COLUMBIA UNIVERSITY

DEPARTMENT OF BIOSTATISTICS

P8109 – STATISTICAL INFERENCE

MIDTERM EXAMINATIONS - SPRING 2020 (MODEL ANSWERS)

INSTRUCTIONS:

- **Time allowed: 60 minutes**
- **Answer ALL questions.**

Question 1

(a) The likelihood function is the joint density of *X* and *Y*, i.e.

$$L(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X-\theta)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(Y-3\theta)^2}{2}}$$

$$= \frac{1}{2\pi} e^{-\left\{\frac{(X-\theta)^2+(Y-3\theta)^2}{2}\right\}}$$

Therefore,

$$\log L(\theta) = -\log(2\pi) - \left\{\frac{(X-\theta)^2+(Y-3\theta)^2}{2}\right\}$$

$$\frac{\partial}{\partial\theta}\log L(\theta) = (x-\theta)+3(Y-3\theta)$$

Setting $\partial \log L(\theta)/\partial\theta = 0$, we have $(x-\theta)+3(Y-3\theta)=0$, so that the MLE of θ is $\hat{\theta}=(X+3Y)/10$.

To prove that $\hat{\theta}$ indeed maximizes the likelihood, we take second derivatives:

$$\frac{\partial^2}{\partial\theta^2}\log L(\theta) = -10 < 0\ ,$$

so that $\hat{\theta}$ is indeed the MLE.

(b) For the estimator $aX + bY$ to be unbiased for θ, we need

$$
\begin{aligned}
\mathrm{E}\left(aX+bY\right)&=\theta\\
a\mathrm{E}X+b\mathrm{E}Y&=\theta\\
a\left(\theta\right)+b\left(3\theta\right)&=\theta\\
\therefore \quad a+3b&=1 \quad [\text{assuming } \theta\neq 0]
\end{aligned}
$$

[8+4 =12 marks]

Question 2

The likelihood function is

$$
\begin{aligned}
L\left(\beta\right)&=\frac{Y_1^{\alpha_0-1}e^{-Y_1/\beta}}{\beta^{\alpha_0}\Gamma\left(\alpha_0\right)}\cdot\frac{Y_2^{\alpha_0-1}e^{-Y_2/\beta}}{\beta^{\alpha_0}\Gamma\left(\alpha_0\right)}\cdots\frac{Y_n^{\alpha_0-1}e^{-Y_n/\beta}}{\beta^{\alpha_0}\Gamma\left(\alpha_0\right)}\\
&=\underbrace{\frac{\left(Y_1Y_2\ldots Y_n\right)^{\alpha_0-1}}{\Gamma^n\left(\alpha_0\right)}}_{h(\mathbf{Y})}\underbrace{\frac{1}{\beta^{n\alpha_0}}\exp\left(-\frac{Y_1+Y_2+\ldots+Y_n}{\beta}\right)}_{g_\beta(T)},
\end{aligned}
$$

Using the Fisher-Neyman factorization, we see that the likelihood can be factored as $L(\beta)=h(\mathbf{Y})g_\beta\left(T\right)$, where a sufficient statistic is $T=Y_1+Y_2+\ldots+Y_n$. [Note this implies that the sample mean $\overline{Y}$ is also sufficient]

[8 marks]

Dr P Gorroochurn 3/27/2020