STA5327 Homework 3. $\boxed{6.16}$ ①

#6.13. Let $Y_1 = \log X_1$

$Y_2 = \log X_2$

$$y = \log x \quad (x>0,\ \alpha>0)$$

$$g^{-1}(y) = e^y \qquad \frac{d}{dy} g^{-1}(y) = e^y \qquad (-\infty < y < \infty,\ \alpha > 0)$$

By Theorem 2.1.5

$$f_Y(y|\alpha) = f_X(g^{-1}(y)) \left|\frac{d}{dy} g^{-1}(y)\right|$$

$$= e^y \cdot \alpha\left(e^{y(\alpha-1)} e^{-e^{\alpha y}}\right) = \alpha \exp\{y(\alpha-1) - e^{\alpha y} + y\}$$

$$= \alpha \exp\{\alpha y - e^{\alpha y}\}$$

$$= \frac{1}{1/\alpha} \exp\{\tfrac{1}{1/\alpha} y - e^{y/1/\alpha}\} \quad (-\infty < y < \infty)$$

Since Y_i is a SF (scale family) with scale parameter $1/\alpha$, so we can rewrite $Y_i = \frac{1}{\alpha} Z_i$, where $Z_i \sim f(z|1)$ by Theorem 3.5.6.

Then
$$\frac{\log X_1}{\log X_2} = \frac{Y_1}{Y_2} = \frac{\frac{1}{\alpha} Z_1}{\frac{1}{\alpha} Z_2} = \frac{Z_1}{Z_2}$$

$X_i = \alpha Z_i + 1$

$Z_i = \frac{1}{\alpha}(X_i - 1)$

Since the distribution of $\frac{Z_1}{Z_2}$ doesn't depend on α, thus $\frac{\log X_1}{\log X_2}$ is an ancillary statistic.

6.14 Given $X_1, X_2, \cdots, X_n$ are a random sample from a LF (location family).

Then by Theorem 3.5.6, we know

$$X_i = Z_i + \mu \quad \text{, where } Z_i \sim N(0,1) \text{ and } \mu \text{ is location parameter}$$

Now We want to show that $M - \bar{X}$ is an ancillary statistic, where M is the sample median

$$X_i = Z_i + \mu \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}(Z_i + \mu)$$

$$\bar{X} = \bar{Z} + \mu$$

Let $M(x)$ be a median function of $\{X_i\}_1^n$.

Then $M(x) = M(Z+\mu) = M(Z) + \mu$

Therefore $M(x) - \bar{X} = M(Z) + \cancel{\mu} - \bar{Z} - \cancel{\mu} = M(Z) - \bar{Z}$.

But the distribution $M(x) - \bar{X}$ doesn't dependent on μ.

Thus, $M(x) - \bar{X}$ is an ancillary statistic as we desire.

#6.16 Consider the multinomial distribution with cell probabilities

$$\left(\frac{1}{2}+\frac{\theta}{4},\ \frac{1}{4}(1-\theta),\ \frac{1}{4}(1-\theta),\ \frac{\theta}{4}\right).$$

(a) Show that this is a curved exponential family.

The joint pmf $f(x_1, x_2, x_3, x_4 | \theta)$

$$= \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\left(\frac{1}{2}+\frac{\theta}{4}\right)^{x_1}\left(\frac{1}{4}(1-\theta)\right)^{x_2+x_3}\left(\frac{\theta}{4}\right)^{x_4} I_{\{x_1,x_2,x_3,x_4\geq 0,\ \sum_{i=1}^{4}x_i=n\}}(x)$$

$$= \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\left(\frac{1}{4}\right)^{n}(2+\theta)^{x_1}(1-\theta)^{x_2+x_3}\,\theta^{x_4}\, I_{\{x_1,x_2,x_3,x_4\geq 0,\ \sum_{i=1}^{4}x_i=n\}}(x)$$

Let the constant part be $K = \frac{n!}{x_1!\,x_2!\,x_3!\,x_4!}\left(\frac{1}{4}\right)^{n} I_{\{x_1,x_2,x_3,x_4\geq 0,\ \sum_{i=1}^{4}x_i=n\}}(x)$

$$= K\exp\left\{x_1\log(2+\theta) + (x_2+x_3)\log(1-\theta) + x_4\log\theta\right\}$$

since $x_2+x_3 = n - x_1 - x_4$, so

$$= K\exp\left\{x_1\log(2+\theta) + (n-x_1-x_4)\log(1-\theta) + x_4\log\theta\right\}$$

$$= K(1-\theta)^{n}\exp\left\{x_1\log\left(\frac{2+\theta}{1-\theta}\right) + x_4\log\left(\frac{\theta}{1-\theta}\right)\right\}$$

Therefore, this is a curved exponential family

6.16

(b) Find a sufficient statistic for θ

Joint pmf

$$f(x_1, x_2, x_3, x_4 \mid \theta) = \underbrace{K}_{h(x)} \underbrace{(1-\theta)^n \exp\left\{x_1 \log\left(\frac{2+\theta}{1-\theta}\right) + x_4 \log\left(\frac{\theta}{1-\theta}\right)\right\}}_{g(T(x) \mid \theta)}$$

Sufficient statistic for θ is $T(X) = (x_1, x_4)$ by FC..

