STAT 661
Shen Qu

HW1

1.5-4

(a). Show that T1 and T2 are equivalent statistics if, and only if, we can write T2 = H(T1) for some 1-1 transformation H of the range of T1 into the range of T2. Which of the following statistics are equivalent? (Prove or disprove.)

If $T_2 = H(T_1)$ for some 1-1 transformation H of the range of T_1 into the range of T_2, then

when $T_1(x) = T_1(y)$, $T_2(x) = H(T_1(x)) = H(T_1(y)) = T_2(y)$;

when $T_2(x) = T_2(y)$, $H(T_1(x)) = T_2(x) = T_2(y) = H(T_1(y))$; then T_1 and T_2 are equivalent.

If T_1 and T_2 are equivalent, then $\exists H$ make $T_2 = H(T_1)$ is a 1-1 transformation of the range of T_1 into the range of T_2.

Therefore, T_1 and T_2 are equivalent statistics $\iff T_2 = H(T_1)$.

(b). $\prod_{i=1}^n x_i$ and $\sum_{i=1}^n \log x_i$, $x_i > 0$

$T_2(x) = \sum_{i=1}^n \ln x_i = \ln(\prod_{i=1}^n x_i) = \ln(T_1)$, $x_i > 0$. $H(x) = \ln x$ is a 1-1 transformation of $T_1 \in (0, \infty)$ into $T_2 \in (-\infty, \infty)$.

Thus, T_1 and T_2 are equivalent.

(c). $\sum_{i=1}^n x_i$ and $\sum_{i=1}^n \log x_i$, $x_i > 0$

$T_2(x) = \sum_{i=1}^n \ln x_i = T_1(\ln(x)) \neq T_1(x)$, $x_i > 0$. Thus, T_1 and T_2 are not equivalent.

(d). $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ and $(\sum_{i=1}^n x_i, \sum_{i=1}^n (x_i - \bar{x})^2)$

Let $T_1 = (T_{11} = \sum_{i=1}^n x_i, T_{12} = \sum_{i=1}^n x_i^2)$, then

$$T_{21} = \sum_{i=1}^n x_i = T_{11}$$

$$T_{22} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n(\bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{2}{n}(\sum_{i=1}^n x_i)^2 + \frac{1}{n}(\sum_{i=1}^n x_i)^2 = T_{12} - \frac{1}{n}T_{11}^2$$

Thus, T_1 and T_2 are equivalent.

(e). $(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^3)$ and $(\sum_{i=1}^n x_i, \sum_{i=1}^n (x_i - \bar{x})^3)$

Let $T_1 = (T_{11} = \sum_{i=1}^n x_i, T_{12} = \sum_{i=1}^n x_i^3)$, then

$$T_{21} = \sum_{i=1}^n x_i = T_{11}$$

$$T_{22} = \sum_{i=1}^n (x_i - \bar{x})^3 = \sum_{i=1}^n x_i^3 - 3\bar{x}\sum_{i=1}^n x_i^2 + 3\bar{x}^2\sum_{i=1}^n x_i - n(\bar{x})^3 =$$

$$\sum_{i=1}^{n} x_i^3 - \frac{3}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 + \frac{3}{n^2}(\sum_{i=1}^{n} x_i)^3 - \frac{1}{n^2}(\sum_{i=1}^{n} x_i)^3 = T_{12} - \frac{3}{n}T_{11}\sum_{i=1}^{n} x_i^2 + \frac{2}{n^2}T_{11}^3$$

There is not statistics in T_1 can represent $\sum_{i=1}^{n} x_i^2$.

Thus, T_1 and T_2 are not equivalent.

1.5-6

Let X take on the specified values $v_1, .., v_k$ with probabilities $\theta_1, .., \theta_k$, respectively. Suppose that $X_1, .., X_n$ are independently and identically distributed as X. Suppose that $\theta = (\theta_1, .., \theta_k)$ is unknown and may range over the set $\Theta = \{(\theta_1, .., \theta_k) : \theta_i \geq 0, 1 \leq i \leq k, \sum_{i=1}^{k} \theta_i = 1\}$. Let N_j be the number of X_i which equal v_j.

