I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0126

A Note on Partial Sufficiency with Connection to the Identification Problem

M. Mouchart and A. Oulhaj

http://www.stat.ucl.ac.be

A Note on Partial Sufficiency with Connection to the Identification Problem

M.Mouchart[1] and A.Oulhaj[1]

July 6, 2001

Abstract

Let ($\mathbf{R}_X$, $\mathcal{X}$, $\mathcal{P}^\Theta = \{P^\theta : \theta \in \Theta\}$) be a parametrized statistical model and $g : \Theta \to G$ be a non injective function characterising a parameter of interest. The basic idea of partial sufficiency is to find a (minimal) statistic containing all the relevant information for the inference on the parameter of interest. Following Fraser (1956), Barndorff-Nielsen (1978) has defined a concept of S-sufficiency. This definition is essentially based on two fundamental concepts: The first one concerns the notion of *sufficient parameter*, and the second one is related to the notion of a *variation free reparametrization.* Our contribution to this area is to give some properties of the S-sufficiency concept. In particular, we establish the connection between S-sufficiency and the identification concept. We show that S-sufficiency preserves the identification property, i.e. if we reduce the complete model by marginalizing on an S-sufficient statistic, we do not lose the identification of the parameter of interest. Furthermore, we review the properties of the standard concept of sufficiency and we check wich ones also hold for S-sufficiency.

[1]Institut de statistique, Université catholique de Louvain.
Acknowledgements. Without implications, the authors gratefully acknowledge useful comments on earlier versions by A.P. Dawid, J.M. Rolin, M. Rémon and J.P.Florens. Support of the contract *Projet d'Actions de Recherche Concertées* ARC 98/03-217, of the Belgium Governement, is gratefully acknowledged.

1 Introduction

We start from a statistical model $\mathbf{M}_{\mathcal{X}}^{\Theta}$ given in the following form:

$$\mathbf{M}_{\mathcal{X}}^{\Theta} \;=\; [(\mathbf{R}_X \;,\; \mathcal{X}) \;,\; \mathcal{P}^{\Theta} = \{P^{\theta} \;:\; \theta \,\in\, \Theta\}] \tag{1}$$

where the measurable space $(\mathbf{R}_X \;,\; \mathcal{X})$ stands for the sample space for the observation of a random vector X, P^{θ} is a probability measure on the sample space for each $\theta \in \Theta$ and Θ represents the parameter space, which may be euclidean or functional or a cartesian product of both. Note that X may be viewed as a measurable function defined on an underlying fundamental probability space, valued on a range space $(R_X, \mathcal{X})$ which may also be euclidean, functional or a cartesian product of both. For the sake of convenience, we use the notation $P(A \mid \theta)$ for the probability of event $A \in \mathcal{X}$ according to the probability measure P^{θ}. We also consider $T = f(X)$ as a statistic, $\mathcal{T} = \sigma(T)$ as the σ-field generated by T and $P_{\mathcal{T}}^{\theta}$ as the trace of P^{θ} on $\mathcal{T}$, this one may also be called the marginal distribution of $\mathcal{T}$. In this paper, script letters $(\mathcal{X}, \mathcal{T}$ etc) will be used to denote σ-fields. Let also $g(\theta)$ be a function defined on the parameter space Θ and representing the parameter of interest. Finally, a submodel, or a reduction, of $\mathbf{M}_{\mathcal{X}}^{\Theta}$ is eventually denoted as $\mathbf{M}_{\mathcal{X}}^{\Theta_0}$ where $\Theta_0 \subset \Theta$.

Let $\mathcal{P}(\mathbf{R}_X, \mathcal{X})$ stand for the set of all probability measures on the sample space. A statistical model, in the form of (1), may also be viewed as a mapping from a parameter space Θ into $\mathcal{P}(\mathbf{R}_X, \mathcal{X})$. Two crucial features should be distinguished. Firstly, the *range* of the mapping underlying (1), i.e the set $\mathcal{P}^{\Theta}$ of possible sampling distributions; this set summarizes all the hypotheses of the model. Secondly, the *properties* of that mapping, e.g. whether it is injective, surjective, continuous etc; these are characteristic of a particular parametrization. Two models sharing a same range space $\mathcal{P}^{\Theta}$ are statistically equivalent, but may correspond to parametrizations with drastically different properties; *for example*, $(X \mid \mu) \sim N(\mu, 1)$ with $\mu \in I\!R$ and $(X \mid \alpha_1, \alpha_2) \sim N(\alpha_1 + \alpha_2, 1)$ with $(\alpha_1, \alpha_2) \in I\!R^2$ are two different parametrizations of statistically equivalent models.

A slightly different definition of a statistical model $\mathbf{M}_{\mathcal{X}}^{\Theta}$, as defined in (1), is that a statistical model is a triplet $(\mathbf{R}_X, \mathcal{X}, \mathcal{P})$ where $\mathcal{P} \subset \mathcal{P}(\mathbf{R}_X, \mathcal{X})$. In such a case, no paramatrization, $\theta \to P^{\theta}$, is made explicit. Thus, a statistical model in the form of (1) is also refered to as a "*parametrized*" statistical model.

In this paper, we are interested in studying the concept of partial sufficiency, heuristically the sufficiency of a statistic relatively to a parameter of interest. Two different contexts should be borne in mind:

Context 1: For a given parameter of interest, say $g(\theta)$, one look for a statistic sufficient (w.r.t the full observation) to make inference about $g(\theta)$ without loss of relevant information. In this case, the sufficiency argument is used to *reduce* the sample space of the statistical model, as far as possible without loss of relevant information. Several motivations should be mentionned, among which is the objective of simplifying and reducing the burden of model specification and, eventually, to improve robustness against specification error, as well as diminishing the volume of the computations involved in the treatement of statistical data.

Context 2: In the case of partial observability, X represents a latent vector and $T = f(X)$ is an observable one. $\mathbf{M}_{\mathcal{X}}^{\Theta}$ is the structural model and $\mathbf{M}_{\mathcal{T}}^{\Theta}$, the image of $\mathbf{M}_{\mathcal{X}}^{\Theta}$ under the transformation f, is the statistical model. Here T is given by a criterion of observability rather than of sufficiency and the problem is now to analyze the identification of the statistical model $\mathbf{M}_{\mathcal{T}}^{\theta}$; in particular, whether a parameter of interest, $g(\theta)$, is identified or not in that statistical model.

This paper will show that the two contexts, although logically quite different, are actually closely related. When the parameter function $g(\theta)$ is injective, the usual concept of sufficiency applies and many results have been established in the statistical literature. When $g(\theta)$ is non injective, a concept of S-sufficiency has been introduced by Barndorff-Nielsen (1978) following Fraser (1956). This paper aims at studying properties and conditions of applicability of S-sufficiency, with a view to compare the properties of sufficiency and of S-sufficiency respectively. Moreover, this paper also establishes connections between the concept of S-sufficiency and the identification properties of the marginal statistical model $\mathbf{M}_{\mathcal{T}}^{\theta}$.

This paper is organised as follows: In section 2, we introduce three basic and relevant concepts: The sufficiency of a statistic for a complete parameter, the sufficiency of a parameter for a statistic and the identification of a parameter. We also establish some connections between these concepts. In section 3, we analyze the S-sufficiency concept. We first introduce a notion of a g-partition. Such a notion offers an alternative view on the concept of a variation free reparametrization, introduced in Barndorff-Nielsen (1978) and used in the definition of S-sufficiency. Next, we discuss the concept of S-sufficiency and analyze its properties. In a final section, we offer some concluding remarks. In the main paper, we reduce the technicalities to a minimum, an appendix gathers the proofs and the analytical aspects.

This paper is concerned with two issues. One is that we want to establish, and prove, new results. Another one is that we want to clarify relationships between concepts and definitions spread in a large number of contributions. These objectives eventually lead to a seemingly "abstract" approach trying to isolate the conditions that are actually required for making a result valid, without gelting involved with unnecessary conditions on irrelevant technicalities. Thus we look for results valid not only for euclidean spaces but also for functional spaces, as many be met when analyzing trajectories in continuous time or in semi-parametric models. The basic theory is (implicitly) written in σ-algebraic terms, identifying freely a statistic with the σ-field generated by a statistic, as long as this should create no ambiguity. Denoting by $\mathcal{X}$ the (σ-field generated by) observations, any sub-σ-field of $\mathcal{X}$ will accordingly be considered as a statistic, as well as any function of the observation ($T = f(X)$) (For the connection between σ-fields and the statistics generating them, see e.g. Bahadur (1955)). Most, but not all, results are valid in the undominated case, an essential requirements for many non-parametric models; as a matter of fact, densities will only be used for illustrative purposes.

