MANA130353.01: Statistic Inference **1st Semester, 2022-2023**

Homework 3

Date: October 14, 2022 *Scribes: Chaij*

Note: This note is a reference answer for the homework.

Disclaimer: *This note is only used as a reference solution for the homework, and the solution to each question is not unique. If you have any questions, you can add my WeChat ID statchaij, or you can find my lecture video here.*

Problem 1 (7′)**.** *Let $X_1, \ldots, X_n$ be independent random variables with densities*

$$f_{X_i}(x \mid \theta) = e^{i\theta - x}\mathbb{I}(x \geq i\theta).$$

Prove that $T = \min_i (X_i/i)$ is a sufficient statistic for θ.

Pf: The jpdf of $(X_1, \cdots, X_n)$ is

$$\begin{aligned}
f(x_1, \cdots, x_n|\theta) &= \prod_{i=1}^{n} e^{i\theta - x}\mathbb{I}(x_i > i\theta) = \exp\left\{\sum_{i=1}^{n}(i\theta - x_i)\right\}\prod_{i=1}^{n}\mathbb{I}\left(\frac{x_i}{i} > \theta\right) \\
&= \exp\left\{-\sum_{i=1}^{n} x_i\right\}\exp\left\{\frac{n(n+1)}{2}\theta\right\}\mathbb{I}\left(\bigcap_{i=1}^{n}\frac{x_i}{i} > \theta\right) \\
&= \exp\left\{-\sum_{i=1}^{n} x_i\right\}\exp\left\{\frac{n(n+1)}{2}\theta\right\}\mathbb{I}\left(\min_i \frac{x_i}{i} > \theta\right) \\
&= \exp\left\{-\sum_{i=1}^{n} x_i\right\}\exp\left\{\frac{n(n+1)}{2}\theta\right\}\mathbb{I}\left(T > \theta\right) \\
&= h(x) * g(T, \theta)
\end{aligned}$$

By factorization criterion we know $T = \min_i (X_i/i)$ is a sufficient statistic for θ. □

Problem 2 (8′)**.** *Suppose $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, show that $\bar{X}$ and $X_{(n)} - X_{(1)}$ are independent.*

Pf: First we fix $\sigma > 0$, the jpdf of $(X_1, \cdots, X_n)$ is

$$f(x_1, \cdots, x_n|\mu) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x_i - \mu)^2}{2\sigma^2}\} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2} + \frac{n\mu\bar{x}}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\}$$

By factorization criterion we have $\bar{X}$ is sufficient for μ, and since the parameter space $\Theta_0 = \{\mu \in \mathbb{R}\}$ is open thus it has interior point, then $\frac{n\bar{X}}{\sigma^2}$ is complete. Because there is a one-to-one correspondence between $\frac{n\bar{X}}{\sigma^2}$ and $\bar{X}$, $\bar{X}$ is also complete.

Let $Y_i = X_i - \mu \sim N(0, \sigma^2)$. and

$$X_{(n)} - X_{(1)} = (X_{(n)} - \mu) - (X_{(1)} - \mu) = Y_{(n)} - Y_{(1)}.$$

Since the distribution of $(Y_1, \cdots, Y_n)$ is independent of μ, so is $Y_{(n)} - Y_{(1)}$. Hence the distribution of $X_{(n)} - X_{(1)}$ is independent of μ, that is, $X_{(n)} - X_{(1)}$ is ancillary.

Then by Basu Thm we have $\bar{X}$ and $X_{(n)} - X_{(1)}$ are independent.

Finaly, since the given σ is arbitrary, we can conclude $\forall \sigma > 0, \mu \in \mathbb{R}$, $\bar{X}$ and $X_{(n)} - X_{(1)}$ are independent. $\square$

Problem 3 (7′). *Let $X_1, \ldots, X_n$ be a random sample from a population with location pdf $f(x-\theta)$ where f has common support. Show that the order statistics, $T(X_1, \ldots, X_n) = \left(X_{(1)}, \ldots, X_{(n)}\right)$, are a sufficient statistic for θ and no further reduction is possible.*

Pf: **Step 1.** First we show the order statistics is sufficient for general location family. The jpdf of $(X_1, \cdots, X_n)$ is

$$f(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} f(x_i - \theta) = \prod_{i=1}^{n} f\left(x_{(i)} - \theta\right).$$

Hence $T(X) = \left(X_{(1)}, \ldots, X_{(n)}\right)$ is sufficient for θ by the factorization criterion.

