AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 121–130

Profile Sufficiency

Abram Kagan
University of Maryland, College Park, U.S.A.

Abstract: Let $\mathcal{P} = \{P_{\theta_1,\theta_2},\ (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\}$ be a family of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ parameterized by a pair of abstract valued parameters θ_1, θ_2. A statistic T_1 is called profile sufficient for θ_1 if for any fixed $\theta_2 \in \Theta_2$, T_1 is sufficient for θ_1.

For a dominated family $\mathcal{P}$, a necessary and sufficient condition in the form of a factorization theorem is proved for T_1 to be profile sufficient for θ_1 and for a statistic T_2 to be profile sufficient for θ_2. The classical (Halmos - Savage) factorization theorem is its special case corresponding to $T_1 = T_2$.

If T_i is profile sufficient for θ_i , $i = 1, 2$ and a statistic S is independent of T_1 and (separately) of T_2 (but S is not assumed independent of (T_1, T_2)) for all θ_1, θ_2, then S is ancillary.

Keywords: Factorization Theorem, Ancillarity.

1 Introduction and Basic Definitions

A triple $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ where $(\mathcal{X}, \mathcal{A})$ is a measurable space and $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is a family of probability measures on $\mathcal{A}$, is called a model for an observation X if

$$P_\theta(X \in A) = P_\theta(A)\,, \quad A \in \mathcal{A}\,.$$

A statistic T is a map of $(\mathcal{X}, \mathcal{A})$ into another measurable space, $T : (\mathcal{X}, \mathcal{A}) \to (\mathcal{T}, \mathcal{B})$. A statistic T is called sufficient for $\mathcal{P}$ (or for the parameter θ) if for every (bounded) statistic $\varphi : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathrm{Bor})$ (Bor stands for the standard Borel sigma-algebra) there exists a statistic $\tilde{\varphi}$ with

$$E_\theta(\varphi|T) = \tilde{\varphi}\,.$$

A sufficient statistic T is often identified with the subalgebra $\tilde{\mathcal{A}} = T^{-1}(\mathcal{B})$ of $\mathcal{A}$ called a sufficient subalgebra.

The concept of sufficiency is due to Fisher and is one of the foundations of statistical inference. Many basic concepts and results (e.g., likelihood, Rao-Blackwellization, completeness, exponential families) are directly related to sufficiency and many more are related indirectly. For the role of sufficiency in different problems of statistical inference, see monographs of Lehmann (1986), Kagan et al. (1973) and Witting (1985). The monograph of Huzurbazar (1976) treats sufficiency from the fiducial point of view.

Suppose that all P_θ are absolutely continuous with respect to a σ-finite measure μ (in this case the family $\mathcal{P}$ is called dominated) with densities

$$p(x; \theta) = \frac{dP_\theta}{d\mu}(x)\,.$$

The classical factorization theorem (Halmos and Savage, 1949) gives a necessary and sufficient condition for a statistic T to be sufficient for $\mathcal{P}$; the density $p(x;\theta)$ is to be factorized as

$$p(x;\theta) = R(T(x);\theta)r(x)\,, \quad x \in \mathcal{X}\,, \quad \theta \in \Theta\,. \tag{1}$$

Being a function of $T(x)$ simply means $\tilde{\mathcal{A}}$-measurability.

In this paper a concept of profile sufficiency is introduced; it is related to sufficiency in the same way as profile likelihood is related to likelihood.

Suppose that a family $\mathcal{P}$ is parameterized by a “bivariate” parameter

$$\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 = \Theta\,.$$

A statistic $T_1 : (\mathcal{X}, \mathcal{A}) \to (\mathcal{T}_1, \mathcal{B}_1)$ is called *profile sufficient* for θ_1 if for any fixed $\theta_2 \in \Theta_2$, T_1 is sufficient for the family $\mathcal{P}_1 = \{P_{\theta_1,\theta_2},\ \theta_1 \in \Theta_1\}$.

Note that profile sufficiency of T_1 for θ_1 is weaker that “*sufficiency of* T_1 *for* θ_1 *in presence of nuisance* θ_2”, a useful tool in constructing most powerful tests of statistical hypotheses (see, Rao, 1965 or Lehmann, 1986, Chapter 3, Problem 31)). The latter requires that (i) T_1 be profile sufficient for θ_1 and (ii) the (marginal) distribution of T_1 depend only on θ_1.