This paper is written in a sampling theory framework. Thus, from a technical point of view the parameters and the observations are treated differently: we reason pointwise on the parameter space but σ-algebraically on the sample space.

2 Basic theory: Identification and sufficiency

In this section we essentially consider the problem of reducing, by a sufficiency argument, a given sampling model $\mathbf{M}_{\mathcal{X}}^{\Theta}$. For that reason three basic concepts are reminded. The first one is the *sufficiency* of a statistic for a *complete* parameter, the second one is the *identification* of a parameter and the third concept concerns the *sufficiency* of a parameter for a statistic. This last concept is dual to the identification concept and is particularily relevant in the case of partial observability.

Let us start by reminding the usual definition of the concept of a *sufficient* statistic for a *complete* parameter.

Definition 1 *(Barra 1981) A statistic $\mathcal{T}$ of $\mathcal{X}$ is said to be sufficient for $\mathbf{M}_{\mathcal{X}}^{\Theta}$ if for every $A \in \mathcal{X}$, a version of $\dot{P}(A \mid \mathcal{T}, \theta)$ exists wich is the same for all $\theta \in \Theta$ or equivalently:*

$$\forall A \in \mathcal{X} : \bigcap_{\theta \in \Theta} \dot{P}(A \mid \mathcal{T}, \theta) \neq \emptyset \tag{2}$$

where for each $\theta, \dot{P}(A \mid \mathcal{T}, \theta) = \dot{I\!\!E}(\mathbf{1}_A \mid \mathcal{T}, \theta)$ is the equivalence class of all versions of the conditional probability of A given $\mathcal{T}$. If the context is clear enough, one simply say that $\mathcal{T}$ is sufficient.

Let us recall some basic properties of sufficiency. Conditions (2) clearly implies that restricting the parameter space does not affect the sufficiency of a σ-algebra. Formally, we have:

Property 1: If $\mathcal{T}$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Theta}$ then $\mathcal{T}$ is also sufficient for any submodel $\mathbf{M}_{\mathcal{X}}^{\Theta_0}$ once $\Theta_0 \subset \Theta$.

Littaye-Petit et al (1969) have shown an important property of transitivity, namely:

Property 2: If $\mathcal{T}_1$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Theta}$ and $\mathcal{T}_2$ is sufficient for $\mathbf{M}_{\mathcal{T}_1}^{\Theta}$, then $\mathcal{T}_2$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Theta}$.

Property 3: If $\mathcal{T}_1$ and $\mathcal{T}_2$ are both sufficient and one of them contains the null-sets, then $\mathcal{T}_1 \cap \mathcal{T}_2$ is also sufficient.

It was shown by Halmos and Savage (1949) that when $\mathbf{M}_{\mathcal{X}}^{\Theta}$ is dominated by a σ- finite measure m, there exists a privileged equivalent probability measure, say P_0, such that for each $A \in \mathcal{X}$, a version of the conditional probability $P_0(A \mid \mathcal{T})$ may be used as the version common to $\{\dot{P}(A \mid \mathcal{T}, \theta) : \ \theta \in \Theta\}$. This common version is used to prove the Neyman factorisation theorem. As a matter of fact, $\mathcal{T}$ is sufficient if and only if for each $\theta \in \Theta$, there exists a version of $\dfrac{dP^\theta}{dP_0}$, say h^θ, wich is $\mathcal{T}$-measurable. The factorisation criteria is then obtained as follows:

$$\frac{dP^\theta}{dm} = \frac{dP^\theta}{dP_0} \cdot \frac{dP_0}{dm} = h^\theta \cdot \frac{dP_0}{dm} \tag{3}$$

This factorization criteria entails the following properties:

Property 4: In the dominated case, if $\mathcal{T}_1$ is sufficient and $\mathcal{T}_2 \supset \mathcal{T}_1$, then $\mathcal{T}_2$ is also sufficient.

Property 5: In the dominated case, the minimal sufficient statistic exists.

In the undominated case, properties 4 and 5 may fail to be satisfied; in particular, a statistic containing a sufficient statistic may failed to be sufficient (see Burkholder (1961)) and a minimal sufficient statistic may not exist (see Pitcher (1957)).

Let us now turn to the concept of identification.

Definition 2 *A parameter function $g(\theta)$ is said to be identified in $\mathbf{M}_{\mathcal{T}}^{\Theta}$ or, equivalently, $\mathcal{T}$-identified, if:*

$$P_{\mathcal{T}}^{\theta_1} = P_{\mathcal{T}}^{\theta_2} \Longrightarrow g(\theta_1) = g(\theta_2) \tag{4}$$

or more explicitly:

$$[\,\forall B \in \mathcal{T} : \ \ P(B \mid \theta_1) = P(B \mid \theta_2)\,] \Longrightarrow g(\theta_1) = g(\theta_2) \tag{5}$$

Definition 2 is eventually extended as follows:

Definition 3 *A parametrization θ is identified if the identity function $g(\theta) = \theta$ is identified in the sense of definition 2 and a model $\mathbf{M}_{\mathcal{X}}^{\Theta}$ is identified if its initial parametrization, $\theta \to P^{\theta}$, is identified.*

Identification of a parameter θ, or of a parameter function $g(\theta)$, is known to be a necessary condition for the existence of a consistent estimator (see e.g. Bunke and Bunke (1974)), or by a similar argument, of an unbiased estimator. It is eventually considered as a basic issue in statistical modelling, particularily for non experimental data, where a (function of a) parameter may be interpreted through being a limit, in probability, or an expectation, of a given statistic (see e.g. Koopmans and Reiersøl (1950) or Roehrig (1988)).

Definition 2 implies that any function of an identified parameter is identified. More specifically:

$$g \text{ is } \mathcal{T}\text{-identified} \Longrightarrow \forall f : f \circ g \text{ is } \mathcal{T}\text{-identified.} \tag{6}$$

Furthermore, if g is $\mathcal{T}_2$-identified then g is $\mathcal{T}_1$-identified for every $\mathcal{T}_1 \supset \mathcal{T}_2$. In particular:

$$g \text{ is } \mathcal{T}\text{-identified} \Longrightarrow g \text{ is } \mathcal{X}\text{-identified.} \tag{7}$$

Remarks

1- Barankin (1961) refers to definition 2 as $g(\theta)$ being a *necessary parameter* for $\mathcal{T}$. We stick however to the more frequent use of an "identified" parameter.

2- The present framework insists on two aspects. Firstly, identification is a property of a parametrization rather than of a model, understood as the set $\mathcal{P}^{\Theta}$ of sampling distributions. Secondly, the identification of θ is treated as a particular case of the more general issue of identifying an arbitrary function of the parameter.

3- At contrast with the sufficiency concept in the sample space wich is unsatisfactory for undominated statistical models, the identification concept is not badly affected by the non-dominance of the sampling model.

As a matter of fact, the concept of sufficiency of a parameter for a statistic may be viewed as a reciprocal of identification, in view of the following definition.

