COLUMBIA UNIVERSITY

DEPARTMENT OF BIOSTATISTICS

P8109 – STATISTICAL INFERENCE

MIDTERM EXAMINATIONS - SPRING 2021

MODEL ANSWERS

Question 1 (1+4=5 marks)

(a) Suppose $X_i = 1$ if an item in a population possess a certain characteristic and let $X_i = 0$ otherwise. Then $\bar{X}_n$ is the proportion of items in the population that possess the characteristic.

(b) Note that $\mathcal{E}\bar{X}_n = \mathcal{E}X = p$ and $\text{var}\,\bar{X}_n = (\text{var}\,X)/n = p(1-p)/n$. With $f(\bar{X}_n) = \bar{X}_n(1-\bar{X}_n)$, a Taylor series expansion about $\bar{X}_n = p$ gives

$$
\begin{aligned}
f\left(\bar{X}_n\right) &= f\left(p\right) + \left(\bar{X}_n - p\right) f^{/}\left(p\right) + \dots \\
&= p\left(1-p\right) + \left(\bar{X}_n - p\right)\left[1 - 2\bar{X}_n\right]_{\bar{X}_n = p} + \dots \\
&\approx p\left(1-p\right) + \left(\bar{X}_n - p\right)\left(1-2p\right)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathcal{E}f\left(\bar{X}_n\right) &= \mathcal{E}\bar{X}_n\left(1-\bar{X}_n\right) \\
&\approx p\left(1-p\right) + \left(1-2p\right)\mathcal{E}\left(\bar{X}_n - p\right) \\
&= p\left(1-p\right) + 0 \\
&= p\left(1-p\right).
\end{aligned}
$$

and

$$\begin{aligned}
\operatorname{var} f\left(\bar{X}_n\right) &= \operatorname{var} \bar{X}_n\left(1-\bar{X}_n\right) \\
&\approx \left(1-2p\right)^2 \operatorname{var} \bar{X}_n \\
&= \frac{\left(1-2p\right)^2\, p\left(1-p\right)}{n}.
\end{aligned}$$

Hence, by the delta theorem,

$$\bar{X}_n\left(1-\bar{X}_n\right) \overset{\cdot}{\underset{\cdot}{\sim}} N\left[\, p\left(1-p\right), \frac{\left(1-2p\right)^2\, p\left(1-p\right)}{n}\right].$$

Question 2 (5+5+5=15 marks)

(a) We have

$$f\left(x;\theta\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) = \frac{1}{\theta^{1/r}\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\theta^{2/r}}\right)$$

The likelihood function is

$$\begin{aligned}
L_{\mathbf{X}}\left(\theta\right) &= \prod_{i=1}^{n} \frac{1}{\theta^{1/r}\sqrt{2\pi}} \exp\left(-\frac{X_i^2}{2\theta^{2/r}}\right) \\
&= \theta^{-n/r}\left(2\pi\right)^{-n/2} \exp\left(-\frac{1}{2\theta^{2/r}} \sum\nolimits_i X_i^2\right)
\end{aligned}$$

Therefore,

$$\ln L_{\mathbf{Y}}\left(\theta\right) = -\frac{n}{r}\ln\theta - \frac{n}{2}\ln\left(2\pi\right) - \frac{1}{2\theta^{2/r}}\sum_i X_i^2$$

$$\frac{\partial}{\partial\theta}\ln L_{\mathbf{Y}}\left(\theta\right) = -\frac{n}{r\theta} + \frac{\theta^{-(2/r)-1}}{r}\sum_i X_i^2$$

$$\Rightarrow \qquad -\frac{n}{r\hat{\theta}} + \frac{\hat{\theta}^{-(2/r)-1}}{r}\sum_i X_i^2 = 0$$

$$\hat{\theta} = \left(\frac{1}{n}\sum_i X_i^2\right)^{r//2}$$

(b) From (a) above, we have for the score function

$$S(\mathbf{X};\theta) = -\frac{n}{r\theta} + \frac{\theta^{-(2/r)-1}}{r}\sum_i X_i^2$$

Therefore,

$$\begin{aligned}
I_{\mathbf{X}}(\theta) &= -\mathcal{E}\frac{\partial}{\partial\theta}S(\mathbf{X};\theta) \\
&= -\mathcal{E}\left\{\frac{n}{r\theta^2} + \frac{\left(\frac{-2}{r}-1\right)\theta^{-(2/r)-2}}{r}\sum_i X_i^2\right\} \\
&= -\frac{n}{r\theta^2} + \frac{(2+r)\theta^{-(2/r)-2}}{r^2}n\mathcal{E}X_i^2 \\
&= -\frac{n}{r\theta^2} + \frac{n(2+r)\theta^{-(2/r)-2}}{r^2}\left(\operatorname{var} X + \mathcal{E}^2 X_i\right) \\
&= -\frac{n}{r\theta^2} + \frac{n(2+r)\theta^{-(2/r)-2}}{r^2}\left(\theta^{2/r} + 0\right) \\
&= \frac{n(2+r)}{r^2\theta^2} - \frac{n}{r\theta^2} \\
&= \frac{2n}{r^2\theta^2}
\end{aligned}$$

and for any unbiased estimator $\hat{\theta}_{UNB}$ of θ we have

$$\operatorname{var}\hat{\theta}_{UNB} \geq \frac{1}{I_{\mathbf{X}}(\theta)} = \frac{r^2\theta^2}{2n}$$

Hence the CRLB is $r^2\theta^2/(2n)$.

(c) Since the exponential distribution belongs to the exponential family, the CRLB is achieved for the unbiased estimator of a particular function of θ. We wish to write the score function in the form

$$S(\mathbf{X};\theta) = g_1(\theta)T - g_1(\theta)\tau = g_1(\theta)\{T(\mathbf{X}) - \tau(\theta)\}$$

where $\tau(\theta)$ is the function of θ sought and $T(\mathbf{Y})$ is the corresponding unbiased estimator. From (c) above,

$$
\begin{aligned}
S(\mathbf{X};\theta) &= -\frac{n}{r\theta} + \frac{\theta^{-(2/r)-1}}{r}\sum_i X_i^2 \\
&= \frac{n\theta^{-(2/r)-1}}{r}\left(\frac{\sum_i X_i^2}{n} - \theta^{2/r}\right)
\end{aligned}
$$

We see that $\tau(\theta) = \theta^{2/r}$ and $T(\mathbf{X}) = \left(\sum X_i^2\right)/n$. Since $T(\mathbf{X})$ is independent of *r*, $\tau(\theta)$ must also be. This happens for *r* = 2.

Dr P Gorroochurn 3/13/2021