Coursework - Fundamentals of Statistical Inference

CID: 02091191

December 1, 2021

Abstract

This document summarises the work conducted for the second Fundamentals of Statistical Inference coursework of the year.

Exercise 1

Let S be a complete sufficient statistic for a parameter $\theta \in \Omega_\theta$ and let C be a distribution constant statistic.

Since C is a distribution constant statistic, the distribution of C does not depend on θ. Moreover, as S is sufficient for $\theta \in \Omega_\theta$, the distribution of $C|S = s$ does not depend on θ either (S being sufficient for $\theta \in \Omega_\theta$ implies that the density function of $C|S = s$ verifies $f_{C|S}(c|s, \theta) = g(c, s)$).

For all event E in sample space of C, let $T(s) = \mathbb{P}[C \in E|S = s] - \mathbb{P}[C \in E]$. Note that

$$\forall \theta \in \Omega_\theta,\ \mathbb{E}_{S,\theta}[\mathbb{P}[C \in E|S]] = \mathbb{P}[C \in E]$$

This implies,

$$\forall \theta \in \Omega_\theta,\ \mathbb{E}_\theta[T(s)] = 0$$

Hence, as S is a complete statistic,

$$\mathbb{E}_\theta[T(s)] = 0 \implies \mathbb{P}_\theta[T(s) = 0] = 1$$

Thus, $T(s) = 0$ almost surely, which is equivalent by definition of T statistic above to $\mathbb{P}[C \in E|S = s] = \mathbb{P}[C \in E]$ almost surely. Therefore, C and S are independent for each $\theta \in \Omega_\theta$. $\square$

We have proved so far that considering S a complete sufficient statistic for a parameter $\theta \in \Omega_\theta$ and C a distribution constant statistic (ancillary statistic), S and C are independent for each $\theta \in \Omega_\theta$. This shows that it is irrelevant to proceed inference on θ for an ancillary statistic when there exists a complete sufficient statistic.

Exercise 2

Let Y be a random variable whose distribution is of multi-parameter exponential family form with density

$$f(y; \theta, \phi) = \exp\left\{\theta U(y) + \phi^T T(y) - \zeta(\theta, \phi)\right\} g(y)$$

where θ is a scalar parameter and ϕ is a vector of nuisance parameter. We wish to test H_0: $\theta \leq \theta_0$ against H_1: $\theta > \theta_0$. Suppose that there exists $V = h(U, T)$ independent of T when $\theta = \theta_0$ and increasing with respect to U for every fixed T.

According to the Lecture Notes [1] (p.50), UMPU test is given by rejecting H_0 if and only if $U > u^*$ where u^* is calculated from

$$\mathbb{P}_{\theta_0}[U > u^* | T_1 = t_1, \ldots, T_m = t_m] = \alpha$$

where $t_1, \ldots, t_m$ are observed values of $T(Y)$. Let fix $T = t$ where $t = (t_1, \ldots, t_m)$. Since V is (strictly) increasing in U,

$$U > u^* \iff V(U, t) > V(u^*, t)$$

Hence

$$\begin{aligned}
\mathbb{P}_{\theta_0}[U > u^* | T_1 = t_1, \ldots, T_m = t_m] = \alpha &\iff \mathbb{P}_{\theta_0}[V(U, t) > V(u^*, t) | T_1 = t_1, \ldots, T_m = t_m] = \alpha \\
&\iff \mathbb{P}_{\theta_0}[V(U, T) > V(u^*, T) | T_1 = t_1, \ldots, T_m = t_m] = \alpha \\
&\iff \mathbb{P}_{\theta_0}[V(U, T) > V(u^*, T)] = \alpha
\end{aligned}$$

as V and T are by assumption independent when $\theta = \theta_0$. Thus, there exists a UMPU test based on the marginal distribution of V.

The test function ϕ can be written as

$$\phi(y) = \begin{cases} 1 & \text{if } V > c \\ \gamma & \text{if } V = c \\ 0 & \text{if } V < c \end{cases}$$

$\square$

Exercise 3

Let $Y_1, \ldots, Y_n$ be independent and identically distributed samples from a normal distribution $N(\mu, \sigma^2)$ with both μ and σ^2 unknown. We wish to test H_0: $\mu \leq \mu_0$ against H_1: $\mu > \mu_0$.

