MATH-503 – Assignment 2
Due Friday February 10, 2017

1. **Let $X_1, \ldots, X_n$ be a random sample with pdf**

$$f(x) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha}, \quad 0 < x < \theta, \quad \alpha, \theta > 0$$

Find a sufficient statistic for (α, θ).

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^n \left[\frac{\alpha x_i^{\alpha-1}}{\theta^\alpha} I_{\{0<x_i<\theta\}}\right] = \frac{\alpha^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1}}{\theta^{n\alpha}} I_{\{x_{(1)}>0\}} I_{\{x_{(n)}<\theta\}}$$

Using the factorization theorem,

$$h(\boldsymbol{x}) = I_{\{x_{(1)}>0\}} \qquad g(T(\boldsymbol{x})|\theta) = \theta^{-n\alpha}\alpha^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} I_{\{x_{(n)}<\theta\}}$$

Therefore, $T(\boldsymbol{X}) = \left(\prod_{i=1}^n X_i, X_{(n)}\right)$ is sufficient for (α, θ).

2. **Let $X_1, \ldots, X_n$ be a random sample from a Pareto distribution with pdf**

$$f(x|\theta) = \frac{\theta}{(1+x)^{\theta+1}}, \qquad x > 0, \qquad \theta > 0$$

Write $f(\boldsymbol{x}|\theta)$ in the exponential family form and deduce a sufficient statistic for θ.

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^n \left[\frac{\theta}{(1+x_i)^{\theta+1}}\right] \\ &= \prod_{i=1}^n \left[\theta(1+x_i)^{-1} \exp\{-\theta \ln(1+x_i)\}\right] \\ &= \left[\prod_{i=1}^n (1+x_i)^{-1}\right] \theta^n \exp\left\{-\theta \sum_{i=1}^n \ln(1+x_i)\right\} \end{aligned}$$

Thus, the Pareto distribution is an exponential family distribution with

$$h(\boldsymbol{x}) = \prod_{i=1}^n (1+x_i)^{-1} \qquad c(\theta) = \theta^n \qquad t(\boldsymbol{x}) = \sum_{i=1}^n \ln(1+x_i) \qquad \omega(\theta) = -\theta$$

Hence, a sufficient statistic for θ is

$$T(\boldsymbol{X}) = \sum_{i=1}^n \ln(1+X_i)$$

3. **Let $X_1, \ldots, X_n$ be iid random variables from a uniform distribution on $(\theta, 2\theta)$**

$$X_i \sim U(\theta, 2\theta), \qquad \theta < x_i < 2\theta, \quad \theta > 0$$

(a) **Find a minimal sufficient statistic for θ.**

$$\begin{aligned}
f(x_i|\theta) &= \frac{1}{2\theta - \theta} I_{\{\theta < x_i < 2\theta\}} = \frac{1}{\theta} I_{\{\theta < x_i < 2\theta\}} \\
f(\boldsymbol{x}|\theta) &= \frac{1}{\theta^n} I_{\{x_{(1)} > \theta\}} I_{\{x_{(n)} < 2\theta\}} \\
\frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{y}|\theta)} &= \frac{\frac{1}{\theta^n} I_{\{x_{(1)} > \theta\}} I_{\{x_{(n)} < 2\theta\}}}{\frac{1}{\theta^n} I_{\{y_{(1)} > \theta\}} I_{\{y_{(n)} < 2\theta\}}} = \frac{I_{\{x_{(1)} > \theta\}} I_{\{x_{(n)} < 2\theta\}}}{I_{\{y_{(1)} > \theta\}} I_{\{y_{(n)} < 2\theta\}}}
\end{aligned}$$

this ratio is independent of θ if and only if $X_{(1)} = Y_{(1)}$ and $X_{(n)} = Y_{(n)}$. Therefore, $T(\boldsymbol{X}) = \left(X_{(1)}, X_{(n)}\right)$ is minimal sufficient for θ.

(b) **Show that $X \sim U(\theta, 2\theta)$ is a scale family distribution.**

Let $Z \sim U(1, 2)$

$$f_Z(z) = 1 \times I_{\{1 < z < 2\}}$$

Let $X = \theta Z \;\Rightarrow\; Z = \frac{X}{\theta}$

$$f_X(x) = f_Z\left(\frac{x}{\theta}\right) \left| \frac{d}{dx} \frac{x}{\theta} \right| = 1 \times I_{\{1 < \frac{x}{\theta} < 2\}} \frac{1}{\theta} = \frac{1}{\theta} I_{\{\theta < x < 2\theta\}}$$

Therefore, X is a scale family distribution.

(c) **Show that $\frac{X_{(1)}}{X_{(n)}}$ is an ancillary statistic, where**

$$X_{(1)} = \min(X_1, \ldots, X_n) \qquad X_{(n)} = \max(X_1, \ldots, X_n)$$

$$\begin{aligned}
P\left(\frac{X_{(1)}}{X_{(n)}} < s\right) &= P\left(\frac{\min(X_1, \ldots, X_n)}{\max(X_1, \ldots, X_n)} < s\right) \\
&= P\left(\frac{\min(\theta Z_1, \ldots, \theta Z_n)}{\max(\theta Z_1, \ldots, \theta Z_n)} < s\right) \\
&= P\left(\frac{\theta \min(Z_1, \ldots, Z_n)}{\theta \max(Z_1, \ldots, Z_n)} < s\right) = P\left(\frac{\min(Z_1, \ldots, Z_n)}{\max(Z_1, \ldots, Z_n)} < s\right)
\end{aligned}$$

which does not depend on θ. Therefore, $\frac{X_{(1)}}{X_{(n)}}$ is an ancillary statistic.