(a). What is the distribution of $(N_1, .., N_k)$?

$(N_1, .., N_k)$ $\sim$Multinomial Distribution

$f_{\vec{\theta}}(\vec{n}) = n! \prod_{i=1}^{k} \frac{\theta_i^{n_i}}{n_i!} \mathbf{1}_{\{\sum n_i = n\}}$, where n_i =the number of times we get outcome $i = 1, .., k$

(b). Show that $\mathbf{N} = (N_1, .., N_{k-1})$ is sufficient for θ.

$f_{\vec{\theta}}(\vec{n}) = n! \prod_{i=1}^{k} (n_i!)^{-1} \exp[\sum_{i=1}^{k} n_i \ln \theta_i] \mathbf{1}_{\{\sum n_i = n\}} = h(\vec{n}) \exp[\sum_{i=1}^{k} \eta_i(\vec{\theta}) T_i(\vec{n}) - B(\vec{\theta})]$, where $\chi = \{\vec{x} \in \{0, .., n\}^k | \sum x_i = n\}$

$h(\vec{n}) = n! \prod_{i=1}^{k} (n_i!)^{-1}$, $B(\vec{\theta}) = 0$

$\eta_i(\vec{\theta}) = (\ln \theta_1, .., \ln \theta_k)$,

$T(\vec{n}) = (n_1, .., n_k)$ is a n.s.s of the family.

$T(\vec{n}) = (n_1, .., n_{k-1}, n - \sum_{i=1}^{k-1} n_i)$ is equivalent with $(N_1, .., N_{k-1})$. Therefore $\mathbf{N}$ is sufficient for θ.

1.5-7

Let $X_1, .., X_n$ be a sample from a population with density $p(x, \theta)$ given by $p(x, \theta) = \begin{cases} \frac{1}{\sigma} \exp\{-\frac{x-\mu}{\sigma}\} & if x \geq \mu \\ 0 & o.w. \end{cases}$

Here $\theta = (\mu, \sigma)$ with $-\infty < \mu < \infty, \sigma > 0$.

(a) Show that $\min(X_1, .., X_n)$ is sufficient for μ when σ is fixed.

When σ is fixed, $p(x_{1:n}, \mu) = \sigma^{-n} \exp[-\sum_{i=1}^{n} x/\sigma] \exp[n\mu/\sigma] \prod_{i=1}^{n} \mathbf{1}_{\{x \geq \mu\}}$, where

$h(x) = \sigma^{-n} \exp[-\sum_{i=1}^{n} x/\sigma]$, $g(T(x), \mu) = \exp[n\mu/\sigma] \prod_{i=1}^{n} \mathbf{1}_{\{x_i \geq \mu\}}$

$\mathbf{1}_{\{x_{(1)} \geq \mu\}}$ contains all the information about μ, then

$T(x) = \min(X_1, .., X_n)$ is sufficient for μ when σ is fixed.

- Another method is that $p(x_{1:n}|t)$ is free of μ

$X \sim Expo(\mu, 1/\sigma)$, $F_{\mu,\sigma}(x) = 1 - e^{-(x-\mu)/\sigma}$,

$\min(X_1, .., X_n) = X_{(1)} = n\frac{1}{\sigma}e^{-(x-\mu)/\sigma}[1 - (1 - e^{-(x-\mu)/\sigma})]^{n-1} = \frac{n}{\sigma}e^{-n(x-\mu)/\sigma}$

$p(x_{1:n}|t) = \frac{1}{n\sigma^{n-1}} e^{\frac{1}{\sigma}(\sum x_i - nx)}$ is free of μ

(b) Find a one-dimensional sufficient statistic for σ when μ is fixed.