Step 2. Then we show the order statistics is minimal sufficient statistic for logistic distribution family with pdf:

$$f(x, \theta) = \frac{\exp\{-(x-\theta)\}}{\left(1 + \exp\{-(x-\theta)\}\right)^2}, \Theta = \mathbb{R}.$$

The jpdf of $(X_1, \cdots, X_n)$ is

$$f(x_1, x_2, \cdots, x_n | \theta) = \prod_{i=1}^{n} \frac{\exp\{-(x_i - \theta)\}}{\left(1 + \exp\{-(x_i - \theta)\}\right)^2}$$

The sufficiency is easily by factorization criterion. Then we want to show for two samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, the ratio

$$\frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{y} \mid \theta)} = \prod_{i=1}^{n} \frac{f(x_i - \theta)}{f(y_i - \theta)} = c(x, y) \Leftrightarrow T(X) = T(Y)$$

Taking logarithm of this ratio, we obtain:

$$\sum_{i=1}^{n}\left[-x_i+y_i+2\log\left(1+e^{-(y_i-\theta)}\right)-2\log\left(1+e^{-(x_i-\theta)}\right)\right]=c' \quad ,\forall\theta\in\mathbb{R}.$$

Since LHS is a constant function of θ, we take the partial derivative of θ on both sides of the equation:

$$\sum_{i=1}^{n}\left[\frac{\exp\{-(y_i-\theta)\}}{1+\exp\{-(y_i-\theta)\}}-\frac{\exp\{-(x_i-\theta)\}}{1+\exp\{-(x_i-\theta)\}}\right]=0 \quad ,\forall\theta\in\mathbb{R},$$

i.e.

$$\sum_{i=1}^{n}\left[\frac{1}{1+e^{y_i-\theta}}-\frac{1}{1+e^{x_i-\theta}}\right]=0\Leftrightarrow\sum_{i=1}^{n}\frac{1}{1+e^{y_i-\theta}}=\sum_{i=1}^{n}\frac{1}{1+e^{x_i-\theta}} \quad ,\forall\theta\in\mathbb{R}.$$

Do not forget our goal is to show this result holds iff $T(X)=T(Y)$. Let $c=e^{-\theta}>0, a_i=e^{y_i}>0, b_i=e^{x_i}>0$. We assume $a_1\leq a_2\leq\cdots\leq a_n, b_1\leq b_2\leq\cdots\leq b_n$ w.l.o.g.. The above equation is equal to

$$\sum_{i=1}^{n}\frac{1}{1+ca_i}=\sum_{i=1}^{n}\frac{1}{1+cb_i}\Leftrightarrow\frac{\sum_{i=1}^{n}\prod_{j\neq i}(1+ca_j)}{\prod_{i=1}^{n}(1+ca_i)}=\frac{\sum_{i=1}^{n}\prod_{j\neq i}(1+cb_j)}{\prod_{i=1}^{n}(1+cb_i)} \quad ,\forall c\in\mathbb{R}^+.$$

Then $\forall c\in\mathbb{R}^+$ we have

$$\sum_{i=1}^{n}\prod_{j\neq i}(1+ca_j)\prod_{i=1}^{n}(1+cb_i)=\sum_{i=1}^{n}\prod_{j\neq i}(1+cb_j)\prod_{i=1}^{n}(1+ca_i)$$

$$\sum_{i=1}^{n}\prod_{j\neq i}(1+ca_j)(1+cb_j)(1+cb_i-1-ca_i)=c\sum_{i=1}^{n}\prod_{j\neq i}(1+ca_j)(1+cb_j)(b_i-a_i)=0.$$

Suppose that there is $k\in\{1,2,\cdots,n\}$ in the above equation such that $a_k\neq b_k$. Then

$$\sum_{i\neq k}^{n}\prod_{j\neq i}(1+ca_j)(1+cb_j)(b_i-a_i)=-\prod_{j\neq k}(1+ca_j)(1+cb_j)(b_k-a_k) \quad ,\forall c\in\mathbb{R}^+.$$

$$\sum_{i\neq k}^{n}\frac{\prod_{j\neq i}(1+ca_j)(1+cb_j)(b_i-a_i)}{\prod_{j\neq k}(1+ca_j)(1+cb_j)(b_k-a_k)}=\sum_{i\neq k}^{n}\frac{(1+ca_k)(1+cb_k)(b_i-a_i)}{(1+ca_i)(1+cb_i)(b_k-a_k)}=-1 \quad ,\forall c\in\mathbb{R}^+.$$

Since the middle of the above equation is the function of c, but the right hand side of the equation is a constant, we can conclude that $a_i=b_i,\forall i\neq k$, but $0\neq-1$ which is a contradiction. Thus we have $a_i=b_i, i=1,2,\cdots,n$, that is, $X_{(i)}=Y_{(i)}, i=1,2,\cdots,n$.

