Statistical Inference

B. Statistical Data Science 2nd Year   Indian Statistical Institute

Teacher: Soham Sarkar

Exercise Series 6

Exercise 1. Let $X_1, X_2 \overset{\text{i.i.d.}}{\sim} f_\theta$, where $f_\theta(x) = \theta x^{\theta-1} e^{-x^\theta}, \quad x > 0,\ \theta > 0$. Show that $\dfrac{\log X_1}{\log X_2}$ is ancillary.

Hint: *Notice that f_θ is the pdf of Y^θ, where $Y \sim \mathsf{Exponential}(1)$.*

Exercise 2. For the following cases, verify whether the minimal sufficient statistic is complete.

(a) $f_p = \mathsf{Bernoulli}(p)$, $p \in (0,1)$. $T = \sum_{i=1}^n X_i$.

(b) $f_\lambda = \mathsf{Poisson}(\lambda)$, $\lambda > 0$. $T = \sum_{i=1}^n X_i$.

(c) $f_p = \mathsf{Geometric}(p)$, $p \in (0,1)$. $T = \sum_{i=1}^n X_i$.

(d) $f_\theta = \mathsf{Uniform}(\theta, 1)$, $\theta < 1$. $T = X_{(1)}$.

(e) $f_\theta = \mathsf{Uniform}(\theta, \theta+1)$, $\theta \in \mathbb{R}$. $\mathbf{T} = \big(X_{(1)}, X_{(n)}\big)$.

Hint: *This is a distribution from the location family. What can you say about $X_{(n)} - X_{(1)}$?*

(f) $f_\sigma = \mathsf{Normal}(0, \sigma^2)$, $\sigma > 0$. $T = \sum_{i=1}^n X_i^2$.

(g) $f_{\mu,\sigma} = \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$. $\mathbf{T} = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

(h) $f_\sigma = \mathsf{Laplace}(0, \sigma)$, $\sigma > 0$. $T = \sum_{i=1}^n |X_i|$.

(i) $f_\theta = \mathsf{Normal}(\theta, \theta^2)$, $\theta \in \mathbb{R}$. $\mathbf{T} = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

(j) $f_\theta = \mathsf{Normal}(\theta, \theta)$, $\theta > 0$. $T = \sum_{i=1}^n X_i^2$.

(k) $\mathsf{Beta}(\alpha, \beta)$, $\alpha > 0, \beta > 0$. $\mathbf{T} = \left(\prod_{i=1}^n X_i, \prod_{i=1}^n (1 - X_i)\right)$.

(l) $\mathsf{Gamma}(\alpha, \lambda)$, $\alpha > 0, \lambda > 0$. $\mathbf{T} = \left(\sum_{i=1}^{n} X_i, \prod_{i=1}^{n} X_i\right)$.

(m) $\mathsf{Multinomial}(n; p_1, \ldots, p_{k-1})$, $p_1, \ldots, p_{k-1} \in (0,1), 0 < p_1 + \cdots + p_{k-1} < 1$.

$\mathbf{T} = \left(\sum_{i=1}^{n} X_{1i}, \ldots, \sum_{i=1}^{n} X_{k-1,i}\right)$.

Note: This is the *full rank* or *non-degenerate* version of the multinomial distribution. The pmf is given by

$$\mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = \frac{n!}{x_1! \cdots x_{k-1}!(n - x_1 - \cdots - x_{k-1})!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} (1 - p_1 - \cdots - p_{k-1})^{n - x_1 - \cdots - x_{k-1}},$$
$$0 \le x_1, \ldots, x_{k-1} \le n, 0 \le x_1 + \cdots + x_{k-1} \le n.$$

This is similar to the pmf we defined earlier, except p_k and x_k are missing, and we have $1 - p_1 - \cdots - p_{k-1}$ and $n - x_1 - \cdots - x_{k-1}$ in their places. Because of these changes, it is no longer true that $\sum_i p_i = 1$ and $\sum_i x_i = n$. This allows us to avoid *degeneracy/non-full-rankness*.

(n) $\mathsf{BVN}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, $\mu_x \in \mathbb{R}, \mu_y \in \mathbb{R}, \sigma_x > 0, \sigma_y > 0, \rho \in (-1, 1)$.

$\mathbf{T} = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} Y_i^2, \sum_{i=1}^{n} X_i Y_i\right)$.

Exercise 3. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Normal}(\theta, a\theta^2)$, where $a > 0$ is a known constant and $\theta \in \mathbb{R}$ is the parameter.

