Assignment 3 Solution

1. Let $X_1, \cdots, X_n$ be *i.i.d.* random variables from the probability density function of the following form:

$$f_X(x|\theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

where $\theta > 0$. Find a minimal sufficient statistic for θ.

Solution: The joint pdf is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \left(\frac{2x_i}{\theta^2} I(0 < x_i < \theta) \right) \\ &= \frac{2^n \prod_{i=1}^{n} x_i}{\theta^{2n}} I(0 < x_{(1)} < x_{(n)} < \theta) \\ &= \frac{2^n \prod_{i=1}^{n} x_i}{\theta^{2n}} I(x_{(1)} > 0) I(x_{(n)} < \theta) \end{aligned}$$

Let $\mathbf{x}, \mathbf{y}$ be two sample points from the sample space $\mathcal{X}$.

$$\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} = \frac{\prod_{i=1}^{n} x_i I(x_{(1)} > 0) I(x_{(n)} < \theta)}{\prod_{i=1}^{n} y_i I(y_{(1)} > 0) I(y_{(n)} < \theta)}$$

The ratio above is constant to θ if and only if $x_{(n)} = y_{(n)}$. To see this, note that if $x_{(n)} = y_{(n)}$, then the ratio above is free of θ. Conversely, suppose $x_{(n)} \neq y_{(n)}$. Without loss of generality, assume $x_{(n)} < y_{(n)}$. Then for $\theta < x_{(n)}$, the above ratio equals

$$\frac{\prod_{i=1}^{n} x_i I(x_{(1)} > 0)}{\prod_{i=1}^{n} y_i I(y_{(1)} > 0)}.$$

But if $x_{(n)} < \theta < y_{(n)}$, then the above ratio equals 0. Hence the above ratio implicitly changes as a function of θ. This establishes the implication in both directions. Thus, by Theorem 6.2.13, $T(\mathbf{X}) = X_{(n)}$ is a minimal sufficient statistic for θ.

2. Suppose that $X_1, \cdots, X_n$ are *i.i.d.* random variables from pdf

$$f_X(x|\theta) = \theta x^{\theta-1} \exp\left(-x^\theta\right)$$

where $\theta > 0$, $x > 0$. Show that $(\log X_{(n)})/(\log X_{(1)})$ is an ancillary statistic.

Solution: Let $Y_i = \log X_i$. By Theorem 2.1.5, the pdf of Y_i is

$$
\begin{aligned}
f_Y(y|\theta) &= f_X(e^y|\theta)\left|\frac{dx}{dy}\right| \\
&= \theta \exp\{(\theta-1)y\}\exp\left\{\left(-e^{\theta y}\right)\right\}e^y \\
&= \theta \exp\left(\theta y - e^{\theta y}\right)
\end{aligned}
$$

where $-\infty < y < \infty$. Let $Z_i = \theta Y_i$, then the pdf of Z_i is

$$
\begin{aligned}
f_Z(z|\theta) &= f_Y(z/\theta|\theta)\left|\frac{dy}{dz}\right| \\
&= \exp\left(z - e^z\right)
\end{aligned}
$$

which is not a function of θ. Because $\log(\cdot)$ is monotonically increasing function, $Z_{(i)} = \theta Y_{(i)} = \theta \log X_{(i)}$. Hence,

$$(\log X_{(n)})/(\log X_{(1)}) = Y_{(n)}/Y_{(1)} = (Z_{(n)}/\theta)/(Z_{(1)}/\theta) = Z_{(n)}/Z_{(1)}$$

can be written as $\mathbf{Z} = (Z_1, \cdots, Z_n)$, whose distribution does not depend on θ. Therefore, $(\log X_{(n)})/(\log X_{(1)})$ is an ancillary statistic.

3. Let $X_1, \cdots, X_n$ be *i.i.d.* random variables from a uniform distribution Uniform$(-\theta, \theta)$ with the pdf given by

$$f_X(x|\theta) = \frac{1}{2\theta} I\left(-\theta < x < \theta\right), \qquad \theta > 0$$

   (a) Is the two dimensional statistic $T_1(\mathbf{X}) = \left(X_{(1)}, X_{(n)}\right)$ a complete sufficient statistic? Justify your answer.

   Solution: Let $h(T_1) = X_{(n)}/X_{(1)}$, $Z_i = X_i/\theta$. Because $X_i \sim$ Uniform$(-\theta, \theta)$, the distribution of $Z_i \sim$ Uniform$(-1, 1)$ does not depend on θ. Define

$$g(T_1) = h(T_1) - \mathrm{E}[h(T_1)] = \frac{X_{(n)}}{X_{(1)}} - \mathrm{E}\left[\frac{X_{(n)}}{X_{(1)}}\right] = \frac{Z_{(n)}}{Z_{(1)}} - \mathrm{E}\left[\frac{Z_{(n)}}{Z_{(1)}}\right]$$

   Then $\mathrm{E}[g(T_1(\mathbf{X}))|\theta] = 0$ for all θ and $\Pr[g(T_1(\mathbf{X})) = 0|\theta] < 1$. Hence, $T_1(\mathbf{X})$ is not a complete statistic¿ One could, however, establish using factorization theorem that T_1 is indeed sufficient.

(b) Is the one-dimensional statistic $T_2(\mathbf{X}) = \max_i \{|X_i|\}$ a complete sufficient statistic? Justify your answer.