Bondesson (1983) showed that if T_1 is profile sufficient for θ_1 and complete in the Lehmann-Scheffé sense, then any statistic $g(T_1)$ with finite second moment is a uniformly minimum variance unbiased estimator (UMVUE).

Similarly to T_1, we say that $T_2 : (\mathcal{X}, \mathcal{A}) \to (\mathcal{T}_2, \mathcal{B}_2)$ is profile sufficient for θ_2 if for any fixed $\theta_1 \in \Theta_1$, T_2 is sufficient for the family $\mathcal{P}_2 = \{P_{\theta_1,\theta_2},\ \theta_2 \in \Theta_2\}$. We shall refer to the subalgebra $\tilde{\mathcal{A}}_i = T_i^{-1}(\mathcal{B}_i)$ of $\mathcal{A}$ as profile sufficient for θ_i, $i = 1, 2$. A statistic $S : (\mathcal{X}, \mathcal{A}) \to (\mathcal{S}, \mathcal{C})$ is called a *subordinate* of T if $S : (\mathcal{T}, \mathcal{B}) \to (\mathcal{S}, \mathcal{C})$. Less formally, S is a subordinate of T if $S = S(T)$.

A statistic $\hat{T}$ is called the *upper subordinate* of T_1 and T_2 if

(i) $\hat{T}$ is a subordinate of T_i, $i = 1, 2$,

(ii) any subordinate S of T_i, $i = 1, 2$ is also a subordinate of $\hat{T}$.

If (i) and (ii) hold, we shall write $\hat{T} = T_1 \wedge T_2$. If $\tilde{\mathcal{A}}_i = T_i^{-1}(\mathcal{B}_i)$, $i = 1, 2$ is a subalgebra of $\mathcal{A}$ generated by T_i, the subalgebra generated by $\hat{T}$ is $\hat{\mathcal{A}} = \tilde{\mathcal{A}}_1 \cap \tilde{\mathcal{A}}_2$.

The main result of the paper is that if $\mathcal{P} = \{P_{\theta_1,\theta_2},\ (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2\}$ is a dominated family of probability measures with positive densities $p(x; \theta_1, \theta_2)$, a necessary and sufficient condition for profile sufficiency of T_i for θ_i, $i = 1, 2$ is the following factorization:

$$p(x; \theta_1, \theta_2) = Q(T_1 \wedge T_2; \theta_1, \theta_2)R_1(T_1; \theta_1)R_2(T_2; \theta_2)r(x)\,. \tag{2}$$

The first factor on the right hand side of (2) may depend on both θ_1 and θ_2 but, as a function of x, is measurable with respect to the subalgebra $\hat{\mathcal{A}}$ generated by the upper subordinate of the profile sufficient statistics. The second and third factors depend each on one parameter component and, as functions of x, are $\tilde{\mathcal{A}}_1$- and $\tilde{\mathcal{A}}_2$-measurable, respectively. The last factor does not contain the parameter and, as function of x, may be an arbitrary $\mathcal{A}$-measurable function. The condition

$$p(x; \theta_1, \theta_2) > 0\,, \quad x \in \mathcal{X}\,,\ (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$$

seems technical and likely can be omitted.

Note that if T_i is profile sufficient for θ_i, $i = 1, 2$, the pair $(T_1, T_2) = T$ is sufficient for (θ_1, θ_2) but the former property is much stronger. That is why the classical factorization

$$p(x; \theta_1, \theta_2) = R(T; \theta_1, \theta_2) r(x) \tag{3}$$

holding for some R and r when (T_1, T_2) is sufficient for (θ_1, θ_2) is a special case of (2). In particular, if $T_1 = T_2 = T$, say, then $(T_1, T_2) = T$ and (2) becomes (3). Another special case of interest is when $T_1 \wedge T_2$ is constant, i.e., the subalgebra $\hat{\mathcal{A}}$ is trivial in which case (2) takes the form of

$$p(x; \theta_1, \theta_2) = C(\theta_1, \theta_2) R_1(T_1; \theta_1) R_2(T_2; \theta_2) r(x)\,. \tag{4}$$

