UC Berkeley
Department of Statistics
Fall 2016

STAT 210A: Theoretical Statistics

Problem Set 2

Student: Ivana Malenica **Due:** Thursday, Sep. 15

Note: All measure-theoretic niceties about conditioning on measure-zero sets, almost-sure equality vs. actual equality, "all functions" vs. "all measurable functions," etc. are disregarded for the time being.

1. Minimal sufficiency for correlated normals

Suppose that $(X_i, Y_i), i = 1, \ldots, n$ are sampled i.i.d. from the bivariate normal distribution

$$(X, Y) \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \theta\\ \theta & 1\end{pmatrix}\right).$$

with $\theta \in \Theta = (-1, 1)$. (Note: this is an example of a curved exponential family)

(a) Find a two-dimensional minimal sufficient statistic and show it is minimal.

Solution:

Recall that the joint density of $(X_i, Y_i)_{i=1}^n$ defined as in the problem setup can be written as:

$$\frac{1}{2\pi\sqrt{1-\theta^2}} e^{-\frac{1}{2(1-\theta^2)}\sum_{i=1}^n (X_i^2+Y_i^2)+\frac{\theta}{1-\theta^2}\sum_{i=1}^n X_iY_i}$$

By the factorization theorem we clearly have that $T = (\sum_{i=1}^n (X_i^2 + Y_i^2), \sum_{i=1}^n X_iY_i)$ is a sufficient statistic. Since this is a two-dimensional curved exponential family, T will also be minimal sufficient. Alternatively, note that if for some (X,Y),(Z,W) the ratio is independent of θ, it goes to 1 as $\theta \to \infty$. Therefore, $C_{(X,Y),(Z,W)} = 1$ and $T_1(X, Y) - T_1(Z, W) = 2\theta(T_2(X, Y)T_2(Z, W))$. Thus, $T_1(X, Y) = T_1(Z, W)$ as $\theta \to 0$.

(b) Prove that the minimal sufficient statistic found in (a) is not complete.

Solution:

Define $f(a, b) = a$, $\forall a, b$. Note that $E_\theta(\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2) = 2n$. Then,

$$E_\theta(f(\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_iY_i)) = E_\theta(\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2) = 2n$$

However, function f is a.s. not a constant. Hence, minimal sufficient statistic found in (a) is not complete.

(c) Prove that $Z_1 = \sum_{i=1}^n X_i^2$ and $Z_2 = \sum_{i=1}^n Y_i^2$ are both ancillary, but that (Z_1, Z_2) is not ancillary.

Solution:

Since $Z_1 \sim \chi^2_n$ and $Z_2 \sim \chi^2_n$, they are not dependent on θ, and hence ancillary. The correlated bivariate chi-square distribution will depend on n and the covariance of Z_1 and Z_2. Recall that $(X_i, Y_i)_i \xrightarrow{d} (X_i, \theta X_i + \epsilon_i\sqrt{1-\theta^2})$ with $\epsilon_i \sim$ Normal(0,1) and $E(X^4) = 3$ for $X \sim$ Normal(0,1). Therefore,

$$\begin{aligned} Cov(Z_1, Z_2) &= \sum_{i=1}^{n} Cov(X_i^2, Y_i^2) \\ &= n(E_\theta X_i^2 Y_i^2 - E_\theta X_i^2 E_\theta Y_i^2) \\ &= nE_\theta X_i^2(\theta X_i + \epsilon_i\sqrt{1-\theta^2}) - n \\ &= n\theta^2 E(X_i^4) + n(1-\theta^2) - n \\ &= 2n\theta^2 \end{aligned}$$

Thus, the joint distribution of (Z_1, Z_2) is not ancillary.

2. Bayesian interpretation of sufficiency

Assume we have a family $\mathcal{P}$ of densities $p_\theta(x)$ with respect to a common measure μ on $\mathcal{X}$, for $\theta \in \Theta \subseteq \mathbb{R}^n$. Additionally, assume the parameter θ is itself random, following *prior density* $q(\theta)$ with respect to the Lebesgue measure on Θ.