Solution: Considering that $|X_i| \sim \text{Uniform}(0, \theta)$ and $T_2(\mathbf{X}) = \max_i |X_i|$ has pdf of $\frac{nt^{n-1}}{\theta^n} I(0 < t < \theta)$, suppose that there exist $g(T_2)$ such that $\mathrm{E}[g(T_2)|\theta] = 0$. Similar to the lecture note, we have

$$
\begin{aligned}
f_{T_2}(t|\theta) &= \frac{nt^{n-1}}{\theta^n} I(0 < t < \theta) \\
E[g(T_2)|\theta] &= \int_0^\theta \frac{nt^{n-1} g(t)}{\theta^n} dt \\
&= \frac{n}{\theta^n} \int_0^\theta g(t) t^{n-1} dt = 0 \\
\int_0^\theta g(t) t^{n-1} dt &= 0 \\
g(\theta)\theta^{n-1} &= 0 \qquad \text{(by taking derivative)} \\
g(\theta) &= 0
\end{aligned}
$$

for all $\theta > 0$. Because $g(T_2) = 0$ and $\Pr[g(T_2) = 0|\theta] = 1$ holds for all $\theta > 0$, T_2 is a complete statistic. Also, the joint pdf of $\mathbf{X}$ can be represented as a function of T_2

$$
f_{\mathbf{X}}(\mathbf{x}|\theta) = \frac{1}{2^n \theta^n} I(-\theta < x_1, \cdots, x_n < \theta) = \frac{1}{2^n \theta^n} I(\max_i |x_i| < \theta) = \frac{1}{2^n \theta^n} I(T_2(\mathbf{x}) < \theta)
$$

Hence, T_2 is also a sufficient statistic by Factorization Theorem (by setting $h(\mathbf{x}) = 1$). Therefore, $T_2(\mathbf{X})$ is a complete sufficient statistic.

4. Let $X_1, \cdots, X_n$ be *i.i.d.* random variables from $N(\mu, \sigma^2)$ population with $\underline{\text{known } \mu}$. Find a one-dimensional minimal sufficient statistic for σ^2.

Solution: The joint pdf is given by:

$$
f_{\mathbf{X}}(\mathbf{x}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right)
$$

For two sample points $\mathbf{x}, \mathbf{y}$, the ratio of the pdf's can be written as:

$$\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma^2)}{f_{\mathbf{X}}(\mathbf{y}|\mu,\sigma^2)} &= \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right) / \exp\left(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right) \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) - \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2)\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2\right) + \frac{\mu}{\sigma^2}\left(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i\right) + \left(\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i\right)\right]
\end{aligned}$$

The ratio above will not depend on σ^2 if and only if

$$\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i = \sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i.$$

Therefore, $\mathbf{T}(\mathbf{X}) = \sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i$ is a minimal sufficient statistic for σ^2 by Theorem 6.2.13. Alternatively, $\mathbf{T}(\mathbf{X}) = \sum_{i=1}^n (x_i-\mu)^2$ can be a minimal sufficient statistics for σ^2.

5. Let $X_1, \cdots, X_n$ be *i.i.d.* observations uniformly drawn from $\{1, 2, \cdots, \theta\}$, where θ is a positive integer. This corresponds to a discrete uniform with pmf

$$f_X(x|\theta) = \begin{cases} 1/\theta & x = 1, 2, \cdots, \theta \\ 0 & \text{otherwise} \end{cases}$$

Show that $T(\mathbf{X}) = \max_i X_i$ is a complete, minimal sufficient statistic.

Solution: Establishing minimal sufficiency is similar to the continuous $Unif(0,\theta)$ case and the details are omitted here. The completeness is proved using the following argument.

First let us find out $P(\Pr(\max_i X_i = k) \quad \text{for } k = \{1, \cdots, \theta\}$. Note that

$$\begin{aligned}
\Pr(\max_i X_i = k) &= \Pr(\max_i X_i \le k) - \Pr(\max_i X_i \le k-1) \\
&= \prod_{i=1}^n \left(\frac{k}{\theta}\right) - \prod_{i=1}^n \left(\frac{k-1}{\theta}\right) \\
&= \theta^{-n}\left[k^n - (k-1)^n\right]
\end{aligned}$$

For an arbitrary function g, the equation $E\left[g(T)\right] = 0$ for all positive integer θ implies

$$
\begin{aligned}
&\theta^{-n} \sum_{k=1}^{\theta} g(k)\left[k^n - (k-1)^n\right] = 0 \\
\Longrightarrow \quad & \sum_{k=1}^{\theta} g(k)\left[k^n - (k-1)^n\right] = 0. \qquad (1)
\end{aligned}
$$

Equation (1) is true for all $k = 1, 2, \ldots, \theta$. So plug in different values of θ successively.

$$
\begin{aligned}
\text{For k} = 1,\ (1) \quad &\Longrightarrow \quad g(1) = 0 \\
\text{For k} = 2,\ (1) \quad &\Longrightarrow \quad g(1) + (2^n - 1)g(2) = 0 \Longrightarrow g(2) = 0 \\
\text{For k} = 3,\ (1) \quad &\Longrightarrow \quad g(1) + (2^n - 1)g(2) + (3^n - 2^n)g(3) = 0 \Longrightarrow g(3) = 0 \\
\cdots \quad & \qquad \cdots\cdots \\
\cdots \quad & \qquad \cdots\cdots \\
\cdots \quad & \qquad \cdots\cdots
\end{aligned}
$$

Hence $g \equiv 0$ and family of pmf's for T is complete.

In order to show that $T(\mathbf{X}) = \max_i X_i$ is indeed minimal sufficient statistic.

$$
\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{y}|\theta)} = \frac{\theta^{-n} I(\max_i x_i \in \{1, 2, .., \theta\})}{\theta^{-n} I(\max_i y_i \in \{1, 2, .., \theta\})}
$$

Because the ratio is a constant as of θ if and only if $\max_i X_i = \max_i Y_i$. Therefore, $T(\mathbf{X}) = \max_i X_i$ is a complete, minimal sufficient statistic.