4. Let $X_1, \ldots, X_n$ be a random sample with pdf

$$f(x|\theta) = \exp(\theta - x), \quad x > \theta, \quad \theta \in \mathbb{R}$$

(a) **Show that $X_{(1)} = \min(X_1, \ldots, X_n)$ is a sufficient statistic for θ using the definition of sufficient statistic. That is, show $f\left(\boldsymbol{x}|x_{(1)}\right)$ does not depend on θ.**

The cdf of X is given by

$$F_X(u) = \int_\theta^u \exp(\theta - x)\,dx = -\left. e^\theta e^{-x}\right|_\theta^u = -e^\theta(e^{-u}) - e^{-\theta}) = 1 - e^{\theta-u}, \quad u > \theta$$

and the pdf of $X_{(1)}$ is

$$q(t|\theta) = f_{X_{(1)}}(u) = n\,(1 - F_X(u))^{n-1} f_X(u) = n\left[e^{\theta-u}\right]^{n-1} e^{\theta-u} = ne^{n(\theta-u)}$$

$$\begin{aligned} P(X_1, \ldots, X_n|T(\boldsymbol{X})) &= \frac{f(\boldsymbol{x}|\theta)}{q(t|\theta)} = \frac{\prod_{i=1}^n \left(e^{\theta-x_i} I_{\{x_i>\theta\}}\right)}{ne^{(\theta-t)} I_{\{t>\theta\}}} \\ &= \frac{e^{n\theta} e^{-\sum_{i=1}^n x_i} I_{\{x_{(1)}>\theta\}}}{ne^{n(\theta-x_{(1)})} I_{\{x_{(1)}>\theta\}}} = \frac{1}{n} e^{-n(\bar{x}-x_{(1)})} \end{aligned}$$

which does not depend on θ. Therefore, $X_{(1)}$ is a sufficient statistic for θ.

(b) **Find a minimal sufficient statistic for θ.**

$$\frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{y}|\theta)} = \frac{e^{n\theta} e^{-\sum_{i=1}^n x_i} I_{\{x_{(1)}>\theta\}}}{e^{n\theta} e^{-\sum_{i=1}^n y_i} I_{\{y_{(1)}>\theta\}}} = e^{-(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i)} \frac{I_{\{x_{(1)}>\theta\}}}{I_{\{y_{(1)}>\theta\}}}$$

which is independent of θ if and only if $X_{(1)} = Y_{(1)}$. Therefore, $X_{(1)}$ is minimal sufficient for θ.

(c) **Show that $X_{(1)}$ is a complete statistic.**

Suppose $E[g(T)] = 0$, that is, $\int_\theta^\infty g(t) \cdot ne^{n(\theta-t)} dt = 0$.
Since $E[g(T)]$ is constant as a function of θ, its derivative with respect to θ is 0.

$$\begin{aligned} 0 = \frac{d}{d\theta} E[g(T)] &= \frac{d}{d\theta} \int_\theta^\infty g(t) \cdot ne^{n(\theta-t)} dt = \frac{d}{d\theta}\left[ne^{n\theta} \int_\theta^\infty g(t) e^{-nt} dt\right] \\ &= n^2 e^{n\theta} \int_\theta^\infty g(t) e^{-nt} dt + ne^{n\theta} \frac{d}{d\theta} \int_\theta^\infty g(t) e^{-nt} dt \qquad \text{by the product rule} \\ &= 0 - ne^{n\theta} g(\theta) e^{-n\theta} \\ &= -ng(\theta) \end{aligned}$$

Thus, $g(\theta) = 0$ for all θ. That is, $P\left(g(T) = 0\right) = 1$. Therefore, $T(\boldsymbol{X}) = X_{(1)}$ is a complete statistic.

(d) **Use Basu's theorem to show that $X_{(1)}$ and S^2 are independent, where**

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Note that $f(x|\theta) = e^{\theta - x} = e^{-(x-\theta)}$, $x > \theta$ is a location family distribution where

$$X_i = Z_i + \theta \qquad \text{with} \qquad Z_i \sim f_Z(z) = e^{-z}$$

$$\begin{aligned}
P\left(S^2 \le u\right) &= P\left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1} \le u\right) \\
&= P\left(\frac{1}{n-1}\sum_{i=1}^n \left[(Z_i + \theta) - \frac{1}{n}\sum_{i=1}^n (Z_i + \theta)\right]^2 \le u\right) \\
&= P\left(\frac{1}{n-1}\sum_{i=1}^n \left(Z_i - \bar{Z}\right)^2 \le u\right)
\end{aligned}$$

which does not depend on θ. Thus, S^2 is an ancillary statistic.
We have shown that $X_{(1)}$ is minimal sufficient and complete. Thus, by Basu's theorem, $X_{(1)}$ is independent of S^2.