STA 522/Solutions to Homework 3

Problem 6.20 (Part a)

We start by finding a sufficient statistic. The joint pdf of $X_1, X_2, \ldots, X_n$ is given by

$$\begin{aligned} f(x_1, \ldots, x_n \mid \theta) &= \prod_{i=1}^{n} \left\{ \frac{2x_i}{\theta^2}\ I(0 < x_i < \theta) \right\} \\ &= \underbrace{\left(2^n \prod_{i=1}^{n} x_i \right)}_{=h(\underline{x})} \underbrace{\left(\frac{1}{\theta^{2n}} I\left(x_{(n)} < \theta\right) \right)}_{g(T(\underline{x})|\theta)} \end{aligned}$$

Therefore, by Factorization theorem, $T(\underline{X}) = X_{(n)} = \max_{1 \le i \le n} X_i$ is sufficient for θ. To obtain the pdf of $T = X_{(n)}$ first note that the common cdf of X_i is

$$F(x \mid \theta) = \int_0^x \frac{2y}{\theta^2}\ dy = \frac{x^2}{\theta^2}; \quad \text{for } 0 < x < \theta.$$

Therefore, it follows from a result discussed in class that the pdf of $T = X_{(n)}$ is given by:

$$f_T(t \mid \theta) = n\ f(t \mid \theta)\ \{F(t \mid \theta)\}^{n-1} = \frac{2n}{\theta^{2n}} t^{2n-1}\ I(0 < y < \theta)$$

To prove completeness start with a function $g(t)$ with

$$u(\theta) := \mathrm{E}_\theta[g(T)] = \int_0^\theta g(t)\ \frac{2n}{\theta^{2n}}\ t^{2n-1}\ dt = 0 \quad \text{for all } \theta > 0.$$

Then $u'(\theta) = 0$ (derivative of a constant function) for all θ, which implies

$$\begin{aligned} 0 &= \frac{d}{d\theta} \left(\frac{2n}{\theta^{2n}} \int_0^\theta g(t)\ t^{2n-1}\ dt \right) \\ &= \left(\frac{d}{d\theta} \frac{2n}{\theta^{2n}} \right) \int_0^\theta g(t)\ t^{2n-1}\ dt + \frac{2n}{\theta^{2n}} \underbrace{\left(\frac{d}{d\theta} \int_0^\theta g(t)\ t^{2n-1}\ dt \right)}_{=g(\theta)\ \theta^{2n-1}, \text{ by Fund. Thm. of Calculus}} \\ &= \left(\frac{d}{d\theta} \frac{2n}{\theta^{2n}} \right) \underbrace{\left(\int_0^\theta g(t)\ \frac{2n}{\theta^{2n}}\ t^{2n-1}\ dt \right)}_{=\ u(\theta)\ =\ 0} \left(\frac{\theta^{2n}}{2n} \right) \\ &\qquad + \frac{2n}{\theta^{2n}} \left(g(\theta)\ \theta^{2n-1}\ dt \right) \\ &= 0 + 2n\ \frac{1}{\theta}\ g(\theta) \quad \text{for all } \theta \in (0, \theta) \\ \implies g(\theta) &= 0 \quad \text{for all } \theta \in (0, \theta) \end{aligned}$$

which implies $P_\theta(g(T) = 0) = 1$ for all $\theta \in (0, \theta)$. This means $T = X_{(n)}$ is complete.

Problem 6.22

Part (a): The joint pdf of $X_1, X_2, \ldots, X_n$ is given by:

$$f(\underline{x} \mid \theta) = \underbrace{\theta^n \exp\left[(\theta-1)\sum_{i=1}^{n} \log x_i\right]}_{=g(T(\underline{x})|\theta)} \underbrace{\prod_{i=1}^{n} I(0 < x_i < 1)}_{=h(\underline{x})}$$

Therefore by the Factorization theorem, $T(\underline{X}) = \sum_{i=1}^{n} \log X_i$ is a sufficient statistics for θ. Since $\sum_{i=1}^{n} X_i$ is not a one-to-one function of $T(\underline{X}) = \sum_{i=1}^{n} \log X_i$, therefore $\sum_{i=1}^{n} X_i$ is NOT a sufficient statistic for θ (from the reverse implication of the Factorization theorem).

Part (b): We note that $f(x \mid \theta)$ is a member of the exponential family:

$$f(x \mid \theta) = \underbrace{\theta^n}_{=c(\theta)} \underbrace{I(0 < x < 1)}_{=h(x)} \exp\left[\underbrace{(\theta-1)}_{=w_1(\theta)} \underbrace{\log x}_{=T_1(x)}\right]$$

and the parameter space $\Theta = \{\theta : \theta > 0\}$ contains the open interval, e.g., $(1, 2)$ in $\mathbb{R}^1$. Therefore, using a theorem on exponential family discussed in class, $\sum_{i=1}^{n} T_1(X_i) = \sum_{i=1}^{n} \log X_i$ is complete sufficient for θ.

Problem 6.40

Say $X_1, X_2, \ldots, X_n$ be iid from the location scale family $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Then we can write $X_i = \mu + \sigma Z_i$ where $Z_i \sim \text{iid}\, f(x)$.

Part (a): We have

$$\frac{T_1(X_1, X_2, \ldots, X_n)}{T_2(X_1, X_2, \ldots, X_n)} = \frac{T_1(\mu + \sigma Z_1, \ldots, \mu + \sigma Z_n)}{T_2(\mu + \sigma Z_1, \ldots, \mu + \sigma Z_n)} = \frac{\sigma\ T_1(Z_1, \ldots, Z_n)}{\sigma\ T_2(Z_1, \ldots, Z_n)} = \frac{T_1(Z_1, \ldots, Z_n)}{T_2(Z_1, \ldots, Z_n)}$$

The right hand side involves random variables whose pdf does not involve the parameters μ and σ. Hence T_1/T_2 is ancillary.

Part (b): We have for $a > 0$, $b \in \mathbb{R}$ and any $\underline{x} = (x_1, \ldots, x_n)$

$$\mathrm{R}(aX_1 + b, \ldots, aX_n + b) = \max_i(aX_i + b) - \min_i(aX_i + b) = a\left(\max_i X_i - \min_i X_i\right) = a\,\mathrm{R}(X_1, \ldots, X_n)$$

and

$$\begin{aligned}\mathrm{S}(aX_1 + b, \ldots, aX_n + b) &= \left(\frac{1}{n-1}\sum_{i=1}^{n}\left((aX_i + b) - \frac{1}{n}\sum_{i=1}^{n}(aX_i + b)\right)^2\right)^{1/2} \\ &= a\ \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)^{1/2} \\ &= a\ S(X_1, \ldots, X_n)\end{aligned}$$