Since then we have shown

$$\frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{y} \mid \theta)}=\prod_{i=1}^{n} \frac{f\left(x_{i}-\theta\right)}{f\left(y_{i}-\theta\right)}=c(x, y) \Leftrightarrow T(X)=\left(X_{(1)}, \ldots, X_{(n)}\right)=\left(Y_{(1)}, \ldots, Y_{(n)}\right)=T(Y)$$

So we know the order statistics is minimal sufficient statistic for logistic distribution family with location parameter.

Step 3. Finally we show the order statistics is minimal sufficient for general location family. Now we know $T(X)$ is sufficient for the general location fmaily $\mathcal{P}$ and is minimal sufficient for logistic distribution family $\mathcal{P}_0$. Obviously, $\mathcal{P}_0 \subset \mathcal{P}$.

By the definition of the minimal sufficient statistic, for any sufficient statistic $U(X)$ for $\mathcal{P}_0$, there is a function H_1 such that $T(X)=H_1(U(X))$. Now for any sufficient statistic $S(X)$ for $\mathcal{P}$, we know it is also sufficient for $\mathcal{P}_0$, and then there is a function H_2 such that $T(X)=H_2(S(X))$. Since $T(X)$ is sufficient for $\mathcal{P}$, we conclude $T(X)=\left(X_{(1)}, \ldots, X_{(n)}\right)$ is also minimal sufficient for $\mathcal{P}$.

$\square$

Bonus 1: Suppose $X_1, \ldots, X_n \stackrel{iid}{\sim} f(x, \theta)=\frac{1}{2} e^{-|x-\theta|}, -\infty<x<\infty, -\infty<\theta<\infty$. Show that $T(X)=\left(X_{(1)}, \ldots, X_{(n)}\right)$ is a minimal sufficient statistic for θ.
Bonus 2: Suppose $X_1, \ldots, X_n \stackrel{iid}{\sim} f(x, \theta)=\frac{1}{\pi} \frac{1}{1+(x-\theta)^2}, -\infty<x<\infty, -\infty<\theta<\infty$. Show that $T(X)=\left(X_{(1)}, \ldots, X_{(n)}\right)$ is a minimal sufficient statistic for θ.

Problem 4 $(8')$. *Suppose X follows a discrete distribution with the following pmf:*

$$P(X=0)=p, \quad P(X=1)=3p, \quad P(X=2)=1-4p, \quad 0<p<1/4.$$

Is the family of distributions of X complete? What about the following family:

$$P(X=0)=p, \quad P(X=1)=p^2, \quad P(X=2)=1-p-p^2, \quad 0<p<1/2$$

Sol: For the first distribution family, let f be a function such that $\mathbb{E}_p f(X)=0$. Then

$$0=\mathbb{E}_p f(X)=pf(0)+3pf(1)+(1-4p)f(2)=(f(0)+3f(1)-4f(2))p+f(2), \quad \forall 0<p<1/4.$$

We find if $f(0)=-3f(1) \neq 0$ and $f(2)=0$ then $\mathbb{E}_p f(X)=0, \forall 0<p<1/4$. Now $P(f(X)=0)=4p \in(0,1)<1$. Thus the distribution family is not complete.

For the second distribution family, let f be a function such that $\mathbb{E}_p f(X)=0$. Then

$$0=\mathbb{E}_p f(X)=pf(0)+p^2 f(1)+\left(1-p-p^2\right) f(2)=(f(1)-f(2))p^2+(f(0)-f(2))p+f(2)$$

for all $0 < p < 1/2$. This is a polynomial of degree 2 in in p. To make it zero for all $p \in (0, \frac{1}{2})$, each coefficient must be zero. Thus,

$$f(0) = f(1) = f(2) = 0,$$

Now $P(f(X) = 0) = 1$. Thus, this distribution family is complete.

$\square$