This occurs, for example, when T_1, T_2 are profile sufficient and independent for all (θ_1, θ_2). The subalgebra $\hat{\mathcal{A}}$ is trivial. Indeed, if $A \in \hat{\mathcal{A}}$, then $A = A \cap A$ where the first A is a set in $\tilde{\mathcal{A}}_1$ while the second is a set in $\tilde{\mathcal{A}}_2$ so that

$$P_{\theta_1,\theta_2}(A) = P_{\theta_1,\theta_2}(A \cap A) = \{P_{\theta_1,\theta_2}(A)\}^2\,,$$

and hence $P_{\theta_1,\theta_2}(A) = 0$ or 1.

To illustrate the concept of profile sufficiency, let

$$\{p(x^{(1)}, x^{(2)}, x^{(3)}; \theta_1, \theta_2), (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2\}$$

be a family of probability densities on a measurable (product) space $(\mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(3)}, \mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)} \otimes \mathcal{A}^{(3)})$.

Example 1. $T_1 = (x^{(1)}, x^{(2)})$ is profile sufficient for θ_1, $T_2 = (x^{(2)}, x^{(3)})$ is profile sufficient for θ_2. Then $\hat{T} = T_1 \wedge T_2 = x^{(2)}$ and (2) becomes

$$\begin{gathered} p(x^{(1)}, x^{(2)}, x^{(3)}; \theta_1, \theta_2) = \\ Q(x^{(2)}; \theta_1, \theta_2) R_1(x^{(1)}, x^{(2)}; \theta_1) R_2(x^{(2)}, x^{(3)}; \theta_2) r(x^{(1)}, x^{(2)}, x^{(3)})\,. \end{gathered} \tag{5}$$

Example 2. $T_1 = x^{(1)}$ is profile sufficient for θ_1, $T_2 = x^{(2)}$ is profile sufficient for θ_2. In this case, $\hat{T} = T_1 \wedge T_2$ is constant and (2) becomes

$$p(x^{(1)}, x^{(2)}, x^{(3)}; \theta_1, \theta_2) = C(\theta_1, \theta_2) R_1(x^{(1)}; \theta_1) R_2(x^{(2)}; \theta_2) r(x^{(1)}, x^{(2)}, x^{(3)})\,. \tag{6}$$

Turn now to a family $\mathcal{P} = \{p(\mathbf{x}; \theta), \theta \in \Theta\}$ of product densities generated by a sample $\mathbf{x} = (x_1, \ldots, x_n)$ from a "univariate" population with density $f(x; \theta)$, i.e.,

$$p(\mathbf{x}; \theta) = \prod_{i=1}^{n} f(x_i; \theta)\,. \tag{7}$$

If $f(x; \theta)$ belongs to an exponential family,

$$f(x; \theta) = \exp\left\{\sum_{i=1}^{M} c_i(\theta)\varphi_i(x) + \varphi_0(x) + c_0(\theta)\right\}\,, \tag{8}$$

the statistic

$$T = (T_1, \ldots, T_M)\,, \quad T_i(x_1, \ldots, x_n) = \sum_{j=1}^{n} \varphi_i(x_j)\,, \quad i = 1, \ldots, M \tag{9}$$

is sufficient for $\mathcal{P}$. As is well known (see, e.g., Barndorff-Nielsen, 1979; Brown, 1985), in the so called regular case the exponential families are the only source of nontrivial sufficient statistics in product families. If $\theta = (\theta_1, \theta_2)$ and some coefficients $c_i(\theta)$ in (8) depend only on θ_1, some other only on θ_2 and the remaining on both θ_1 and θ_2, the sufficient statistic (9) can be split into profile sufficient statistics T_1, T_2. Using sub- and superscripts for convenience, assume that $M = J + K + L$ and

$$\begin{aligned} c_j(\theta) &= c_j^{(1)}(\theta_1)\,, \quad j = 1, \ldots, J \\ c_{J+k}(\theta) &= c_k^{(2)}(\theta_2)\,, \quad k = 1, \ldots, K \\ c_{J+K+l}(\theta) &= c_l^{(12)}(\theta_1, \theta_2)\,, \quad l = 1, \ldots, L\,. \end{aligned}$$