When μ is fixed, $p(x_{1:n}, \sigma) = \sigma^{-n} \exp[-(\sum_{i=1}^n x - n\mu)/\sigma] \prod_{i=1}^n \mathbf{1}_{\{x_i \geq \mu\}}$, where
$h(x) = \prod_{i=1}^n \mathbf{1}_{\{x \geq \mu\}}$, $g(T(x), \sigma) = \sigma^{-n} \exp[-(\sum_{i=1}^n x_i - n\mu)/\sigma]$, then
$T(x) = \sum_{i=1}^n x_i$ is sufficient for μ when σ is fixed.

- Another method is that $p(x_{1:n}|t)$ is free of σ

$X \sim Expo(\mu, 1/\sigma)$, $F_{\mu,\sigma}(x) = 1 - e^{-(x-\mu)/\sigma}$,
$Y = X - \mu \sim Exp(1/\sigma)$, $T = \sum Y_i \sim Gamma(n, \sigma)$
$p(x_{1:n}|t) = \Gamma(n)t^{1-n}$ is free of σ

(c) Exhibit a two-dimensional sufficient statistic for θ.

$p(x_{1:n}, \mu, \sigma) = \sigma^{-n} \exp[-(\sum_{i=1}^n x - n\mu)/\sigma] \prod_{i=1}^n \mathbf{1}_{\{x_i \geq \mu\}}$, where
$h(x) = 1$, $g(T(x), \mu, \sigma) = \sigma^{-n} \exp[-(\sum_{i=1}^n x_i - n\mu)/\sigma] \prod_{i=1}^n \mathbf{1}_{\{x_{(1)} \geq \mu\}}$, then
$T(x) = (x_{(1)}, \sum_{i=1}^n x_i)$ is a two-dimensional sufficient statistic for θ.

1.5-9

Let $X_1, .., X_n$ be a sample from a population with density $f_\theta(x) = \begin{cases} a(\theta)h(x) & if \theta_1 \leq x \leq \theta_2 \\ 0 & o.w. \end{cases}$ where $h(x) \geq 0$, $\theta = (\theta_1, \theta_2)$ with $-\infty < \theta_1 \leq \theta_2 < \infty$, and $a(\theta) = [\int_{\theta_1}^{\theta_2} h(x)dx]^{-1}$ is assumed to exist. Find a two-dimensional sufficient statistic for this problem and apply your result to the $U[\theta_1, \theta_2]$ family of distributions.

Let $H'(x) = h(x)$, $a(\theta) = [\int_{\theta_1}^{\theta_2} h(x)dx]^{-1} = [H(\theta_2) - H(\theta_1)]^{-1}$
$f_{\theta_1,\theta_2}(x_{1:n}) = \prod_{i=1}^n a(\theta)h(x) = \mathbf{1}_{\{x \in [\theta_1,\theta_2]\}}[H(\theta_2) - H(\theta_1)]^{-n} \prod_{i=1}^n h(x)$, where
$g(T(x), \theta_1, \theta_2) = \mathbf{1}_{\{x \in [\theta_1,\theta_2]\}}[H(\theta_2) - H(\theta_1)]^{-n}$, $h'(x) = \prod_{i=1}^n h(x)$
$\mathbf{1}_{\{x_{(n)} \leq \theta_2)\}} \mathbf{1}_{\{x_{(1)} \geq \theta_2)\}}$ contains all the information about μ, then
$T(x) = (x_{(1)}, x_{(n)})$ is a two-dimensional sufficient statistic for θ.
For $U[\theta_1, \theta_2]$, let $h(x) = 1$, $a(\theta) = (\theta_2 - \theta_1)^{-1}$
$f_{\theta_1,\theta_2}(x_{1:n}) = \prod_{i=1}^n a(\theta)h(x) = \mathbf{1}_{\{x \in [\theta_1,\theta_2]\}}[\theta_2 - \theta_1]^{-n} \prod_{i=1}^n 1$, where
$g(T(x), \theta_1, \theta_2) = \mathbf{1}_{\{x \in [\theta_1,\theta_2]\}}[\theta_2 - \theta_1]^{-n}$, $h'(x) = 1$
$T(x) = (x_{(1)}, x_{(n)})$ is a two-dimensional sufficient statistic for θ in the $U[\theta_1, \theta_2]$ family.