: It is possible that it does not exist. (5 pts)

Solution:

For part (a), the joint likelihood is:

$$L(\theta|\underline{y}) = \theta^n \left(\prod_{i=1}^{n} y_i\right)^{-2} I_{\{y_{(1)} \geq \theta\}}$$

Based on Factorization Thm, $Y_{(1)}$ is sufficient for θ.

For part (b), since the likelihood is increasing in θ. Hence $L(\theta|\underline{y})$ can be maximized at the maximum value θ can take, $Y_{(1)}$.

For part (c), in order to figure our the method of moment estimator, we need to calculate E(Y):

$$E(Y) = \int_{\theta}^{\infty} \frac{\theta}{y} dy = \theta ln(y)\Big|_{\theta}^{\infty} = \infty$$

Hence, mean of Y does not exist implies method of moment estimator does not exist.

****

4. Let $Y_1, \ldots, Y_n$ be i.i.d with pdf:

$$f(y|\theta) = \theta y^{\theta - 1}, \qquad , 0 \leq y \leq 1, \quad , 0 < \theta < \infty$$

a. Prove the MLE of θ is $\hat{\theta} = \dfrac{n}{\sum -log(y_i)}$. (10 pts)

b. Is the MLE in part (a) biased? Hint: Find out the distribution of $-log(y_i)$ first. (10 pts)

c. Find the method of moments estimator for θ. (5 pts)

d. Find a complete sufficient statistic for θ. (10 pts)

e. Find the MVUE for θ. (10 pts)

Solution:

For part (a),

$$
\begin{aligned}
&L(\theta|\underline{y}) = \theta^n \left(\prod y_i\right)^{\theta-1} \\
\implies &l(\theta|\underline{y}) = log(L(\theta|\underline{y})) = nlog\theta + (\theta-1)log\left(\prod y_i\right) \\
\implies &\frac{\partial}{\partial\theta} l(\theta|\underline{y}) = \frac{n}{\theta} + log\left(\prod y_i\right) \overset{set}{=} 0 \\
\implies &\hat{\theta} = \frac{n}{-log\left(\prod y_i\right)} = \frac{n}{-\sum log(y_i)} = \frac{n}{\sum -log(y_i)} \qquad \left(\frac{\partial^2}{\partial\theta^2} l(\theta|\underline{y}) = -\frac{n}{\theta^2} < 0\right)
\end{aligned}
$$

Hence, MLE for θ is $\hat{\theta} = \frac{n}{\sum -log(y_i)}$.

For part (b), using transformation technique, $-log(y_i) \sim exp(\frac{1}{\theta})$, hence $\sum -log(y_i) \sim Gamma(n, \frac{1}{\theta})$. Now, let $U = \sum -log(y_i)$, let's find out $E(\frac{n}{U})$:

$$
\begin{aligned}
E\left(\frac{n}{U}\right) &= \int_0^\infty \frac{n}{u} \cdot \frac{1}{\Gamma(n)\left(\frac{1}{\theta}\right)^n} u^{n-1} e^{-u\theta} du \\
&= \frac{n}{\Gamma(n)\left(\frac{1}{\theta}\right)^n} \int_0^\infty u^{(n-1)-1} e^{-u\theta} du \\
&= \frac{n}{\Gamma(n)\left(\frac{1}{\theta}\right)^n} \cdot \Gamma(n-1)\left(\frac{1}{\theta}\right)^{n-1} \\
&= \frac{n\theta}{n-1} \neq \theta
\end{aligned}
$$

Hence, the MLE is biased.

For part (c),

$$
E(Y) = \int_0^1 \theta y^\theta dy = \frac{\theta}{\theta+1} y^{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}
$$

Hence, set $E(Y) = \bar{Y}$ implies $\frac{\theta}{\theta+1} = \bar{Y} \implies \tilde{\theta} = \frac{\bar{Y}}{1-\bar{Y}}$.

For part (d), the pdf can be viewed alternatively:

$$
f(y|\theta) = \theta e^{(\theta-1)log(y)} I_{\{0 \leq y \leq 1\}}, \qquad \theta > 0
$$

Clearly this distribution is a member of exponential distribution, hence, $\sum log(y_i)$ is complete sufficient statistic.

For part (e), we see that the MLE of θ: $\hat{\theta} = \frac{n}{\sum -log(y_i)}$ is a function of complete sufficient statistic $\sum log(y_i)$. Since $E\left(\frac{n}{\sum -log(y_i)}\right) = \frac{n\theta}{n-1}$, simple algebra and Lehmann-Scheffe Thm indicates that $\frac{n-1}{n}\hat{\theta} = \frac{n-1}{\sum -log(y_i)}$ is the MVUE for θ.

****

5. Let $Y_1, \ldots, Y_n$ be a random sample from the pdf $f(y|\mu) = e^{-(y-\mu)}$, where $-\infty < \mu < y < \infty$.

a. Show that $Y_{(1)} = min(Y_1, \ldots, Y_n)$ is a complete sufficient statistic. Hint: You will see that $Y_{(1)}$ is not a member of exponential family. So first prove it is sufficient, then prove it is complete. (10 pts)

b. Find the MVUE for μ. (10 pts)

Solution:

For part (a),

$$L(\mu|\underline{y}) = e^{-\sum y_i} e^{n\mu} I_{\{Y_{(1)} > \mu\}}$$

By Factorization Thm, $Y_{(1)}$ is sufficient statistic for μ. Now let's prove it is also complete using definition of completeness. The distribution of $Y_{(1)}$ is:

$$f_{Y_{(1)}}(y) = ne^{-n(y-\mu)} I_{\{y>\mu\}}$$

Now,

$$\begin{aligned}
E[g(Y_{(1)})] &= \int_{\mu}^{\infty} g(y) n e^{-n(y-\mu)} dy = 0 \\
\implies \int_{\mu}^{\infty} g(y) e^{-ny} dy = 0 & \qquad \text{for all } \mu \\
\implies \frac{\partial}{\partial \mu}\left[\int_{\mu}^{\infty} g(y) e^{-ny} dy\right] = 0 & \qquad \text{for all } \mu \\
\implies -g(\mu) e^{-n\mu} = 0 & \qquad \text{for all } \mu
\end{aligned}$$

This implied $P(g(\mu) = 0) = 1$. Hence, $Y_{(1)}$ is complete and sufficient statistic.

For part (b), it is very easy to prove that $E(Y_{(1)}) = \mu + \frac{1}{n}$. Hence, $Y_{(1)} - \frac{1}{n}$ is a function based on complete sufficient statistic and also unbiased for μ, which means it is the MVUE for μ.