With these notations,

$$p(\mathbf{x}; \theta_1, \theta_2) = \exp\left\{ \sum_{j=1}^{J} c_j^{(1)}(\theta_1) T_j^{(1)} + \sum_{k=1}^{K} c_k^{(2)}(\theta_2) T_k^{(2)} + \sum_{l=1}^{L} c_l^{(12)}(\theta_1, \theta_2) T_l^{(12)} + T_0 + n c_0(\theta_1, \theta_2) \right\} \tag{10}$$

where

$$\begin{aligned} T_0 &= \sum_{i=1}^{n} \varphi_0(x_i)\,; \quad T_j^{(1)} = T_j\,,\ j = 1, \ldots, J; \\ T_k^{(2)} &= T_{J+k}\,,\ k = 1, \ldots, K\,; \quad T_l^{(12)} = T_{J+K+l}\,, \quad l = 1, \ldots, L \end{aligned}$$

and $T_1, \ldots, T_{J+K+L}$ are given in (9). On setting

$$T^{(1)} = (T_1^{(1)}, \ldots, T_J^{(1)}\,; T_1^{(12)}, \ldots, T_L^{(12)})\,, \quad T^{(2)} = (T_1^{(2)}, \ldots, T_K^{(2)}\,; T_1^{(12)}, \ldots, T_L^{(12)})$$

one sees from (10) that $T^{(i)}$ is profile sufficient for θ_i, $i = 1, 2$ and $\hat{T} = T^{(1)} \wedge T^{(2)} = (T_1^{(12)}, \ldots, T_L^{(12)})$.

Example 3. $\mathbf{x} = (x_1, \ldots, x_n)$; $x_i = (x_i^{(1)}, x_i^{(2)})$ is a sample from a population with density

$$f(x^{(1)}, x^{(2)}; \theta_1, \theta^2) = \exp\left\{ \left(\theta_1 x^{(1)} + \theta_2 x^{(2)}\right)^2 + \varphi_0(x^{(1)}, x^{(2)}) + c_0(\theta_1, \theta_2) \right\}$$

where $(\theta_1, \theta_2) \in \mathbb{R}^2$, $\varphi_0(x^{(1)}, x^{(2)})$ makes the integral $\int\int f(x^{(1)}, x^{(2)}; \theta_1, \theta_2) dx^{(1)} dx^{(2)}$ convergent and $C_0(\theta_1, \theta_2)$ makes f a probability density. The profile sufficient statistic for θ_1 is $T^{(1)}(\mathbf{x}) = \left(\sum (x_i^{(1)})^2, \sum x_i^{(1)} x_i^{(2)}\right)$, the profile sufficient statistic for θ_2 is $T^{(2)}(\mathbf{x}) = \left(\sum (x_i^{(2)})^2, \sum x_i^{(1)} x_i^{(2)}\right)$ and $\hat{T}(\mathbf{x}) = \sum x_i^{(1)} x_i^{(2)}$.

There is a relation between profile sufficiency and ancillarity. We call a family $\mathcal{P} = \{P_{\theta_1,\theta_2},\ (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\}$ *linked* if no two elements of $\mathcal{P}$ are mutually singular. If (i) $\mathcal{P}$ is linked, (ii) T_i is profile sufficient for θ_i, $i = 1, 2$ and (iii) a statistic S is independent of T_1 for all (θ_1, θ_2) and of T_2, then S is ancillary, i.e., its distribution is the same for all (θ_1, θ_2). Note that independence of S of the pair (T_1, T_2) is not required; neither is the family $\mathcal{P}$ assumed dominated.

Similarly, if $\mathcal{P}$ is linked and T_1, T_2 are profile sufficient and independent, then the distribution of T_i depends only on θ_i, $i = 1, 2$.

In Section 2 the factorization theorem is proved and in Section 3 the relation between profile sufficiency and ancillarity is discussed.

