Minimal Sufficient Statistics in Exponential Families
Stat 543 Spring 2005

Showing that the natural sufficient statistic in an exponential family is minimal sufficient is carried out by showing that the statistic has another property sometimes of independent interest, a property called "completeness."

Definition 1 *A statistic $T(X)$ is called (boundedly) complete if for any (bounded) real valued function $h(t)$*

$$E_\theta h(T(X)) = 0\ \forall\theta \Longrightarrow P_\theta\left[h(T(X)) = 0\right] = 1\ \forall\theta$$

It follows from some hard real analysis (essentially establishing a uniqueness property of Laplace transforms) that provided the parameter space (in the natural parameterization $\boldsymbol{\eta}$) under consideration includes an open rectangle in $\Re^k$, the natural sufficient statistic in an exponential family is complete (and therefore boundedly complete). That is, it is a "complete sufficient statistic." The fact that it is also minimal sufficient then follows from a theorem (also sometimes of independent interest) due to Bahadur.

Theorem 2 *(Bahadur's Theorem) Suppose that $T(X)$ taking values in $\Re^k$ is sufficient for θ and boundedly complete. Then $T(X)$ is minimal sufficient.*

Proof. Without loss of generality, we may assume that each coordinate of $T(X) \doteq (T_1(X), T_2(X), ..., T_k(X))$ takes values in $(0,1)$. If not, for example

$$T^*(X) \doteq \left(\frac{1}{1+\exp(T_1(X))}, ..., \frac{1}{1+\exp(T_k(X))}\right)$$

is equivalent to $T(X)$, and takes values in $(0,1)^k$.

Let $S(X)$ be any other sufficient statistic. We want to show that $T(X)$ can be realized as a function of $S(X)$. Define

$$H_i(s) \doteq \mathrm{E}\left[T_i(X)\,|S(X) = s\right]$$

(note that by sufficiency, the expectation here doesn't depend on θ). Further, let

$$L_i(t) \doteq \mathrm{E}\left[H_i(S(X))\,|T(X) = t\right] \quad ,$$

(and again note that by sufficiency, the expectation doesn't depend on θ). We will show that $P_\theta[T_i(X) = H_i(S(X))] = 1\,\forall\theta$, and then have the desired conclusion.

Now the fact that each $T_i(X)$ takes values in $(0,1)$ implies that $0 \le H_i(s) \le 1$ and that $0 \le L_i(t) \le 1$. Note too that

$$\begin{aligned}
\mathrm{E}_\theta T_i(X) &= \mathrm{E}_\theta\left(\mathrm{E}\left[T_i(X)\,|S(X)\right]\right) \\
&= \mathrm{E}_\theta H_i(S(X)) \\
&= \mathrm{E}_\theta\left(\mathrm{E}\left[H_i(S(X))\,|T(X)\right]\right) \\
&= \mathrm{E}_\theta L_i(T(X))
\end{aligned}$$

So

$$\mathrm{E}_\theta\left(T_i(X) - L_i(T(X))\right) = 0 \quad \forall\theta \ .$$

Bounded completeness then implies that

$$P_\theta[T_i(X) = L_i(T(X))] = 1\ \forall\theta\ . \tag{1}$$

So

$$\mathrm{E}\left[L_i(T(X))\,|S(X) = s\right] = \mathrm{E}\left[T_i(X)\,|S(X) = s\right] = H_i(s) \ . \tag{2}$$

Then for any θ

$$\begin{aligned}
\mathrm{Var}_\theta L_i(T(X)) &= \mathrm{E}_\theta \mathrm{Var}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] + \mathrm{Var}_\theta \mathrm{E}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] \\
&= \mathrm{E}_\theta \mathrm{Var}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] + \mathrm{Var}_\theta H_i\left(S\left(X\right)\right) \\
&= \mathrm{E}_\theta \mathrm{Var}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] + \mathrm{E}_\theta \mathrm{Var}\left[H_i\left(S\left(X\right)\right)|T\left(X\right)\right] + \mathrm{Var}_\theta \mathrm{E}\left[H_i\left(S\left(X\right)\right)|T\left(X\right)\right] \\
&= \mathrm{E}_\theta \mathrm{Var}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] + \mathrm{E}_\theta \mathrm{Var}\left[H_i\left(S\left(X\right)\right)|T\left(X\right)\right] + \mathrm{Var}_\theta L_i\left(T\left(X\right)\right) \quad .
\end{aligned}$$

(The second equality above follows from (2).) Now the fact that the random variables $\mathrm{Var}[L_i\left(T\left(X\right)\right)|S\left(X\right)]$ and $\mathrm{Var}[H_i\left(S\left(X\right)\right)|T\left(X\right)]$ are both nonnegative implies that both of the expectations on the right side of the last line above are nonnegative. This in turn implies that they are both 0, which in turn implies that

$$P_\theta[\mathrm{Var}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right] = 0] = 1 \quad ,$$

so that

$$P_\theta\left(L_i(T(X) = \mathrm{E}\left[L_i\left(T\left(X\right)\right)|S\left(X\right)\right]\right) = 1 \ .$$

But again $\mathrm{E}[L_i\left(T\left(X\right)\right)|S\left(X\right)] = H_i(S(X))$, so by (1)

$$P_\theta[T_i(X) = L_i(T(X)) = H_i(S(X))] = 1 \ .$$

■