#6.16

(c) Let $X=(x_1, x_2, x_3, x_4)$ and $Y=(y_1, y_2, y_3, y_4)$ denote two multinomial vectors with the cell probability $(P_1^x, P_2^x, P_3^x, P_4^x)$ and $(P_1^y, P_2^y, P_3^y, P_4^y)$ respectively, where

$$P_1^x = \frac{1}{2}+\frac{\theta}{4} \qquad P_1^y = \frac{1}{2}+\frac{\theta}{4}$$
$$P_2^x = \frac{1}{4}(1-\theta) \qquad P_2^y = \frac{1}{4}(1-\theta)$$
$$P_3^x = \frac{1}{4}(1-\theta) \qquad P_3^y = \frac{1}{4}(1-\theta)$$
$$P_4^x = \frac{\theta}{4} \qquad P_4^y = \frac{\theta}{4}.$$

Then, the ratio of two densities is

$$\frac{f(X|\theta)}{f(Y|\theta)} = \frac{K_x (1-\theta)^n \exp\{x_1 \log\left(\frac{2+\theta}{1-\theta}\right) + x_4 \log\left(\frac{\theta}{1-\theta}\right)\}}{K_y (1-\theta)^n \exp\{y_1 \log\left(\frac{2+\theta}{1-\theta}\right) + y_4 \log\left(\frac{\theta}{1-\theta}\right)\}}.$$

where

$$K_x = \frac{n!}{x_1!x_2!x_3!x_4!}\left(\frac{1}{4}\right)^n I_{\{x_1,x_2,x_3,x_4\geq 0,\ \sum_{i=1}^4 x_i = n\}}(x)$$

$$K_y = \frac{n!}{y_1!y_2!y_3!y_4!}\left(\frac{1}{4}\right)^n I_{\{y_1,y_2,y_3,y_4\geq 0,\ \sum_{i=1}^4 y_i = n\}}(y)$$

This ratio will be constant as a function of θ iff $x_1=y_1$ and $x_4=y_4$. Thus, $T(X)=(x_1, x_4)$ is a minimal sufficient statistic for θ by Lehmann-Scheffe Theorem.

6.25

$$\frac{f(x|\mu,\sigma^2)}{f(y|\mu,\sigma^2)} = \frac{\cancel{(2\pi\sigma^2)^{-n/2}} \cdot \exp\left(-\left(n(\bar{x}-\mu)^2 + (n-1)S_x^2\right)/2\sigma^2\right)}{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\left(-\left(n(\bar{y}-\mu)^2 + (n-1)S_y^2\right)/2\sigma^2\right)}$$

Note: $(n-1)S_x^2 = \Sigma x_i^2 - n\bar{x}^2$

$$= \exp\left[-\left(n\cdot(\bar{x}-\mu)^2 + \left(\Sigma x_i^2 - n\bar{x}^2\right)\right) + \left(n(\bar{y}-\mu)^2 + (\Sigma y^2 - n\bar{y}^2)\right)\right]$$

$$= \exp\left[\frac{\mu}{\sigma^2}\left(\Sigma x_i - \Sigma y_i\right) - \frac{1}{2\sigma^2}\left(\Sigma x_i^2 - \Sigma y_i^2\right)\right] \quad \text{— eqn (1)}$$

(a) The statistic $(\Sigma x_i, \Sigma x_i^2)$ is sufficient, but not minimal sufficient, in the $n(\mu,\mu)$ family. (b) The statistic Σx_i^2 is minimal sufficient in the $n(\mu,\mu)$ family.

soln) From eqn(1), we can obtain

$$\frac{f(x|\mu,\mu)}{f(y|\mu,\mu)} = \exp\left[\frac{\mu}{\mu}\left(\Sigma x_i - \Sigma y_i\right) - \frac{1}{2\mu}\left(\Sigma x_i^2 - \Sigma y_i^2\right)\right]$$

$$= \exp\left[\left(\Sigma x_i - \Sigma y_i\right) - \frac{1}{2\mu}\left(\Sigma x_i^2 - \Sigma y_i^2\right)\right]$$

The ratio is constant as a function of μ iff $\Sigma x_i^2 = \Sigma y_i^2$. Thus Σx_i^2 is minimal sufficient in the $n(\mu,\mu)$ family. However,

$T(x) = \left(\Sigma x_i, \Sigma x_i^2\right)$ is not a function of Σx_i^2. By Definition 6.2.11,

$T(x)$ is not minimal sufficient in the $n(\mu,\mu)$ family

although $T(x)$ is sufficient.