2 Factorization Theorem: A Necessary and Sufficient Condition for Profile Sufficiency

We assume in this section that the elements P_{θ_1,θ_2} of a family

$$\mathcal{P} = \{P_{\theta_1,\theta_2},\ (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\}$$

of probability measures on a measurable space $(\mathcal{X}, \mathcal{A})$ are absolutely continuous with respect to a σ-finite measure μ and the densities

$$p(x;\theta_1,\theta_2) = \frac{dP_{\theta_1,\theta_2}}{d\mu}(x)$$

are positive,

$$p(x;\theta_1,\theta_2) > 0\,, \quad x \in \mathcal{X}\,,\ (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2\,.$$

Theorem 2.1 *For a statistic $T_1 : (\mathcal{X}, \mathcal{A}) \to (\mathcal{T}_1, \mathcal{B}_1)$ to be profile sufficient for θ_1 and a statistic $T_2 : (\mathcal{X}, \mathcal{A}) \to (\mathcal{T}, \mathcal{B}_2)$ to be profile sufficient for θ_2, it is necessary and sufficient that the following factorization hold:*

$$p(x;\theta_1,\theta_2) = Q(\hat{T}(x);\theta_1,\theta_2)R_1(T_1(x);\theta_1)R_2(T_2(x);\theta_2)r(x) \tag{11}$$

where $\hat{T} = T_1 \wedge T_2$ is the upper subordinate of T_1 and T_2.

In terms of subalgebras, if $\tilde{\mathcal{A}}_i = T_i^{-1}(\mathcal{B}_i)$, $i = 1, 2$ and $\hat{\mathcal{A}} = \tilde{\mathcal{A}}_1 \cap \tilde{\mathcal{A}}_2$, the factorization means that the first factor on the right hand side of (11), as a function of x, is $\hat{\mathcal{A}}$-measurable, the second and third factors are $\tilde{\mathcal{A}}_1$- and $\tilde{\mathcal{A}}_2$-measurable, respectively, while the last factor is simply $\mathcal{A}$-measurable but does not depend on either parameter.

Proof. Sufficiency. Assume that (11) holds. Then for any fixed θ_2^*, (11) can be written as

$$p(x;\theta_1,\theta_2^*) = R(T_1(x);\theta_1)r_1(x)$$

with

$$R(T_1(x);\theta_1) = Q(\hat{T}(x);\theta_1,\theta_2^*)R_1(T_1(x);\theta_1)\,, \quad r_1(x) = R_2(T_2(x);\theta_2^*)r(x)\,.$$

Note that since $\hat{T}$ is a subordinate of T_1,

$$Q(\hat{T}(x);\theta_1,\theta_2^*) = \tilde{Q}(T_1(x);\theta_1)\,.$$

Hence, by virtue of the classical (Halmos-Savage) factorization theorem, T_1 is profile sufficient for θ_1. □

Necessity. From the classical factorization theorem, profile sufficiency of T_1 for θ_1 and of T_2 for θ_2 imply the two factorizations of $p(x;\theta_1,\theta_2)$,

$$p(x;\theta_1,\theta_2) = U(T_1;\theta_1,\theta_2)u(x;\theta_2)\,, \tag{12}$$

$$p(x;\theta_1,\theta_2) = V(T_2;\theta_1,\theta_2)v(x;\theta_1)\,. \tag{13}$$

(Here and in what follows we are skipping the argument when it does not lead to a confusion.) Dividing (12) by (13) leads to

$$\frac{U(T_1;\theta_1,\theta_2)}{V(T_2;\theta_1,\theta_2)} = \frac{v(x;\theta_1)}{u(x;\theta_2)} \tag{14}$$

which for $\theta_2 = \theta_2^*$ becomes

$$\frac{U(T_1;\theta_1,\theta_2^*)}{V(T_2;\theta_1,\theta_2^*)} = \frac{v(x;\theta_1)}{u(x;\theta_2^*)}\,. \tag{15}$$

Combining (14) with (15) gives

$$\frac{U(T_1;\theta_1,\theta_2)}{V(T_2;\theta_1,\theta_2)}\frac{V(T_2;\theta_1,\theta_2^*)}{U(T_1;\theta_1,\theta_2^*)} = \frac{u(x;\theta_2^*)}{u(x;\theta_2)}\,. \tag{16}$$

