Statistical Inference Assignment 3

Junhao Yuan (20307130129)

October 10, 2022

Problem 1.

Let $X_1, \ldots, X_n$ be independent random variables with densities

$$f_{X_i}(x|\theta) = e^{i\theta - x}\mathbb{I}(x \geq i\theta). \tag{1}$$

Prove that $T = \underset{i}{min}(X_i/i)$ is a sufficient statistics for θ.

Solution.

The joint pdf of the samples X is

$$\begin{aligned} f_X(\mathbf{x}|\theta) &= \prod_{i=1}^{n} e^{i\theta - x_i}\mathbb{I}(x_i \geq i\theta) \\ &= exp\left(\frac{n(n+1)}{2}\theta - \sum_{i=1}^{n} x_i\right)\mathbb{I}(T(\mathbf{x}) \geq 0) \\ &= g(T(\mathbf{x})|\theta)h(\mathbf{x}), \end{aligned}$$

where

$$\begin{aligned} g(T(\mathbf{x})|\theta) &= e^{\frac{n(n+1)}{2}\theta}\mathbb{I}(T(\mathbf{x}) \geq 0), \\ h(\mathbf{x}) &= exp^{-\sum_{i=1}^{n} x_i}. \end{aligned}$$

Here, we use the fact that $i > 0$ and $x_i\theta \geq i\theta$ (for $1 \leq i \leq n$) if and only if $min_i(x_i/i) \geq \theta$. Thus, by the factorization theorem, we have $T(X) = min_i(X_i/i)$ is a sufficient statistics.

Problem 2.

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

Solution.

First consider σ^2 fixed and let μ vary, $-\infty < \mu < \infty$. It's obvious that $\bar{X}$ is a sufficient statistics for μ. Since the pdf of $\bar{X}$ is (σ is known)

$$\begin{aligned} f_{\bar{X}}(\bar{x}) &= \sqrt{\frac{n}{2\pi}}\frac{1}{\sigma}e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} \\ &= \sqrt{\frac{n}{2\pi}}e^{-\frac{n\bar{x}^2}{2\sigma^2}} \cdot \frac{1}{\sigma}e^{-\frac{n\mu^2}{2\sigma^2}} \cdot e^{\frac{n\mu}{\sigma^2}\bar{x}}, \end{aligned}$$

which has an open set in $\mathbb{R}$, we can know that $\bar{x}$ is complete. Thus, $\bar{X}$ is a complete and sufficient statistics for μ. Since $X_{(n)} - X_{(1)}$ is an ancillary statistics for μ when σ is known, by the Basu's theorem, we know that $\bar{X}$ and $X_{(n)} - X_{(1)}$ are independent. But since σ is arbitrary, we have that $\bar{X}$ and $X_{(n)} - X_{(1)}$ are independent for any choice of μ and σ.

PROBLEM 3.

Let $X_1, \dots, X_n$ be a random sample from a population with location pdf $f(x-\theta)$. Show that the order statistics, $T(X_1, \dots, X_n) = (X_{(1)}, \dots, X_{(n)})$, are a sufficient Statistics for θ and no further reduction is possible.

SOLUTION.

The pdf of X is

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^n f(x_i|\theta) \\ &= \prod_{i=1}^n f(x_{(i)}|\theta), \end{aligned}$$

which shows that $T(X) = (X_{(1)}, \dots, X_{(n)})$ is a sufficient statistics. Consider two samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$, then the ratio is

$$\begin{aligned} \frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{\prod_{i=1}^n f(x_i|\theta)}{\prod_{j=1}^n f(y_j|\theta)} \\ &= \prod_{i=1}^n \frac{f(x_{(i)}|\theta)}{f(y_{(i)}|\theta)}. \end{aligned}$$

It's obvious that the ratio is a constant function of θ is and only if $T(X) = T(Y)$. Thus, $T(X)$ is a minimal sufficient for θ, which means no further reduction is possible without further restrictions on f.

PROBLEM 4.

Suppose X follows a discrete distribution with the following pmf:

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

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

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

SOLUTION.

Let g be a function such that $\mathbb{E}_p g(X) = 0$. Then

$$\begin{aligned} 0 = \mathbb{E}_p g(X) &= pg(0) + 3pg(1) + (1-4p)g(2) \\ &= (g(0) + 3g(1) - 4g(2)) + g(2), \end{aligned}$$

for all $0 < p < 1/4$. Thus, we have

$$\begin{aligned} g(2) &= 0 \\ g(0) + 3g(1) - 4g(2) &= 0, \end{aligned}$$

which implies that $g(X)$ need to statistics that $g(2) = 0$ and $g(0) + 3g(1) = 0$. Thus, $g(X)$ don't have to be identically zero,and this family of distribution of X is not complete.

For the second distribution, we have

$$\begin{aligned} 0 = \mathbb{E}_p g(X) &= pg(0) + p^2 g(1) + (1 - p - p^2) g(2) \\ &= (g(1) - g(2))p^2 + (g(0) - g(2))p + g(2), \end{aligned}$$

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

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

or in other words,

$$P_p(g(X) = 0) = 1.$$

Thus, we have this family of distribution is complete.