#6.25

(c) The statistic $(\Sigma x_i, \Sigma x_i^2)$ is minimal sufficient in the $n(\mu, \mu^2)$ family.

(soln) From eqn (1), we can obtain

$$\frac{f(x|\mu,\mu^2)}{f(y|\mu,\mu^2)} = \exp\left[\frac{\mu}{\mu^2}\left(\Sigma x_i - \Sigma y_i\right) - \frac{1}{2\mu^2}\left(\Sigma x_i^2 - \Sigma y_i^2\right)\right]$$

$$= \exp\left[\frac{1}{\mu}\left(\Sigma x_i - \Sigma y_i\right) - \frac{1}{2\mu^2}\left(\Sigma x_i^2 - \Sigma y_i^2\right)\right]$$

The ratio is constant as a function of μ iff $\Sigma x_i = \Sigma y_i$ and $\Sigma x_i^2 = \Sigma y_i^2$. Thus $T(x) = (\Sigma x_i, \Sigma x_i^2)$ is a minimal sufficient statistic in the $n(\mu, \mu^2)$.

(d) The statistic $(\Sigma x_i, \Sigma x_i^2)$ is minimal sufficient in the $n(\mu, \sigma^2)$ family

From the eqn (1), we can notice that the ratio is constant as a function of (μ, σ^2) iff $\Sigma x_i = \Sigma y_i$ and $\Sigma x_i^2 = \Sigma y_i^2$. Thus, $T(x) = (\Sigma x_i, \Sigma x_i^2)$ is minimal sufficient statistics for (μ, σ^2) in the $n(\mu, \sigma^2)$ family.

#6.40 Suppose $X_1, X_2, \cdots, X_n = \{X_i\}_1^n \overset{iid}{\sim}$ a location scale family (LSF)

Let $T_1(\{X_i\}_1^n)$ and $T_2(\{X_i\}_1^n)$ satisfy

$$T_i(ax_1+b, \cdots, ax_n+b) = aT_i(x_1, \cdots, x_n) \text{ for all values of } x_1, \cdots, x_n \text{ and any } a>0 \text{ and } b.$$

(a) Show that $\frac{T_1}{T_2}$ is an ancillary statistic.

Let $X_i \overset{iid}{\sim} f(x|\mu, \sigma)$. Then (LSF) $Z \overset{d}{=} \frac{X-\mu}{\sigma} \Leftrightarrow X \overset{d}{=} \sigma Z + \mu$

$$\frac{T_1(\{X_i\}_1^n)}{T_2(\{X_i\}_1^n)} = \frac{T_1(\sigma Z_1+\mu, \cdots, \sigma Z_n+\mu)}{T_2(\sigma Z_1+\mu, \cdots, \sigma Z_n+\mu)}$$

$$= \frac{\cancel{\sigma} T_1(\{Z_i\}_1^n)}{\cancel{\sigma} T_2(\{Z_i\}_1^n)} = \frac{T_1(\{Z_i\}_1^n)}{T_2(\{Z_i\}_1^n)}$$

Since $\frac{T_1}{T_2}$ is a function of $\{Z_i\}_1^n$, so the distribution of $\frac{T_1}{T_2}$ does not depend on μ and σ. Therefore, $\frac{T_1}{T_2}$ is an ancillary statistic.

(b) Let R be the sample range and S be the sample standard deviation. Verify that R and S satisfy the above condition so that R/S is an ancillary statistic.

$R(\{x_i\}_1^n)$ = sample range = $\max(\{x_i\}_1^n) - \min(\{x_i\}_1^n) = x_{(n)} - x_{(1)}$

for $a>0$,

$$R(ax_1+b, \cdots, ax_n+b) = [a x_{(n)} + \cancel{b}] - [a x_{(1)} + \cancel{b}] = a(x_{(n)} - x_{(1)}) = aR$$

$S(\{x_i\}_1^n)$ = sample standard deviation $= \sqrt{\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2}$

$$S^2(ax_1+b, \cdots, ax_n+b) = \frac{1}{n-1}\sum_{i=1}^n [(ax_i+\cancel{b}) - (a\bar{x}+\cancel{b})]^2 = a^2 \cdot \underbrace{\frac{1}{n-1}\sum(x_i-\bar{x})^2}_{S^2(\{x_i\}_1^n)} = a^2 S^2(\{x_i\}_1^n).$$

Thus, R and S satisfy the above condition. It leads to

$$\frac{R}{S} = \frac{\cancel{a}(x_{(n)} - x_{(1)})}{\cancel{a}\, S(x_1, \cdots, x_n)} = \frac{R(x_1, \cdots, x_n)}{S(x_1, \cdots, x_n)}$$ is an ancillary

statistic by the resuet of the part (a).