The right hand side of (16) does not depend on θ_1 and, hence, neither does the left hand side:

$$\frac{U(T_1;\theta_1,\theta_2)}{U(T_1;\theta_1,\theta_2^*)}\frac{V(T_2;\theta_1,\theta_2^*)}{V(T_2;\theta_1,\theta_2)} = \frac{U(T_1;\theta_1^*,\theta_2)}{U(T_1;\theta_1^*,\theta_2^*)}\frac{V(T_2;\theta_1^*,\theta_2^*)}{V(T_2;\theta_1^*,\theta_2)} \tag{17}$$

whence

$$\frac{U(T_1;\theta_1,\theta_2)}{U(T_1;\theta_1,\theta_2^*)}\frac{U(T_1;\theta_1^*,\theta_2^*)}{U(T_1;\theta_1^*,\theta_2)} = \frac{V(T_2;\theta_1,\theta_2)}{V(T_2;\theta_1,\theta_2^*)}\frac{V(T_2;\theta_1^*,\theta_2^*)}{V(T_2;\theta_1^*,\theta_2)}\,. \tag{18}$$

But the left hand side of (18) is a function of $T_1(x)$ (i.e., as function of x, is $\tilde{\mathcal{A}}_1$-measurable) while the right hand side is a function of $T_2(x)$ (i.e., $\tilde{\mathcal{A}}_2$-measurable). Thus, they both are functions of $\hat{T}(x) = T_1(x) \wedge T_2(x)$, i.e., $\hat{\mathcal{A}}$-measurable. On setting

$$Q_1(\hat{T};\theta_1,\theta_2) = \frac{V(T_2;\theta_1,\theta_2)}{V(T_1;\theta_1,\theta_2^*)}\frac{V(T_2;\theta_1^*,\theta_2^*)}{V(T_2;\theta_1^*,\theta_2)} \tag{19}$$

one gets from (18)

$$U(T_1;\theta_1,\theta_2) = Q_1(\hat{T};\theta_1,\theta_2)U(T_1;\theta_1,\theta_2^*)U(T_1;\theta_1^*,\theta_2)/U(T_1;\theta_1^*,\theta_2^*) \tag{20}$$

that combined with (12) results in

$$p(x;\theta_1,\theta_2) = Q_1(\hat{T};\theta_1,\theta_2)\tilde{U}(T_1;\theta_1)\tilde{u}(x;\theta_2) \tag{21}$$

with

$$\tilde{U}(T_1;\theta_1) = U(T_1;\theta_1,\theta_2^*)\,,$$
$$\tilde{u}(x;\theta_2) = u(x;\theta_2)U(T_1;\theta_1^*,\theta_2)/U(T_1;\theta_1^*,\theta_2^*)\,.$$

By exchanging θ_1 and θ_2 and proceeding as above, one comes to

$$p(x;\theta_1,\theta_2) = Q_2(\hat{T};\theta_1,\theta_2)\tilde{V}(T_2;\theta_2)\tilde{v}(x;\theta_1) \tag{22}$$

with $\tilde{V}$ and $\tilde{v}$ obtained from $\tilde{U}$ and $\tilde{u}$ by replacing U with V, u with v, T_1 with T_2 and θ_1 with θ_2 (actually, the explicit form of the functions on the right hand sides of (21), (22) does not matter; only their arguments count). Now dividing (21) by (22) gives us

$$\frac{Q_1(\hat{T};\theta_1,\theta_2)\tilde{u}(x;\theta_2)}{Q_2(\hat{T};\theta_1,\theta_2)\tilde{V}(T_2;\theta_2)} = \frac{\tilde{v}(x;\theta_1)}{\tilde{U}(T_1;\theta_1)}\,. \tag{23}$$