Definition 4 *(Barankin 1961) A parameter function $g(\theta)$ is said to be sufficient in $\mathbf{M}_{\mathcal{T}}^{\Theta}$ or, equivalently, $\mathcal{T}$-sufficient, if:*

$$g(\theta_1) = g(\theta_2) \implies P_{\mathcal{T}}^{\theta_1} = P_{\mathcal{T}}^{\theta_2} \tag{8}$$

Heuristically, let us assume that $\mathcal{T}$ is generated by a statistic T, the sampling distribution of wich is representable by a density $p(t \mid \theta)$. The T-sufficiency of $g(\theta)$ means that the density of T depends only on the value of g and may accordingly be parametrized by $\gamma = g(\theta)$, viz:

$$p(t \mid \theta) = p(t \mid \gamma) \quad \forall \theta \in g^{-1}(\gamma) \tag{9}$$

or equivalently:

$$p(x \mid \theta) = p(t \mid \gamma)\, p(x \mid t, \theta) \quad \forall \theta \in g^{-1}(\gamma) \tag{10}$$

Clearly, definition 4 enjoys the following property: If g is $\mathcal{T}_1$-sufficient then g is $\mathcal{T}_2$-sufficient for every $\mathcal{T}_2 \subset \mathcal{T}_1$, in particular:

$$g \text{ is } \mathcal{X}\text{-sufficient} \Longrightarrow g \text{ is } \mathcal{T}\text{-sufficient.} \tag{11}$$

Property (8) actually describes a binary relationship between $\mathcal{T}$ and g. Thus Picci (1977) refers to (8) as a function g being *unresolvable* for $\mathcal{T}$ and Basu (1977) as a statistic $\mathcal{T}$ being *g-oriented*. Later on, we shall also make use of Ξ_g, the set of all g-oriented statistics, i.e. the set of all $\mathcal{T}$ satisfying (8) for a given function g. Next theorem, useful for later results, gives a characteristic property for the set Ξ_g of all g-oriented statistics. It states that if $\mathcal{T} \in \Xi_g$ and θ is identified in $\mathbf{M}_{\mathcal{X}}^{\Theta}$, then for every two distinct parameter values having a same value of g, there is no common version of their $\mathcal{T}$-conditional distributions.

Theorem 1 *If θ is $\mathcal{X}$-identified and $\mathcal{T} \in \Xi_g$, then $\forall\, \theta_1 \neq \theta_2$:*

$$g(\theta_1) = g(\theta_2) \implies \exists A \in \mathcal{X} : \; \dot{P}(A \mid \mathcal{T}, \theta_1) \cap \dot{P}(A \mid \mathcal{T}, \theta_2) = \emptyset \tag{12}$$

The meaning of theorem 1 may be viewed, in the simplified framework of statistics representable by densities, as follows: Once T is a g-oriented statistics, then $p(t \mid \theta_1) = p(t \mid \theta_2)$ for each $(\theta_1, \theta_2) \in \Theta^2$ such that $g(\theta_1) = g(\theta_2)$. Consider now the conditional densities $p(x \mid t, \theta_1)$ and $p(x \mid t, \theta_2)$. If they were equal, θ would not be identified; equivalently, once θ is identified, these two conditional densities should be different.

Basu (1977) has called "T is *specific ancillary* for β" the situation where T is g-oriented along with $\beta = h(\theta)$ is a variation free complement of $\gamma = g(\theta)$ (See next section for details on the concept of a variation free complement). Note however that the fact that θ may be reparametrized into (β, γ) does not affect the structure of (10), but gives only a characteristic property of β. In such case, equation (10) becomes:

$$p(x \mid \theta) = p(t \mid \gamma)\, p(x \mid t, \beta, \gamma) \tag{13}$$

Definitions (2) and (4) clearly refer to parametrized statistical models in the form of (1). This implies that the initial parametrization, $\theta \rightarrow P^\theta$, is trivially sufficient. *For example*, in a Gauss-Markov approach to a standard linear model such as $y = Z\beta + \epsilon$, a sufficient parametrization would be $\theta = (\beta, P_\epsilon^{\sigma^2}) \in I\!R^k \times \mathcal{L}_{2,0}$ (where $\mathcal{L}_{2,0}$ is the set of zero-expectation finite variance real-valued random variable) whereas an identified function of interest would be $g(\theta) = (\beta, \sigma^2)$. Thus, it is not unusual to consider partially specified models where an identified function of interest is explicitly specified without making a sufficient parametrization explicit.

Combining definitions 2 and 4, Barankin (1961) introduces the concept of a *minimal sufficient* parameter. Such a concept is also called *maximal identifiable* (or *maximal unresolvable*) parameter in Picci (1977).

Definition 5 *(Barankin 1961) A parameter function $g(\theta)$ is said to be minimal sufficient in $\mathbf{M}_{\mathcal{T}}^{\Theta}$ or, equivalently, $\mathcal{T}$-minimal sufficient, if it is both sufficient and identified in $\mathbf{M}_{\mathcal{T}}^{\Theta}$.*

Thus, θ is $\mathcal{X}$-identified is equivalent to θ is $\mathcal{X}$-minimal sufficient because θ is trivially $\mathcal{X}$-sufficient, whereas an $\mathcal{X}$-identified function $g(\theta)$ is not, in general, $\mathcal{X}$-sufficient unless $g(\theta)$ were also injective. Next theorem gives a condition under which the two properties in (7) are equivalent for $g(\theta) = \theta$. Florens et al. (1990, pp: 162) established a similar results in the bayesian framework.

Theorem 2 *If the statistic $\mathcal{T}$ is sufficient then:*

$$\theta \text{ is } \mathcal{T}\text{-identified} \Longleftrightarrow \theta \text{ is } \mathcal{X}\text{-identified.} \tag{14}$$

According to the two contexts given in the introduction, theorem 2 may be understood as follows. From *context* 2, the theorem states that when we reduce an "identified" complete model by marginalizing w.r.t a sufficient statistic $\mathcal{T}$, we do not lose the identification property. In other words, if the observable vector $\mathcal{T}$ is sufficient for the structural model $\mathbf{M}^{\Theta}_{\mathcal{X}}$, it is equivalent to check the identification of θ in that structural model or in the statistical one.

From *context* 1, theorem 2 may be interpreted as follows: For a given identified model $\mathbf{M}^{\Theta}_{\mathcal{X}}$, a necessary condition for a statistic $\mathcal{T}$ to be sufficient for $\mathbf{M}^{\Theta}_{\mathcal{X}}$ is that θ be identified in $\mathbf{M}^{\Theta}_{\mathcal{T}}$. This fact may provide an easy way for detecting non-sufficient statistics. Next example provides an illustration:

Example 1 Let $X_i \sim IN(\mu_i, 1)$ $\;i = 1, 2, ..., n$. Let $T_i = X_1 - X_{i+1}$ with $i = 1, 2, ..., n-1$ and consider the statistic $T = (T_1, T_2, ..., T_{n-1})$. To show that the statistic T is not sufficient, it suffices (by theorem 2) to show that θ is not T-identified. As matter of fact:

$$T_i \sim N(\mu_1 - \mu_{i+1}, 2) \;\; \forall i = 1, 2, .., n. \tag{15}$$

$$\forall i \neq j : \;\; cov(T_i, T_j) = 1 \tag{16}$$

Now let us consider $\theta^{'} = (\mu^{'}_1, \mu^{'}_2, ..., \mu^{'}_n)$ and $\theta^{"} = (\mu^{"}_1, \mu^{"}_2, ..., \mu^{"}_n)$ such that $\mu^{"}_i = \mu^{'}_i + k$ $\forall i = 1, 2, ..., n$, with k is a constant different from zero. Clearly, $\theta^{'} \neq \theta^{"}$ but the marginal distribution of T is the same for these two different values. Then θ is not T-identified and consequently T is not sufficient. Notice that this argument is an alternative easier than computing the T-conditional distribution of X and checking whether it depends or not on θ.

3 Partial sufficiency

In many situations, only a non injective function of θ, say $g(\theta)$, is of interest. A general theory,i.e. basic concepts and general theorems, similar to the injective case, as reminded in section 2, seems to be desirable. In this section we first discuss the concept of a statistic sufficient for a paramater of interest, $g(\theta)$ and call it partial sufficiency; next we examine how far the propositions of section 2 still apply for the case of partial sufficiency.

The basic idea of partial sufficiency is to find a statistic sufficient (w.r.t the complete observation) to make inference on the parameter of interest, defined here as a non injective function of the complete parameter. Such a statistic should accordingly contain all the relevant information about the parameter of interest. Following Fraser (1956), Barndorff-Nielsen (1978) has defined a concept of S-sufficiency. This definition is essentially based on two fundamental concepts. The first one, introduced in section 2, concerns the notion of *sufficient parameter* and the second one is related to the notion of a *variation free reparametrization.* This definition has been criticised by some authors like Basu (1977) and Dawid (1975). Our contribution to this area is to give some further properties of S-sufficiency. In particular, we establish the connection between S-sufficiency and the identification concept (Extension of theorem 2). Furthermore, we review the properties of the standard concept of sufficiency, as displayed in section 2, and we check wich ones also hold for S-sufficiency.

Firstly, let us introduce the notion of a g-partition. Such a technical tool provides us with a deeper understanding of the variation free concept introduced by Barndorff-Nielsen (1978) and makes the conditions of applicability of S-sufficiency more explicit.

