Exercises Relating To Sufficient Statistics

Arya Pourzanjani

May 31, 2016

1 A Minimal Sufficient Statistic That is Not Complete

Suppose that (X_i, Y_i), $i = 1, \cdots, n$ are sampled i.i.d from the two-dimensional normal distribution

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix} \right) \tag{1}$$

with $\theta \in \Omega = (-1, 1)$.

1.1 A Two-Dimensional Minimal Sufficient Statistic

By independence the joint density of our n sample is

$$\begin{aligned}
& \prod_{i=1}^{n} p_\theta \left(\begin{pmatrix} x_i \\ y_i \end{pmatrix} \right) \\
= \;& (2\pi\sqrt{1-\theta^2})^{-n} \exp \left\{ -\frac{1}{2} \sum \begin{pmatrix} x_i & y_i \end{pmatrix} \begin{pmatrix} 1/(1-\theta^2) & -\theta/(1-\theta^2) \\ -\theta/(1-\theta^2) & 1/(1-\theta^2) \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right\} \\
= \;& (2\pi\sqrt{1-\theta^2})^{-n} \exp \left\{ -\frac{1}{2(1-\theta^2)} \sum x_i^2 - 2\theta x_i y_i + y_i^2 \right\} \\
= \;& (\pi\sqrt{1-\theta^2})^{-n} \exp \left\{ \underbrace{\frac{\theta}{1-\theta^2}}_{\eta_1(\theta)} \underbrace{\sum x_i y_i}_{T_1(\mathcal{D})} + \underbrace{-\frac{1}{2(1-\theta^2)}}_{\eta_2(\theta)} \underbrace{\sum x_i^2 + y_i^2}_{T_2(\mathcal{D})} \right\}
\end{aligned} \tag{2}$$

thus our sufficient statistic is

$$T(\mathcal{D}) = T\left(\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix}\right) = \begin{pmatrix} \sum x_i y_i \\ \sum x_i^2 + y_i^2 \end{pmatrix} \tag{3}$$

and our parameter of interest η in this canonical form is

$$\eta(\theta) = \begin{pmatrix} \theta/(1-\theta^2) \\ -1/2(1-\theta^2) \end{pmatrix} \tag{4}$$

1.2 Minimality of the Sufficient Statistic

Note that our parameter of interest η is a two-dimensional vector, i.e. an element of $\mathbb{R}^2$. The possible values of η are any $\eta(\theta)$ where $\theta \in (-1, 1)$. Thus we have a whole continuum of vectors in $\mathbb{R}^2$ to consider, and each of these vector corresponds to a value of the original parameter θ. This is all plotted in the figure below

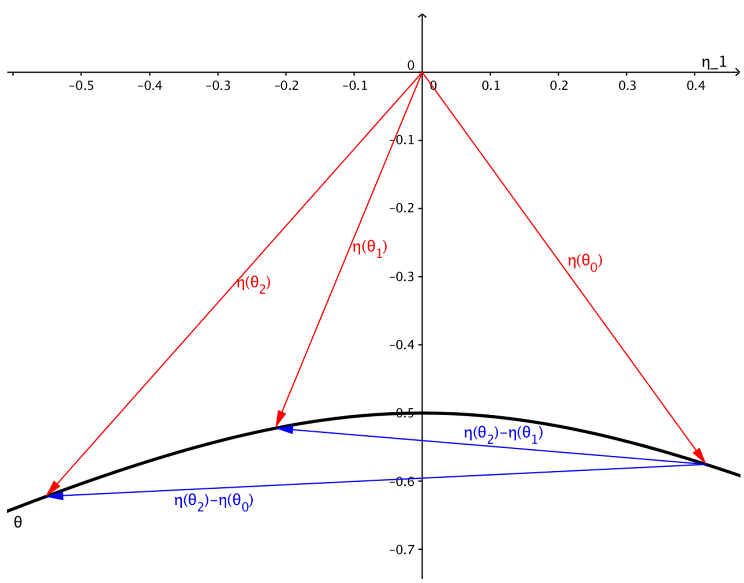

Figure 1: Plot of all possible values of η

The red two-dimensional vectors are different possible η's, they each correspond to a different value of θ. Note that they all lie on the thick black curve parameterized by θ. We can take the difference of any two to create a new vector. The blue vectors are two examples of this. We can pick any two to subtract and thus we have a continuum of possible blue vectors. In Keener this continuum of possible blue vectors is referred to as $\eta(\Omega) \ominus \eta(\Omega)$. If the span of this set is equal to the span of the parameter space in which the vector $\eta(\theta)$ lives, then our statistic will be minimally sufficient.

1.3 Completeness of the Sufficient Statistic

For our sufficient minimal sufficient statistic to be complete we must have that for any possible θ we choose, there is no non-trivial function $f(T)$ such that

$$E_\theta f(T(\mathcal{D})) = \int_{\mathcal{D}} f(T(\mathcal{D}))\, dP_\theta(\mathcal{D}) = c \tag{5}$$

for some constant c. The expected value value of

$$T_2(\mathcal{D}) = \sum x_i^2 + y_i^2 \tag{6}$$

is $2n$. Thus the function

$$f(T(\mathcal{D})) = T_2(\mathcal{D}) - 2n \tag{7}$$

is a non-trivial function of the sufficient statistic T that will have an expected value of zero for any $\mathcal{D}$ and thus our statistic is not complete.