Since the right hand side of (23) does not depend on θ_2, neither does the left hand side and thus,

$$\frac{Q_1(\hat{T};\theta_1,\theta_2)\tilde{u}(x;\theta_2)}{Q_2(\hat{T};\theta_1,\theta_2)\tilde{V}(T_2;\theta_2)} = \frac{Q_1(\hat{T};\theta_1,\theta_2^*)\tilde{u}(x;\theta_2^*)}{Q_2(\hat{T};\theta_1,\theta_2^*)\tilde{V}(T_2;\theta_2^*)}$$

whence

$$\tilde{u}(x;\theta_2) = \tilde{Q}(\hat{T}(x);\theta_1,\theta_2)\tilde{V}(T_2(x);\theta_2)r(x) \tag{24}$$

where

$$\tilde{Q}(\hat{T};\theta_1,\theta_2) = \frac{Q_2(\hat{T};\theta_1,\theta_2)Q_1(\hat{T};\theta_1,\theta_2^*)}{Q_1(\hat{T};\theta_1,\theta_2)Q_2(\hat{T};\theta_1,\theta_2^*)}\,,$$

$$r(x) = \frac{\tilde{u}(x;\theta_2^*)}{\tilde{V}(T_2(x);\theta_2^*)}\,.$$

Substituting (24) into (21) gives the sought factorization (11) with

$$\begin{aligned} Q(\hat{T}(x);\theta_1,\theta_2) &= Q_1(\hat{T}(x);\theta_1,\theta_2)\tilde{Q}(\hat{T}(x);\theta_1,\theta_2)\,, \\ R_1(T_1(x);\theta_1) &= \tilde{U}(T_1(x);\theta_1)\,, \\ R_2(T_2(x);\theta_2) &= \tilde{V}(T_2(x);\theta_2)\,, \end{aligned} \tag{25}$$

and all the functions on the right hand sides of relations (25) are defined above. □

As in the classical factorization theorem, factorization (11) is not unique. For example, one may multiply $R_2(T_2(x);\theta_2)$ by an arbitrary function $h(T_2(x))$ and divide $r(x)$ by the same function without affecting (11).

3 Profile Sufficiency and Ancillarity

Let T be sufficient for a family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. If $\mathcal{P}$ is linked, i.e., no two elements $P_{\theta'}$, $P_{\theta''}$ are mutually singular and a statistic S is independent of T for all $\theta \in \Theta$, then S is ancillary. It is a well known fact.

Note in passing that the condition that $\mathcal{P}$ is linked can not simply be omitted as the following example demonstrates. Let $\mathcal{P} = \{P_1, P_2\}$ consists of two mutually singular distributions so that for some A, $P_1(A) = 1$, $P_2(A) = 0$. Plainly, the indicator $\chi_A = T$ is sufficient for $\mathcal{P}$ (knowing T means knowing the distribution). Let now $B \subset A$ with $P_1(B) > 0$. Then

$$P_1\{B \cap (T = 1)\} = P_1(B)P_1(T = 1)\,, \quad P_1\{B \cap (T = 0)\} = P_1(B)P_1(T = 0)\,,$$

$$P_2\{B \cap (T = 1)\} = P_2(B)P_2(T = 1)\,, \quad P_2\{B \cap (T = 0)\} = P_2(B)P_2(T = 0),$$

while $P_1(B) > 0$, $P_2(B) = 0$.

Theorem 3.1 *Let $\mathcal{P} = \{P_{\theta_1,\theta_2}, (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2\}$ be a linked family and T_i is profile sufficient for θ_i, $i = 1, 2$. If a statistic S is independent of T_1 for all (θ_1, θ_2) and (separately) of T_2, then S is ancillary.*

Proof. From profile sufficiency of T_1 for θ_1 and T_2 for θ_2 one has for any bounded $h = h(S)$

$$E_{\theta_1,\theta_2}(h|T_1) = C'(T_1, \theta_2) \quad \text{a.e. } P_{\theta_1,\theta_2}\,, \tag{26}$$

$$E_{\theta_1,\theta_2}(h|T_2) = C''(T_2, \theta_2) \quad \text{a.e. } P_{\theta_1,\theta_2}\,. \tag{27}$$