3.1 g-partitions and variation free reparametrization

As mentionned above, the concept of S-sufficiency is based on two basic issues. That of a parameter of interest being sufficient for a statistic and that of a variation free reparametrization. A deeper understanding of this last concept seems therefore appropriate. This is the object of this subsection.

The following theorem offers an alternative view to the concept of a variation free reparametrization.

Theorem 3 *Let h and g be two functions defined on Θ, with range space $h(\Theta)$ and $g(\Theta)$ respectively. The following two properties are equivalent:*

i) (h, g) *is injective on Θ and surjective on a factor space, i.e.*

$$\{(h(\theta)\,,\,g(\theta)):\ \theta \in \Theta\} = h(\Theta) \times g(\Theta)$$

ii) $\forall(\beta, \gamma) \in h(\Theta) \times g(\Theta):\ Card\,\{\,h^{-1}(\beta) \cap g^{-1}(\gamma)\,\} = 1$

where the symbol $Card\{.\}$ means the Cardinality of a set.

Definition 6 *Under any one of the two conditions in theorem 3, the transformation (h, g) is said to be a variation free reparametrization or, equivalently, h and g are said to be (mutually) variation free complements.*

This concept, introduced by Barndorff-Nielsen (1978), was motivated by property (i), hence the expression *variation free* or *variation independent.* Under the conditions of theorem 3, if h (resp. g) is injective, $g(\Theta)$ (resp. $h(\Theta)$) reduces to a singleton, i.e. g (resp. h) is a constant function. In the sequel, we denote by V_g as the set of all the variation free complements of a given parameter of interest g.

As a first use of the concept of a variation free complement of a given function, let us consider a particular case of theorem 1, namely when the statistical model is dominated, so that the sampling distributions are representable by densities, and when there exists a variation free complement of the parameter of interest. We then obtain next proposition wich is a corollary of theorem 1, as a matter of fact a rewritting for a particular case.

Proposition 1 *Suppose that, in a dominated model, the statistic T is g-oriented, and that $h(\theta)$ is a variation free complement of $g(\theta)$, namely:*

$$p(x \mid \theta) = p(t \mid \gamma)\, p(x \mid t, \beta, \gamma) \quad (\beta, \gamma) \in h(\Theta) \times g(\Theta) \tag{17}$$

Then, if θ is identified in $p(x \mid \theta)$, $\beta = h(\theta)$ is identified in $p(x \mid t, \beta, \gamma)$ for any fixed value of $\gamma = g(\theta)$, i.e.

$$\beta_1 \neq \beta_2 \Longrightarrow p(x \mid t, \beta_1, \gamma) \neq p(x \mid t, \beta_2, \gamma) \quad \forall \gamma \in g(\Theta) \tag{18}$$

In other words, theorem 1 generalises proposition 1 to the non dominated and the non variation free cases.

Let us now construct a variation free complement parameter from a given non injective function g. We start with the following equivalence relation defined on Θ:

$$\theta_1 \sim_g \theta_2 \iff g(\theta_1) = g(\theta_2) \tag{19}$$

The quotient set generated by (19) is a partition on Θ, denoted by Θ/g, and may be represented as $\Theta/g = \left\{ g^{-1}(\gamma) : \quad \gamma \in g(\Theta) \right\}$. Let us accordingly introduce a concept of a g-section. This is a subset of Θ made of exactly one element from each equivalence class in Θ/g, namely:

Definition 7 *A subset Θ^*_g of Θ is said to be a g-section of Θ if:*

$$\forall \gamma \in g(\Theta) : \; Card\{g^{-1}(\gamma) \cap \Theta^*_g\} = 1 \tag{20}$$

On every g-section, g is injective and has the same image, namely $g(\Theta^*_g) = g(\Theta)$. So that all the g-sections are in bijection. This means that if $\Theta^*_{g,1}$ and $\Theta^*_{g,2}$ are two g-sections of Θ, then $\forall \theta_1 \in \Theta^*_{g,1}$ there exists a unique $\theta_2 \in \Theta^*_{g,2}$ such that $\theta_1 \sim_g \theta_2$.

Remark An alternative approach is to define a g-section as any function $m : \Theta/g \longrightarrow \Theta$ such that $v \circ m$ is the identity on Θ/g, where v is the canonical surjection from Θ to Θ/g. That the two approaches are equivalent may be viewed through the equality: $m(\Theta/g) = \Theta^*_g$.

In next definition, we introduce a notion of a g-partition, viewed as a technical tool for constructing a variation free complement of a given parameter function $g(\theta)$, namely:

Definition 8 *Let $\Pi = \{\Theta^*_{g,\beta} : \quad \beta \in B\}$ be a parametrized family of g-sections of Θ. Π is said to be a g-partition of Θ if the elements of Π form a partition of Θ.*

As matter of fact, a g-partition is one way of considering a variation free complement of a given function g. In particular their existence is equivalent, in view of the following theorem.

Theorem 4 *The following two statements are equivalent:*

i) A variation free complement of g exists.

ii) A g-partition of Θ exists.

When the parameter space Θ is finite, a necessary and sufficient condition for the existence of a g-partition (and thus by theorem 4 of a variation free complement of g) is that all the equivalence classes defined by g have the same cardinality. That a variation free complement may not exist for a given function $g(\theta)$, may be viewed in the following example.

Example 2 Suppose that $\theta = (\mu_1, \mu_2) \in I\!R^2$ and $g(\theta) = \mu_1 - \mu_2$; then $h(\theta) = \mu_1 + \mu_2$ is a variation free complement of $g(.)$. Suppose now that $(\mu_1, \mu_2) \in [0,1]^2$ along with the same $g(\theta) = \mu_1 - \mu_2$. The absence of a variation free complement of $g(.)$ is due to the fact that the sets $g^{-1}(\gamma)$ $\gamma \in [-1,1]$ have not a constant cardinality. For instance $g^{-1}(1) = \{(1,0)\}$ and $g^{-1}(-1) = \{(0,1)\}$ are each reduced to a singleton making accordingly impossible the construction of disjoints g-sections and then of a g-partitions.

3.2 S-sufficiency

Let us now consider a parameter of interest described by a non injective function, $g(\theta)$. A simple corollary of theorem 4 entails the following:

Proposition 2 *Let $\mathcal{T}$ be a statistic of $\mathcal{X}$. The following two statements are equivalent:*

i) There exists a variation free complement of g, say h, such that:

$$\forall \beta \in h(\Theta) : \bigcap_{h^{-1}(\beta)} \dot{P}(A \mid \mathcal{T}, \theta) \neq \emptyset \tag{21}$$

ii) There exists a g-partition of Θ*, say* $\Pi = \{\Theta^*_{g,\beta} : \ \beta \in B\}$*, such that* $\mathcal{T}$ *is sufficient for each submodel* $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,\beta}}$.

Basu (1977) has called "T is *specific sufficient* for $\gamma = g(\theta)$" the condition (21). Following Fraser (1956), Barndorff-Nielsen (1978) has defined a concept of S-sufficiency as follows:

Definition 9 *A statistic* $\mathcal{T}$ *of* $\mathcal{X}$ *is* S*-sufficient for* g *if* $\mathcal{T}$ *is* g*-oriented (i.e.* $\mathcal{T} \in \Xi_g$*) and any one of conditions of Proposition 2 holds.*

Heuristically, to say that a statistic T is S-sufficient for a given g means the following: θ accepts a variation free reparametrization $\theta \leftrightarrow (\beta, \gamma)$, where $\beta = h(\theta)$ is a variation free complement of $\gamma = g(\theta)$, such that the joint density $p(x \mid \theta)$ can be factorized into:

$$p(x \mid \theta) = p(t \mid \gamma)\, p(x \mid t, \beta) \qquad (\beta, \gamma) \in h(\Theta) \times g(\Theta) \tag{22}$$

Thus, the new requirement added to the g-orientedness of the statistic T, is that T operates a marginal-conditional decomposition of the sampling process, characterized by a variation free reparametrization, one of element of wich is the parameter of interest $\gamma = g(\theta)$. The same structure as (22) has also been called a *Cut* by Barndorff-Nielsen (1978) when γ is not necessarily a parameter of interest. He also noticed that the standard concept of sufficiency (resp. ancillarity) corresponds to a cut when $g(\theta)$ is injective (resp. $g(\theta)$ is a constant function). Also, Basu (1977) has called the structure of (22) as T is *p-sufficient* for γ or T is *p-ancillary* for β. Next example illustrates the concept of S-sufficiency.