Recall that the pdf of a $N(\mu, \sigma^2)$ distribution can be written as

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\}$$

It implies that $Y_1, \ldots, Y_n$ are samples from a full exponential family distribution with natural parameters $\frac{\mu}{\sigma^2}$, $-\frac{1}{2\sigma^2}$. We may also notice that this distribution is also member of a transformation family since $X \sim N(\mu, \sigma^2) \implies aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Considering the vector $y = (Y_1, \ldots, Y_n)^T$ of n observations, we have

$$\begin{aligned}
f(y; \mu, \sigma) &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}U + \frac{n\mu}{\sigma^2}\bar{Y} - \frac{n\mu^2}{2\sigma^2}\right\} \\
&\equiv \exp\left\{\theta\bar{Y} + \phi U - h(\theta, \phi)\right\}
\end{aligned}$$

where $U = \sum_{i=1}^{n} Y_i^2$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\theta = \frac{n\mu}{\sigma^2}$ and $\phi = -\frac{1}{2\sigma^2}$.

Recall that we wish to test H_0: $\mu \leq \mu_0$ against H_1: $\mu > \mu_0$. In fact, we may reduce the problem to considering $\mu_0 = 0$ by transforming all samples Y_i to $Y_i - \mu_0$ (i.e. center data). Hence, there is no loss of information in assuming $\mu_0 = 0$. Therefore, testing H_0: $\mu \leq \mu_0$ against H_1: $\mu > \mu_0$ is equivalent to testing H_0: $\theta \leq 0$ against H_1: $\theta > 0$ (according to the definition of θ above).

Moreover, the exponential family form of f implies that the UMPU test is based on the conditional distribution of $\bar{Y}$ given $T = t$ where $T = \sum_{i=1}^n Y_i^2$. Let then consider the statistic

$$V = \frac{\bar{Y}}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{U}{\sqrt{T - nU^2}} = V(U, T)$$

where $T = \sum_{i=1}^n Y_i^2$ and $U = \bar{Y}$. Note that V is equivalent to the Student t-statistic

$$t(y) = \frac{\sqrt{n}\bar{Y}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2}} \sim t_{n-1}$$

when $\theta = 0$ (equivalent to $\mu = 0$).

First, notice that V is an increasing function of U for every fixed T (this can be shown by calculating $\frac{\partial V}{\partial U}$ and notice trivially that this quantity is always positive).

V is independent of T when $\theta_1 = 0$ (i.e. $\mu = 0$): U is sufficient for $\theta_1 = \frac{n\mu}{\sigma^2}$ by factorisation theorem (according to the likelihood written above) and complete by exponential family properties. The distribution of $U|T = t$ is member of of exponential family and does not depend on $\theta_2 = -\frac{1}{2\sigma^2}$. Thus, according to the Lecture Notes [1] (p.35), T is ancillary for $\theta_1 = \frac{n\mu}{\sigma^2}$. Therefore, T is transformation invariant with respect to θ_1, i.e. T is distribution constant with respect to θ_1.

V is complete sufficient for $\theta_1 = 0$: since (U, T) is complete, $V(U, T)$ is complete. Moreover, for fixed $T = t$, $V(U, T)$ is a one-to-one transformation and since U is sufficient, V is sufficient as well.

Hence, by applying the result of Exercise (1): V is complete sufficient for $\theta_1 = 0$ and T is distribution constant implies that V and T are independent for all θ in sample space. In particular, V and T are independent for $\theta_1 = 0$.

Then, we apply the result of Exercise (2): $V(U, T)$ is independent of T when $\theta_1 = 0$ and increasing with respect to U for every fixed T implies that there exists a UMPU test of H_0 against H_1 based on the marginal distribution of V, i.e. the t-test of H_0 (as defined above with the t-statistic) is UMPU.

References

[1] Pr G.A. Young (October 2021) *MATH70078 - Fundamentals of Statistical Inference Lecture Notes*, Imperial College London MSc Statistics resources