J. Ital. Statist. Soc. (1994)
1, pp. 125-151

ON IDENTIFIABILITY OF PARAMETRIC STATISTICAL MODELS

Carlos Daniel Mimoso Paulino*
Universidade Técnica de Lisboa, Depto. de Matemática, Lisboa, Portugal
Carlos Alberto de Bragança Pereira
Universidade de São Paulo, Depto. de Estatística, São Paulo, Brasil

Summary

This is a review article on statistical identifiability. Besides the definition of the main concepts, we deal with several questions relevant to the statistician: parallelism between parametric identifiability and sample sufficiency; relationship of identifiability with measures of sample information and with the inferential concept of estimability; several strategies of making inferences in unidentifiable models with emphasis on the distinct behaviour of the classical and Bayesian approaches. The concepts, ideas and methods discussed are illustrated with simple examples of statistical models.

Keywords: Bayesian inference; categorical data; estimability; linear model; sample information; sample sufficiency.

1. Introduction

The so-called problem of identification occurs in many scientific areas, including some where the physical structure of the phenomenon to be modelled does not involve stochasticity. This diversity justifies the several formulations and meanings that have been given to the terms identifiability/identification. We shall concern ourselves with the most relevant concept to statistics which is related to the possibility of the model parameters being uniquely determined from the distribution of the observed random variables. This concept, which sometimes is termed statistical identifiability or distribution identifiability, will just be called identifiability.

Although this problem occurs in various statistical modelling fields, little attention has been devoted to it in the statistical literature, and even so, very

* *Address for correspondence*: C. D. M. Paulino, Universidade Técnica de Lisboa, Depto. de Matemática, IST, Av. Rovisco Pais, 1096 Lisboa Codex, Portugal, Centro de Análise e Processamento de Sinais da UTL.

dispersely. The study and illustration of the identifiability concept are usually confined to specific fields like econometric models. Reference to this concept sometimes is made in a little incisive way. It is often masked by reference to inferential concepts that are not equivalent to it, in general. In our opinion, all these aspects have somehow restricted the comprehension of central issues of identifiability theory, at least in some sections of the statistical community.

In this article we intend to review the subject matter providing a systematic and unified description of the aspects of identifiability theory, for parametric models, of most relevance to statisticians (see, e.g., Koopmans and Reiersol [1950] and Basu [1983], for non-parametric settings). Readers most interested in the topic for econometric models are advised to read the «classical» book by Fisher (1966) and/or some more recent reference, for instance, Hsiao (1983), which devotes a special attention to dynamical models relevant to time series and contains an extensive bibliography on the subject.

We will deal with the following questions: what is actually an identifiable statistical model? Is it possible to visualize any parallelism between sample sufficiency theory and parametric identifiability theory? What connection can one establish between identifiability and sample information? Are identifiability and estimability only synonyms? Which are the limitations on the inferential procedures for unidentifiable models? How to make inferences possible in unidentifiability situations? If some kind of prior information is to be used, how does it act on the model identification? Are the inferential implications of the lack of identifiability indifferent to the kind of approach (Classical or Bayesian) to be used?

Illustration of the concepts and results to be presented is made through examples, hopefully familiar to the statistician, that describe real situations without unnecessary complications. Some emphasis is given to examples of categorical data because they constitute a potential source of unidenfiability problems stemming from the very genesis of the Multinomial model (from Poisson processes), the use of the log-linear structure in frequent data analyses and, on a less specific plane, from the possible incomplete nature of certain observations. Some problems with diversely incomplete data are analised in detail to the extent that the identifiability question is an indispensable element of the foundations of the incomplete data analysis.

2. Concepts and basic results of the identifiability theory

Let us consider the statistical model $(\mathcal{Y}, \mathcal{A}, \mathcal{P})$ where $\mathcal{Y}$ is the sample space, $\mathcal{A}$ is a σ-algebra defined on it and $\mathcal{P}$ is a family of probability measures on $(\mathcal{Y}, \mathcal{A})$ specified by $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$, where Θ denotes the parametric space.