Example 3 Let $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{(2)}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$ with $\theta = (\mu_1, \mu_2) \in I\!R^2$ and consider $g(\theta) = \mu_1 - \mu_2$ as the parameter of interest. To show that the statistic $T = X_1 - X_2$ is S-sufficient for $g(\theta)$, Let us consider the following g-partition:

$$\Pi = \{\Theta^*_{g,\beta} : \ \beta \in I\!R\} \text{ with } \Theta^*_{g,\beta} = \{(\mu_1, \mu_2) \in I\!R^2 : \ \mu_1 + \mu_2 = \beta\}. \tag{23}$$

Clearly (22) is satisfied with $h(\theta) = \mu_1 + \mu_2$ because $(T \mid \theta) \sim N(g(\theta), 2)$, i.e. $T \in \Xi_g$ and the T-conditional distribution of X depends only on $h(\theta)$, namely:

$$(X \mid T, \theta) \sim N_{(2)} \left(\frac{1}{2} \left(\begin{array}{c} h(\theta) + T \\ h(\theta) - T \end{array} \right), \left(\begin{array}{cc} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{array} \right) \right) \tag{24}$$

Thus, $\forall \, \Theta^*_{g,\beta} \in \Pi \, : \, T$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,\beta}}$. The g-partition in (23) satisfies condition (ii) of proposition 2. Thus, by definition 9, T is S-sufficient for the parameter function $\mu_1 - \mu_2$ when $(\mu_1, \mu_2) \in I\!R^2$.

■

For many other examples, see, among others, Basu (1977), Barndorff-Nielsen (1978), Fraser (1956) or Godambe (1980). In what follows, we give an example of statistics satisfying only one condition in definition 9.

Example 4 Let $X = (X_1, X_2, ..., X_n)$ with $X_i \sim IN(\mu, \sigma^2)$ and consider the statistics $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$. Here $\theta = (\mu, \sigma^2)$. Classical results for the multivariate normal distribution show that:

$$\overline{X} \mid \mu, \sigma^2 \;\; \sim \;\; N(\mu, \frac{\sigma^2}{n}) \tag{25}$$

$$X \mid \overline{X}, \mu, \sigma^2 \;\; \sim \;\; N_{(n)}(\overline{X} \, i_n, \, \sigma^2 [I_n - \frac{1}{n} i_n i_n^{'}]) \tag{26}$$

$$S^2 \mid \mu, \sigma^2 \;\; \sim \;\; \frac{\sigma^2}{n-1} \chi^2_{n-1} \tag{27}$$

Where I_n is the $(n \times n)$ identity matrix, $i_n = (1, 1, ..., 1)' \in I\!R^n$.

Case 1: The parameter of interest is $g(\theta) = \mu$

Relation (26) shows that the $\overline{X}$-conditional probability depend only on σ^2 wich is a variation free complement of μ. So, according to Basu's terminology, $\overline{X}$ is *specific sufficient* for μ (satisfies conditions of proposition 2). However, relation (25) shows that the sampling probabilities of $\overline{X}$ depend on both μ and σ^2. So $\overline{X}$ is not μ-oriented and consequently not S-sufficient for μ.

Case 2: The parameter of interest is $g(\theta) = \sigma^2$

For each (μ, σ^2) the S^2-conditional probability of X may be decomposed as follows:

$$
\begin{aligned}
p(x \mid s^2, \mu, \sigma^2) &= p(x \mid \overline{x}, s^2, \mu, \sigma^2)\, p(\overline{x} \mid s^2, \mu, \sigma^2) \\
&= p_0(x \mid \overline{x}, s^2)\, p(\overline{x} \mid \mu, \sigma^2) \qquad (28)
\end{aligned}
$$

Where $p_0(x \mid \overline{x}, s^2)$ represents a common version of all (w.r.t (μ, σ^2)) the $(\overline{X}, S^2)$-conditional probabilities (this common version exists because $(\overline{X}, S^2)$ is *sufficient* for (μ, σ^2)). The second term in the right hand side of (28) is obtained by using the sampling independence property between $\overline{X}$ and S^2. Relation (27) shows that S^2 is σ^2-oriented and, according to Basu's terminology, *specific ancillary* for μ. However, relations (25) and (28) show that the S^2-conditional probability of X depend on μ but also on the parameter of interest σ^2. So S^2 is not S-sufficient for σ^2.

■

In what follows, we consider some properties of the concept of S-sufficiency. A first question to be asked is the following: Why S-sufficiency does not require equation (21) to be satisfied for all the variation free complements of g, i.e. $\forall h \in V_g$? Next Theorem shows that, for a standard family of statistical models, this requirement is not possible, because such an h is essentially unique.

Theorem 5 *Let $\mathbf{M}_{\mathcal{X}}^{\Theta}$ be an identified homogeneous model, i.e. all the probability measures in the model have the same null-sets. Let $\mathcal{T}$ be a g-oriented statistic and h a variation free complement of g making of $\mathcal{T}$ an S-sufficient statistic for g. Then, h is unique up to an injective transformation.*

Next theorem can be considered as a generalization of theorem 2, for the case where the parameter of interest may be a non injective function of the complete parameter and a variation free complement of the parameter of interest is made explicit, as in the S-sufficiency. This theorem also makes the relationship between S-sufficiency and the identification explicit.

Theorem 6 *If the statistic $\mathcal{T}$ is S-sufficient for g then:*

$$g \text{ is } \mathcal{T}\text{-identified} \Longleftrightarrow g \text{ is } \mathcal{X}\text{-identified}. \qquad (29)$$

In the case of partial observability, $\mathbf{M}_{\mathcal{X}}^{\Theta}$ represents a structural model where X involves latent variables, along with other manifest variables. The statistic $T = f(X)$ represents the manifest variables. $\mathbf{M}_{\mathcal{T}}^{\Theta}$, the image of $\mathbf{M}_{\mathcal{X}}^{\Theta}$ under the transformation f, is the statistical model. A sufficient parametrization in that statistical model $\mathbf{M}_{\mathcal{T}}^{\Theta}$ is often described by a non injective function of the complete parameter (characterizing the structural model), say $g(\theta)$. From *context* 2, the practical problem is to obtain conditions ensuring that $g(\theta)$ is identified in $\mathbf{M}_{\mathcal{T}}^{\Theta}$. Theorem 6 gives such conditions. It states that if we reduce a complete model $\mathbf{M}_{\mathcal{X}}^{\Theta}$ by marginalizing w.r.t an S-sufficient statistic for g, say $\mathcal{T}$, then we do not loose the identification of g in that reduced model $\mathbf{M}_{\mathcal{T}}^{\Theta}$. Moreover, from *context* 1, theorem 6 gives necessary conditions for a statistic $\mathcal{T}$ to be S-sufficient for g.

Let us now check wich properties of sufficiency for a complete parameter, as reviewed in section 2, also hold for S-sufficiency.

Property 1 does not hold any more for S-sufficiency because S-sufficiency crucially depends on the non-emptiness of the set V_g. Thus in example 3, V_g is not empty for $\Theta = I\!R^2$ but is empty for $\Theta = [0,1]^2 \subset I\!R^2$.

Property 2 still holds for S-sufficiency, in view of the following theorem.

Theorem 7 *Let $\mathcal{T}_1$ be S-sufficient for g in $\mathbf{M}_{\mathcal{X}}^{\Theta}$ and let $\mathcal{T}_2$ be S-sufficient for g in the reduced model $\mathbf{M}_{\mathcal{T}_1}^{\Theta}$. Then $\mathcal{T}_2$ is S-sufficient for g in the complete model $\mathbf{M}_{\mathcal{X}}^{\Theta}$.*

Theorem 7 means the following: Let us reduce the complete model by marginalizing on an S-sufficient statistic, say $\mathcal{T}_1$, and consider the reduced model $\mathbf{M}_{\mathcal{T}_1}^{\Theta}$. If there exists a statistic, say $\mathcal{T}_2$, wich is S-sufficient for g in that reduced model, then it will also be S-sufficient in a complete model. For a particular case of statistics representable by densities, let T_1 and T_2 be two statistics such that $T_2 = f(T_1)$. The two conditions of theorem 7, i.e. T_1 is S-sufficient for g in the complete model and T_2 is S-sufficient for g in the reduced model are respectively described as follows:

$$p(x \mid \theta) = p(t_1 \mid \gamma)\, p(x \mid t_1, \beta) \tag{30}$$
$$p(t_1 \mid \gamma) = p(t_2 \mid \gamma)\, p(t_1 \mid t_2) \tag{31}$$

Combining equations (30) and (31) and using the fact that T_2 is a reduction of T_1 implies that:

$$\begin{aligned} p(x \mid \theta) &= p(t_2 \mid \gamma)\, p(t_1 \mid t_2)\, p(x \mid t_1, \beta) \\ &= p(t_2 \mid \gamma)\, p(x \mid t_2, \beta) \end{aligned} \tag{32}$$

Equation (32) clearly means that T_2 is S-sufficient for g in the complete model.