From independence of S and T_1 and (separately) of T_2, (26), (27) imply

$$C'(T_1, \theta_2) = c'(\theta_1, \theta_2) \quad \text{a.e. } P_{\theta_1,\theta_2}\,, \tag{28}$$

$$C''(T_2, \theta_2) = c''(\theta_1, \theta_2) \quad \text{a.e. } P_{\theta_1,\theta_2} \tag{29}$$

for some $c'(\theta_1, \theta_2)$, $c''(\theta_1, \theta_2)$. Let (28) hold on a set $A(\theta_1, \theta_2)$ with $P_{\theta_1,\theta_2}\{A(\theta_1, \theta_2)\} = 1$. Since $\mathcal{P}$ is linked, for any $\theta_1', \theta_1'', \theta_2$,

$$P_{\theta_1',\theta_2}\{A(\theta_1', \theta_2)\} = 1 \to P_{\theta_1'',\theta_2}\{A(\theta_1', \theta_2)\} > 0$$

so that due to $P_{\theta_1'',\theta_2}\{A(\theta_1'', \theta_2)\} = 1$,

$$P_{\theta_1'',\theta_2}\{A(\theta_1', \theta_2) \cap A(\theta_1'', \theta_2)\} > 0\,.$$

On the intersection $A(\theta_1', \theta_2) \cap A(\theta_1'', \theta_2)$

$$C'(T_1, \theta_2) = c'(\theta_1', \theta_2) = c'(\theta_1'', \theta_2)$$

proving that

$$c'(\theta_1, \theta_2) = c'(\theta_2)$$

does not depend on θ_1. Similarly,

$$c''(\theta_1, \theta_2) = c''(\theta_1)$$

does not depend on θ_1. From (26), (27) one has

$$E_{\theta_1,\theta_2}\{h(S)\} = c'(\theta_2) = c''(\theta_1)$$

implying

$$E_{\theta_1,\theta_2}\{h(S)\} = \text{const.} \tag{30}$$

Since (30) holds for any bounded $h(S)$, the distribution of S does not depend on (θ_1, θ_2) so that S is an ancillary statistic. □

Theorem 3.2 *If $\mathcal{P} = \{P_{\theta_1,\theta_2}, (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2\}$ is linked and T_1, T_2 are profile sufficient and independent for all (θ_1, θ_2), then the distribution of T_i depends only on θ_i, $i = 1, 2$.*

Proof. Take an arbitrary bounded function $h(T_1)$. Since T_2 is profile sufficient for θ_2,

$$E_{\theta_1,\theta_2}\{h(T_1)|T_2\} = \tilde{h}(T_2, \theta_1)\,.$$

Independence of T_1 and T_2 implies that there is a constant (with respect to T_2) $\tilde{c}(\theta_1, \theta_2)$ such that

$$P_{\theta_1,\theta_2}\{\tilde{h}(T_2, \theta_1) = \tilde{c}(\theta_1, \theta_2)\} = 1\,.$$

Following the arguments used in Theorem 2 leads to

$$\tilde{c}(\theta_1, \theta_2) = \tilde{c}(\theta_1)$$

whence

$$E_{\theta_1,\theta_2}\{h(T_1)\} = \tilde{c}(\theta_1)\,.$$

Since h is arbitrary, it means that the distribution of T_1 depends only on θ_1. □.

Acknowledgements

The work on profile sufficiency began when the author visited Dept. of Statistics, University of Leeds with financial support from Engineering and Physical Sciences Research Council of the U.K. Many thanks to John Kent and Lutz Mattner for their hospitality and useful discussions. Thanks to Marek Omelka for his careful reading of the manuscript and pointing a number of typos.

References

Barndorff-Nielsen, O. E. (1979). *Information and Exponential Families in Statistical Theory*. New York: J. Wiley.

Bondesson, L. (1983). On uniformly minimum variance unbiased estimation when no complete sufficient statistics exist. *Metrika*, *30*, 49-54.

Brown, L. D. (1985). *Foundations of Exponential Families*. Hayward, California: IMS Monograph Series.

Halmos, P. R., and Savage, J. L. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Annals of Mathematical Statistics*, *20*, 225-241.

Huzurbazar, V. S. (1976). *Sufficient Statistics*. New York: Marcel Dekker.

Kagan, A. M., Linnik, Y. V., and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*. New York: J. Wiley.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2 ed.). New York: J. Wiley.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, *52*, 447-452.

Witting, H. (1985). *Mathematische Statistik* (Vol. I). Berlin: Teubner.

Author's address:

Abram Kagan
Department of Mathematics
University of Maryland
College Park, MD 20742
USA

E-mail: amk@math.umd.edu