(a) Show that $\mathbf{T} = \left(\overline{X}, S^2\right)$ is minimal sufficient for θ.

(b) Show that the parameter space $\{(\theta, a\theta^2) : \theta \in \mathbb{R}\}$ does not contain a two-dimensional open set.

Hint: *Try to visualize the parameter space in* $\mathbb{R}^2$.

(c) Show that the minimal sufficient statistic $\mathbf{T}$ is not complete.

Hint: *Start with* $\mathbb{E}_\theta(\overline{X})$, $\mathbb{E}_\theta(\overline{X}^2)$, $\mathbb{E}_\theta(S^2)$.

Exercise 4. Suppose that the random variable X takes the values 0, 1 and 2 according to the following pmf

$$\mathbb{P}_\theta(X = 0) = \theta, \quad \mathbb{P}_\theta(X = 1) = 3\theta, \quad \mathbb{P}_\theta(X = 2) = 1 - 4\theta, \qquad 0 < \theta < \frac{1}{4}.$$

Verify whether X is complete.

Is X complete with the following pmf?

$$\mathbb{P}_\theta(X = 0) = \theta, \quad \mathbb{P}_\theta(X = 1) = \theta^2, \quad \mathbb{P}_\theta(X = 2) = 1 - \theta - \theta^2, \qquad 0 < \theta < \frac{1}{2}.$$

Exercise 5. As we mentioned in the class, completeness is a property of the entire family of distributions of a statistic. In particular, the parameter space plays a vital role in determining whether a statistic is complete. Let $X \sim \mathsf{Poisson}(\theta)$, where $\theta \in \{1, 2\}$. Show that X is not complete.

Note: We have already seen that if $T \sim \mathsf{Poisson}(n\theta)$, $\theta > 0$, then T is complete. Here, the completeness is lost when the parameter space is changed.

Exercise 6. Let $X_1, \dots, X_n$ be i.i.d. with pdf $f_\theta(x) = e^{-(x-\theta)}, \quad x > \theta,\ \theta \in \mathbb{R}$.

(a) Show that $X_{(1)}$ is complete sufficient for θ.

(b) Define $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ to be the sample variance based on $X_1, \dots, X_n$. Show that $X_{(1)}$ and S^2 are independent.

Hint: *Use Basu's theorem. The pdf f_θ is a member of the location family. S^2 can be written as*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{2n(n-1)}\sum_{1\le i,j\le n}(X_i - X_j)^2.$$

Exercise 7. Let $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Gamma}(\alpha, \theta)$, where $\alpha > 0$ is a known constant and $\theta > 0$ is the parameter.

(a) Show that $T = \sum_{i=1}^{n} X_i$ is sufficient for θ.

(b) Show that $T \sim \mathsf{Gamma}(n\alpha, \theta)$.

Hint: *Find the moment generating function/characteristic function of X_1. What is the characteristic function of $X_1 + \cdots + X_n$? Be careful about the domain of the moment generating function. It is not defined everywhere.*

(c) Use part (b) to show that T is complete.

(d) Verify that $\mathsf{Gamma}(\alpha, \theta)$ with known α is a scale family.

(e) Use part (d) to show that $\dfrac{X_{(i)}}{T}$ is an ancillary statistic.

(f) Use Basu's theorem to conclude that

$$\mathbb{E}\left(X_{(i)} \mid T\right) = \mathbb{E}\left(\frac{X_{(i)}}{T}\, T \,\Big|\, T\right) = \mathbb{E}\left(\frac{X_{(i)}}{T} \,\Big|\, T\right) T = \frac{\mathbb{E}\left(X_{(i)}\right)}{\mathbb{E}(T)}\, T.$$

Hint: *If X/Y and Y are independent, then*

$$\mathbb{E}(X) = \mathbb{E}\left(\frac{X}{Y}\, Y\right) = \mathbb{E}\left(\frac{X}{Y}\right)\mathbb{E}(Y) \quad \Rightarrow \quad \mathbb{E}\left(\frac{X}{Y}\right) = \frac{E(X)}{\mathbb{E}(Y)}.$$

Exercise 8. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. with pdf $f_\theta(x, y) = e^{-(\theta x + y/\theta)}, \quad x > 0, y > 0,\ \theta > 0$.
Define $T = \sqrt{\dfrac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}}$ and $U = \sqrt{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}$.

(a) Show that T and U are jointly sufficient for θ.

(b) Show that (T, U) is not complete.