Property 4 does not hold any more for S-sufficiency because of the following simple corrolary of theorem 6.

Corollary 1 *If θ is $\mathcal{X}$-identified and $\mathcal{T}$ is sufficient, then $\mathcal{T}$ is not S-sufficient for any non injective function $g(\theta)$.*

Indeed, the fact that θ is $\mathcal{X}$-identified and $\mathcal{T}$ is sufficient implies by theorem 2 that θ is a minimal sufficient parameter for $\mathcal{T}$. This also implies that every non injective function, say g, cannot be minimal sufficient for $\mathcal{T}$, wich also implies by theorem 6 that $\mathcal{T}$ is not S-sufficient for that function g. Therefore if $\mathcal{T}$ is an S-sufficient statistic for an non injective function g, then the whole sample $\mathcal{X}$ will never be S-sufficient for such function g because of corollary 1.

4 Conclusion

The S-sufficiency concept crucially depends on the nonemptiness of the two sets V_g and Ξ_g, i.e. the set of all variation free complements of g and the set of all g-oriented statistics. If for a given parameter of interest $\gamma = g(\theta)$, any one of these two sets is empty, S-sufficiency is not applicable.

Condition $\Xi_g \neq \emptyset$ ensures that there exists a g-oriented statistic $\mathcal{T}$ such that the marginal model generating $\mathcal{T}$ contains information only about the parameter of interest or, equivalently, that $\gamma = g(\theta)$ is a sufficient parameter for $\mathcal{T}$. That condition $\Xi_g \neq \emptyset$ is not sufficient in itself is due, among others, to the fact that a statistic containing $\mathcal{T}$ may not be g-oriented or that a statistic sufficient for the complete parameter may not be sufficient for a non injective function of θ (see corollary 1).

Adding the condition $V_g \neq \emptyset$ to the condition $\Xi_g \neq \emptyset$ is not enough to produce usefull concept: S-sufficiency requires furthermore that the $\mathcal{T}$-conditional model admits a variation free complement of $\gamma = g(\theta)$ as a sufficient parameter. This property ensures that neglecting the $\mathcal{T}$-conditional model does not lose information on the parameter of interest. S-sufficiency is accordingly equivalent to a cut.

S-sufficiency raises several difficulties, one is the proof of the existence and the construction of an S-sufficient statistic. Another one is the existence, or the non existence, and the construction of an optimal cut, i.e. a minimal S-sufficient statistic.

However, as a good and practical property of S-sufficiency (when conditions of its applicability are satisfied) is its connection to the identification problem. In fact, from context 2, given in the introduction, S-sufficiency preserves the identification of the parameter of interest, say $g(\theta)$. In other words, if the observable vector is S-sufficient for $g(\theta)$, then it is equivalent to identify that parameter of interest from the structural model or from the statistical one. This fact gives us an alternative approach to identify the parameter of interest in partially observable models. As a matter of fact, instead of identifying directly $g(\theta)$ from the statistical model, wich is not usually easy, one may first identify $g(\theta)$ in the structural model (often easy to do, because the complete parameter is usually identified in the structural model) and thereafter find conditions under wich the observation is S-sufficient for the parameter of interest. This approach happens to be promising when studying the identification of discrete choice models.

Another aspect of this paper is to show the abundance of overlapping concepts in the statistical literature on partial sufficiency. We payed a particular attention to connect among themselves the many concepts proposed earlier, and eventually we showed that a small number of them are actually different, even when they have been introduced from different perspectives.

Appendix

Proof of Theorem 1

Let us assume that there exist two distinct values $(\theta_1, \theta_2) \in \Theta^2$ such that $g(\theta_1) = g(\theta_2)$ and a statistic $\mathcal{T} \in \Xi_g$ such that, for each $A \in \mathcal{X}$ there exists a $\mathcal{T}$-measurable common version $f_A \in \dot{P}(A \mid \mathcal{T}, \theta_1) \cap \dot{P}(A \mid \mathcal{T}, \theta_2)$. This means that:

$$\begin{aligned} \forall A \in \mathcal{X}, \forall B \in \mathcal{T} :\ P(A \cap B \mid \theta_1) &= \int_B f_A \, dP_{\mathcal{T}}^{\theta_1} \\ &\overset{\mathcal{T} \in \Xi_g}{=} \int_B f_A \, dP_{\mathcal{T}}^{\theta_2} \\ &= P(A \cap B \mid \theta_2) \end{aligned} \tag{33}$$

Taking $B = \mathbf{R}_X$, equation (33) implies that:

$$\forall A \in \mathcal{X} :\ P(A \mid \theta_1) = P(A \mid \theta_2) \tag{34}$$

wich in turn implies that $\theta_1 = \theta_2$ because θ is identified in $\mathbf{M}_{\mathcal{X}}^{\Theta}$.

■

Proof of Theorem 2

The $\Longrightarrow$ is trivial; Thus we only proof the $\Longleftarrow$ part. Thus, let us assume that θ is $\mathcal{X}$-identified but is not $\mathcal{T}$-identified; this means by definition 2 that there exist $(\theta_1, \theta_2) \in \Theta^2$ such that:

$$\theta_1 \neq \theta_2 \text{ and } P_{\mathcal{T}}^{\theta_1} = P_{\mathcal{T}}^{\theta_2} \tag{35}$$

As $\mathcal{T}$ is sufficient in $\mathbf{M}_{\mathcal{X}}^{\Theta}$, there exists, for each $A \in \mathcal{X}$ and each (θ_1, θ_2) satisfying (35), a $\mathcal{T}$-measurable common version f_{A,θ_1,θ_2} such that:

$$\forall B \in \mathcal{T} :\ P(A \cap B \mid \theta_i) = \int_B f_{A,\theta_1,\theta_2} \, dP_{\mathcal{T}}^{\theta_i} \qquad i = 1, 2 \tag{36}$$

Combining (35) and (36), we have:

$$\forall A \in \mathcal{X}, \forall B \in \mathcal{T} :\ P(A \cap B \mid \theta_1) \overset{(36)}{=} \int_B f_{A,\theta_1,\theta_2} \, dP_{\mathcal{T}}^{\theta_1}$$

$$\begin{aligned}
&\overset{(35)}{=} \int_B f_{A,\theta_1,\theta_2}\, dP_{\mathcal{T}}^{\theta_2} \\
&\overset{(36)}{=} P(A \cap B \mid \theta_2)
\end{aligned} \tag{37}$$

Equation (37) implies for $B = \mathbf{R}_X$ that:

$$\forall A \in \mathcal{X} : \;\; P(A \mid \theta_1) = P(A \mid \theta_2) \tag{38}$$

wich in turn implies that $\theta_1 = \theta_2$ because θ is identified in $\mathbf{M}_{\mathcal{X}}^{\Theta}$.