Then, we can write the *posterior density* (distribution of θ given $X = x$) as

$$q_{\text{post}}(\theta \mid x) = \frac{p_\theta(x)q(\theta)}{\int_\Theta p_\zeta(x)q(\zeta)\,d\zeta}.$$

(Note: this is a measure-theoretically naïve manipulation of the densities, but one that "usually" works. Feel free to make such non-rigorous manipulations yourself in the problem).

In this setting, prove the following claims:

(a) Suppose a statistic $T(X)$ has the property that, for any prior distribution $q(\theta)$, the posterior distribution $q_{\text{post}}(\theta \mid x)$ depends on x only through $T(x)$. Show that $T(X)$ is sufficient for $\mathcal{P}$.

(b) Conversely, show that, if $T(X)$ is sufficient for $\mathcal{P}$ then, for any prior q, the posterior depends on x only through $T(x)$.

(c) Explain in your own words why part (b) gives a strong heuristic motivation for the sufficiency principle (Note: there isn't a unique correct answer to this part and, accordingly, the grading will be lenient).

3. Montonicity and minimal sufficiency

For a family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, we say that a property holds $\mathcal{P}$-a.s. if it holds almost surely under P_θ, for all $\theta \in \Theta$. Consider two nested families $\mathcal{P}_0 \subseteq \mathcal{P}$ such that any property holding $\mathcal{P}_0$-a.s. also holds $\mathcal{P}$-a.s. (i.e., events that are "impossible" in $\mathcal{P}_0$ do not become "possible" in the larger family).

(a) Prove that if T is sufficient for $\mathcal{P}$ and minimal sufficient for $\mathcal{P}_0$, it is also minimal sufficient for $\mathcal{P}$.

(b) Show that the condition above, that $\mathcal{P}_0$-a.s. implies $\mathcal{P}$-a.s., is necessary for the conclusion in (a) by giving a counterexample in the case of general nested families $\mathcal{P}_0 \subseteq \mathcal{P}$.

4. Ancillarity in location-scale families

In a parameterized family where $\theta = (\zeta, \lambda)$, we say a statistic T is *ancillary for* ζ if its distribution is independent of ζ; that is, if $T(X)$ is ancillary in the subfamily where λ is known, for each possible value of λ.

Suppose that $X_1, \ldots, X_n \in \mathcal{X} = \mathbb{R}$ are an i.i.d. sample from a *location-scale family* $\mathcal{P} = \{F_{a,b}(x) = F((x-a)/b) : \ a \in \mathbb{R}, b > 0\}$, where $F(\cdot)$ is a known cumulative distribution function. The real numbers a and b are called the *location* and *scale* parameters respectively.

(a) Show that the vector of differences $(X_1 - X_i)_{i=2}^n$ is ancillary for a.

(b) Show that the vector of ratios $\left(\frac{X_1-a}{X_i-a}\right)_{i=2}^n$ is ancillary for b. (Note: this is only a statistic when a is known).

(c) Show that the vector of difference ratios $\left(\frac{X_1-X_i}{X_2-X_i}\right)_{i=3}^n$ is ancillary for (a, b).

5. Uniform location family

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, with $\theta \in \mathbb{R}$ unknown.

(a) Find a two-dimensional minimal sufficient statistic and show it is minimal.

(b) Show that the minimal sufficient statistic is not complete sufficient.

(c) Suppose that we wish to estimate θ under the squared error loss $L(\theta, d) = (\theta - d)^2$. The sample mean $\bar{X}$ might seem to be one reasonable estimator of θ, but turns out to be inadmissible.

Find a strictly better estimator $\delta(X_1, \ldots, X_n)$, and compute the MSE of each estimator, as a function of n and θ. What happens to the ratio $\text{MSE}(\bar{X})/\text{MSE}(\delta)$ as $n \to \infty$?