■

Proof of Theorem 3

We first prove that (ii) implies that (h, g) is surjective on the factor space.

$$\begin{aligned}
(\beta, \gamma) \in h(\Theta) \times g(\Theta) &\overset{\text{(ii)}}{\Longrightarrow} Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1 \\
&\Longrightarrow \exists! \theta \in \Theta : \;\; (\beta, \gamma) = (h(\theta), g(\theta)) \\
&\Longrightarrow (\beta, \gamma) \in \{(h(\theta)\,,\, g(\theta)) : \; \theta \in \Theta\}
\end{aligned} \tag{39}$$

Equation (39) implies that $h(\Theta) \times g(\Theta) \subset \{(h(\theta)\,,\, g(\theta)) : \; \theta \in \Theta\}$. But trivially, $h(\Theta) \times g(\Theta) \supset \{(h(\theta)\,,\, g(\theta)) : \; \theta \in \Theta\}$ and so the two sets are equal. Next we show that the function $f = (h, g)$ is injective on Θ, more specifically:

$$\begin{aligned}
f(\theta_1) = f(\theta_2) &\Longrightarrow h(\theta_1) = h(\theta_2) \text{ and } g(\theta_1) = g(\theta_2) \\
&\Longrightarrow \exists (\beta, \gamma) \in h(\Theta) \times g(\Theta) : \;\; (\theta_1, \theta_2) \in h^{-1}(\beta) \cap g^{-1}(\gamma) \\
&\overset{(ii)}{\Longrightarrow} \theta_1 = \theta_2
\end{aligned}$$

Conversely, let us show that (i) implies (ii).

$$\begin{aligned}
(\beta, \gamma) \in h(\Theta) \times g(\Theta) &\overset{(i)}{\Longrightarrow} \exists! \theta \in \Theta : \;\; (h(\theta), g(\theta)) = (\beta, \gamma) \\
&\Longrightarrow h^{-1}(\beta) \cap g^{-1}(\gamma) = \{\theta\} \\
&\Longrightarrow Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1
\end{aligned} \tag{40}$$

■

Proof of Theorem 4

Let h be a variation free complement of g; then the quotient set generated by h i.e. Θ/h is a partition, the elements of wich are trivially g-sections. So, condition (i) implies condition (ii).

Conversely, let us show that every g-partition defines, on the parameter space, a unique function, say h, wich is a variation free complement of $g(\theta)$. Indeed, Let $\Pi = \{\Theta^*_{g,\beta} : \ \beta \in B\}$ be a given g-partition of Θ and let us construct, on Θ, a function h as follows, $h : \ \Theta \longrightarrow B$ such that $h(\theta) = \beta, \ \forall\theta \in \Theta^*_{g,\beta}$. Clearly, $\forall\beta \in B : \ h^{-1}(\beta) = \Theta^*_{g,\beta}$ and Π coincides with the quotient set generated by h, i.e. $\Pi = \{h^{-1}(\beta) : \ \beta \in B\}$. Because Π is a g-partition, the function h satisfies by condition (20):

$$\forall(\beta,\gamma) \in h(\Theta) \times g(\Theta) : \ Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1 \tag{41}$$

Then, h satisfies condition (ii) of theorem 3.

■

Proof of Theorem 5

To prove theorem 5, it suffices to show that if there exist two different g-sections $\Theta^*_{g,1}$ and $\Theta^*_{g,2}$ such that $\mathcal{T}$ is sufficient for both $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,1}}$ and $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,2}}$, then $\Theta^*_{g,1} \cap \Theta^*_{g,2} = \emptyset$.

Let us assume that $\mathcal{T}$ is sufficient for both $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,1}}$ and $\mathbf{M}_{\mathcal{X}}^{\Theta^*_{g,2}}$ with $\Theta^*_{g,1}$ and $\Theta^*_{g,2}$ being two non-disjoints g-sections. This means that for each $A \in \mathcal{X}$, there exist two $\mathcal{T}$-measurable function $f_{i,A} \in \bigcap_{\theta\in\Theta^*_{g,i}} \dot{P}(A \mid \mathcal{T},\theta) \ \ i = 1,2$.

Let $\theta_0 \in \Theta^*_{g,1} \cap \Theta^*_{g,2}$, then:

$$\begin{aligned}\forall A \in \mathcal{X}, \forall B \in \mathcal{T} : \ P(A \cap B \mid \theta_0) &= \int_B f_{1,A}\, dP^{\theta_0}_{\mathcal{T}} \\ &= \int_B f_{2,A}\, dP^{\theta_0}_{\mathcal{T}}\end{aligned} \tag{42}$$

This implies that:

$$\forall A \in \mathcal{X}, \forall B \in \mathcal{T}: \int_B (f_{1,A} - f_{2,A}) dP^{\theta_0}_{\mathcal{T}} = 0 \tag{43}$$

wich also implies that:

$$\forall A \in \mathcal{X}: \; f_{1,A} = f_{2,A} \qquad P^{\theta_0}_{\mathcal{T}} \text{ a.s} \tag{44}$$

Because $\mathbf{M}^{\Theta}_{\mathcal{X}}$ is a homogeneous model, equation (44) implies that:

$$\forall A \in \mathcal{X}: \; f_{1,A} = f_{2,A} \qquad P^{\theta} \text{ a.s } \forall \; \theta \in \Theta^*_{g,1} \cup \Theta^*_{g,2} \tag{45}$$

Equation (45) implies that for each $A \in \mathcal{X}$, $\{f_{1,A}, f_{2,A}\}$ are common versions of $\{\dot{P}(A \mid \mathcal{T}, \theta): \; \theta \in \Theta^*_{g,1} \cup \Theta^*_{g,2}\}$, wich implies that:

$$\forall A \in \mathcal{X}: \quad \bigcap_{\Theta^*_{g,1} \cup \Theta^*_{g,2}} \dot{P}(A \mid \mathcal{T}, \theta) \neq \emptyset \tag{46}$$

Now, equation (46) implies that within each equivalence class defined by g, there exist two distinct values having a same $\mathcal{T}$-measurable common version, more specifically:

$$\begin{aligned} \forall \gamma \in g(\Theta) \qquad & \exists (\theta_1, \theta_2) \in (g^{-1}(\gamma))^2 \text{ with } \theta_1 \neq \theta_2 \text{ and such that :} \\ & \forall A \in \mathcal{X}: \; \dot{P}(A \mid \mathcal{T}, \theta_1) \cap P(A \mid \mathcal{T}, \theta_2) \neq \emptyset \end{aligned} \tag{47}$$

Clearly, equation (47) contradicts Theorem 1 because $\mathcal{T} \in \Xi_g$ and θ is $\mathbf{M}^{\Theta}_{\mathcal{X}}$ identified.

■

Proof of Theorem 6

Because equation (7) makes the other implication trivial, one should only prove that the $\mathcal{T}$-identification of $g(\theta)$ is implied by its $\mathcal{X}$-identification. Let us assume that $g(\theta)$ is not $\mathcal{T}$-identified; this means by definition 2 that there exist two distinct values $(\theta_1, \theta_2) \in \Theta^2$ such that:

$$g(\theta_1) \neq g(\theta_2) \text{ and } P^{\theta_1}_{\mathcal{T}} = P^{\theta_2}_{\mathcal{T}} \tag{48}$$

The S-sufficiency of $\mathcal{T}$ implies the existence of a g-partition, say Π, satisfying: $\forall\, \Theta^*_g \in \Pi, \mathcal{T}$ is sufficient for $\mathbf{M}^{\Theta^*_g}_{\mathcal{X}}$. Let us now consider $(\theta_1, \theta_2) \in \Theta^2$

satisfying (48). Then there exist two g-sections $(\Theta^*_{1,g}, \Theta^*_{2,g}) \in \Pi^2$ such that $(\theta_1, \theta_2) \in \Theta^*_{1,g} \times \Theta^*_{2,g}$, because a g-partition cover all Θ.

Let $\theta_2^{'}$ be an element of $\Theta^*_{1,g}$ satisfying $\theta_2^{'} \sim_g \theta_2$ ($\theta_2^{'}$ exists and is unique by definition of a g-section); then $P_{\mathcal{T}}^{\theta_2} = P_{\mathcal{T}}^{\theta_2^{'}}$ (because $T \in \Xi_g$) wich implies by equation (48) that $P_{\mathcal{T}}^{\theta_1} = P_{\mathcal{T}}^{\theta_2^{'}}$.

The S-sufficiency of $\mathcal{T}$ implies, in particular for $(\theta_1, \theta_2^{'})$, that for each $A \in \mathcal{X}$ there exists a $\mathcal{T}$-measurable common version $f_{A,\theta_1,\theta_2^{'}}$ or more specifically;

$$\begin{aligned} \forall A \in \mathcal{X}, \forall B \in \mathcal{T} : \; P(A \cap B \mid \theta_1) &= \int_B f_{A,\theta_1,\theta_2^{'}} \, dP_{\mathcal{T}}^{\theta_1} \\ &= \int_B f_{A,\theta_1,\theta_2^{'}} \, dP_{\mathcal{T}}^{\theta_2^{'}} \\ &= P(A \cap B \mid \theta_2^{'}) \end{aligned} \tag{49}$$

Equation (49) implies for $B = \mathbf{R}_X$ that:

$$\forall A \in \mathcal{X} : \;\; P(A \mid \theta_1) = P(A \mid \theta_2^{'}) \tag{50}$$

wich in turn implies a contradiction of (48), namely $g(\theta_1) = g(\theta_2^{'}) = g(\theta_2)$ because $g(\theta)$ is $\mathcal{X}$-identified.

■

Proof of Theorem 7

To prove theorem 7 we need the following Lemma.

Lemma 1 *Let $\mathcal{T}_1$ be a g-oriented σ-algebra of $\mathcal{X}$ and $\mathcal{T}_2 \subset \mathcal{T}_1$. If there exists a g-section, say $\Theta^*_{g,0}$ such that:*

$$\forall B \in \mathcal{T}_1 : \;\; \bigcap_{\theta \in \Theta^*_{g,0}} \dot{P}(B \mid \mathcal{T}_2, \theta) \neq \emptyset. \tag{51}$$

Then, every g-section satisfies (51).

Proof of Lemma 1

Let $\Theta^*_{g,1}$ be an arbitrary g-section. Then, $\forall \theta_1 \in \Theta^*_{g,1}$ there exists a unique $\theta_0 \in \Theta^*_{g,0}$ such that $\theta_1 \sim_g \theta_0$ and then $P^{\theta_1}_{\mathcal{T}_1} = P^{\theta_0}_{\mathcal{T}_1}$, because $\mathcal{T}_1 \in \Xi_g$. Wich also implies that:

$$\forall B \in \mathcal{T}_1 : \ \dot{P}(B \mid \mathcal{T}_2, \theta_1) = \dot{P}(B \mid \mathcal{T}_2, \theta_0).$$

and therefore, by equation (51) that:

$$\forall B \in \mathcal{T}_1 : \quad \bigcap_{\theta \in \Theta^*_{g,1}} \dot{P}(B \mid \mathcal{T}_2, \theta) \neq \emptyset. \tag{52}$$

■

Let us now prove theorem 7. $\mathcal{T}_1$ is S-sufficient for g in $\mathbf{M}^{\Theta}_{\mathcal{X}}$ means that there exists a g-partition, say $\Pi = \{\Theta^*_\beta : \ \beta \in H\}$ such that:

$$\forall \beta \in H, \forall A \in \mathcal{X} : \ \bigcap_{\theta \in \Theta^*_\beta} \dot{P}(A \mid \mathcal{T}_1, \theta) \neq \emptyset. \tag{53}$$

$\mathcal{T}_2$ is S-sufficient for g in $\mathbf{M}^{\Theta}_{\mathcal{T}_1}$ implies the existence of a g-section satisfying (51), because $\mathcal{T}_2 \subset \mathcal{T}_1$. This implies by lemma 1 that equation (51) holds for every g-section in Π, more specifically:

$$\forall \beta \in H, \forall B \in \mathcal{T}_1 : \ \bigcap_{\theta \in \Theta^*_\beta} \dot{P}(B \mid \mathcal{T}_2, \theta) \neq \emptyset. \tag{54}$$

Finally, let $\beta \in H$ and $A \in \mathcal{X}$. Let us also denote by $f_A^{\Theta^*_\beta}$ any common version of $\{\dot{P}(A \mid \mathcal{T}_1, \theta) : \ \theta \in \Theta^*_\beta\}$ (This common version exists by (54)). In order to show that $\mathcal{T}_2$ is S-sufficient in $\mathbf{M}^{\Theta}_{X}$, we notice:

$$\begin{aligned} \bigcap_{\theta \in \Theta^*_\beta} \dot{P}(A \mid \mathcal{T}_2, \theta) &= \bigcap_{\theta \in \Theta^*_\beta} \dot{\mathbb{E}}(\dot{P}(A \mid \mathcal{T}_1, \theta) \mid \mathcal{T}_2, \theta) \\ &= \bigcap_{\theta \in \Theta^*_\beta} \dot{\mathbb{E}}(f_A^{\Theta^*_\beta} \mid \mathcal{T}_2, \theta) \neq \emptyset \end{aligned} \tag{55}$$

■

References

AMEMIYA, T., (1954), *Advanced Econometrics*, Basil Blackwell, Oxford.

BAHADUR, R. R., (1954), sufficiency and statistical decision functions, *The Annals of Mathematical Statistics*, Vol. 25, pp. 423-462.

BAHADUR, R. R., LEHMAN, E. L., (1955), Two comments on "sufficiency and statistical decision functions", *The Annals of Mathematical Statistics*, Vol. 26, pp. 139-142.

BAHADUR, R. R., (1955a), A characterization of sufficiency, *The Annals of Mathematical Statistics*, Vol. 26(2), pp. 286-293.

BAHADUR, R. R., (1955b), Statistics and Subfields, *The Annals of Mathematical statistics*, Vol. 26(3), pp. 490-497.

BARANKIN, E. W., (1955b), Sufficient parameters: solution of the minimal dimensionality problem, *Annals of the Institute of Statistical Mathematics*, Vol. 12, pp. 91-118.

BARNDORFF-NIELSEN, O., (1978), *Information and Exponential Families in Statistical Theory*, John Wiley, New York.

BARRA, J.-R., (1981), *Mathematical Basis of Statistics*, Academic Press, New York.

BASU, D., (1977), On the elimination of the nuisance parameters, *Journal of the American Statistical Association*, Vol. 72, pp. 355-366.

BURKHOLDER, D.L., (1961), Sufficiency in the undominated case, *The Annals of the Mathematical Statistics*, Vol. 32, pp. 1191-1200.

BUNKE, H. and O.BUNKE, (1974), Identification and Estimability, *Mathematische Operationsforschung und Statistik*, **5**, 223-233.

DAWID, A. P., (1975), On the concepts of Sufficiency and Ancillarity in the Presence of Nuisance Parameters, *The Journal of the Royal Statistical Society,*, Series B (Methodological), Vol. 37(2), pp. 248-258.

Dawid, A. P., (1980), A Bayesian Look at Nuisance Parameters, In: *Bayesian Statistics. Proceedings of the First International Meeting on Bayesian Staistics,*, edited by Bernardo, J.-M., De Groot, M. H., Lindley, D. V. and Smith, A. F. M, Valencia: University Press, pp. 167-184.

Florens, J.-P., M.Mouchart and J.-M.Rolin, (1990), *Elements of Bayesian Statistics*, Marcel Dekker, New York.

Fraser, D.A.S., (1956), Sufficient statistics with nuisance parameters, *The Annals of the Mathematical Statistics*, Vol. 27, pp. 838-842.

Godambe, V. P., (1980), On sufficiency in the presence of a nuisance parameter, *Biometrika*, Vol. 67, No. 1, pp. 155-162.

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

Hájek, J., (1967), On basic concepts of statistic, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilities*, Vol. 1, pp. 139-162.

Hille, J., (1997), On sufficiency and Blackwell sufficiency, *Statistics & Decisions*, Vol. 15, pp. 319-347.

Kabaila, P., (1998), A note on sufficiency and information loss , *Statistics & Probability Letters*, Vol. 37, pp. 111-114.

Koopmans, T. C. and O.Reiersøl, (1950), The Identification of Structural Characteristics, *The Annals of Mathematical Statistics*, 165-181.

Littaye-Petit, M., Piednoir, J.-L. and Van Custem, B., (1969), Exhaustivité , *Annales de l'Institut Henri Poincaré*, vol V, No. 4, pp. 289-322.

Picci, G., (1977), Some connections between the theory of sufficient statistics and the identifiability problem, *SIAM Journal of applied mathematics*, Vol. **33**(3), pp. 383-398.

Pitcher, T.S., (1957), Sets of measures not admitting necessary and sufficient statistics or subfields, *The Annals of the Mathematical Statistics*, Vol. 28, pp. 267-268.

Ramamoorthi, R. V., Yamada, S., (1983), Neyman factorization theorems for experiments admitting densities, *The Indian Journal of Statistics*, Vol. 45, Series A, No.2, pp. 168-180.

Rémon, M., (1984), On a concept of partial sufficiency: L-sufficiency, *International Statistical Review*, vol 52, pp. 127-136.

Roehrig, C. S., (1988), Conditions for Identification in Nonparametric and Parametric Models, *Econometrica*, Vol. 56, No. 2, 433-447.