Definition 2.1. Two points of Θ, ϑ_0 and ϑ_1 are said observationally equivalent (and we write $\vartheta_0 \sim \vartheta_1$) if

$$P_{\vartheta_0}(A) = P_{\vartheta_1}(A), \quad \forall A \in \mathcal{A}.$$

Letting $F(y|\vartheta)$ denote the distribution function of the observed random vector Y, we have $\vartheta_0 \sim \vartheta_1$ iff $F(y|\vartheta_0) = F(y|\vartheta_1)$, $\forall y \in \mathcal{Y}$. In the usual dominated case, the likelihood function is, essentially for all data, constant on all observationally equivalent points. And for each pair of points not equivalent in this sense, it assumes different values at least for one $y \in \mathcal{Y}$.

It is easy to see that $\sim$ is an equivalent relation, so inducing in Θ a partition in the equivalence classes $[\vartheta_0] = \{\vartheta_1 \in \Theta : P_{\vartheta_1} = P_{\vartheta_0}\}$. This partition, termed quotient set of Θ according to $\sim$, will be denoted by $\Theta/\sim$.

Definition 2.2.(a): i) The point $\vartheta_0 \in \Theta$ is said (globally) identifiable if $[\vartheta_0] = \{\vartheta_0\}$. ii) Θ (or the model P) is said identifiable if $[\vartheta] = \{\vartheta\}$, $\forall \vartheta \in \Theta$, i.e., if $\Theta/\sim$ is the finest possible partition, $\Theta/\sim = \{\{\vartheta\}, \vartheta \in \Theta\}$.

Definition 2.2.(a) ii) shows that Θ is not identifiable if there is at least one $\vartheta_1 \in \Theta$ such that $[\vartheta_1]$ is not singular (formed by a single element). Thus, it is possible that Θ is not identifiable and there are points satisfying Definition 2.2.(a) i). Nevertheless, in the usual unidentifiability situations, no equivalence class is singular. When $F(\cdot|\vartheta)$ stands for a family of mixture distributions resulting from mixing a given parametric model by a class Θ of probability distributions, the above definition becomes that of identifiability of mixtures. See Everitt (1985) for this special case which will not be developed here.

The set Θ commonly is a subset of a metric space which allows us to define the weaker concept of local identifiability.

Definition 2.2.(b) iii) The point $\vartheta_0 \in \Theta$ is said locally identifiable if $[\vartheta_0] \cap V(\vartheta_0) = \{\vartheta_0\}$, where $V(\vartheta_0)$ stands for an open neighbourhood of ϑ_0.

This concept can be useful because in certain unidentifiable models, even with no globally identifiable point, a given equivalence class may be singular on a certain neighbourhood. This enables us particulary to establish a relationship to measures of sample information, as seen in Section 5.

According to Definition 2.2, the model identifiability is defined in terms of the mapping $\vartheta \to F(\cdot|\vartheta)$ being injective. This definition clearly shows that the presence or absence of identifiability is a feature of the specification adopted for the data generator model, and so, is independent of the inferen-

tial procedure to be used. Yet, the limitations of this procedure are strongly dependent on that feature, as one will see in Section 6.

The inferential problems stemming from an unidentifiable model does not necessarily imply that nothing can be done. Many unknown aspects of the model can be expressed by functions $\varphi(\vartheta) : \Theta \to \Phi$ that are constant over each observational equivalence class. This means that one may establish a correspondence, not necessarily one-to-one, between $\Theta/\sim$ and each $\varphi(\Theta)$ such that distinct values of $\varphi(\vartheta)$ correspond to distinct values of $F(y|\vartheta)$, $\forall y \in \mathcal{Y}$. These functions are thus identified by the data and then unambiguous inferences about them can be drawn.

The argument above shows that the concept of partition, rather than that of function, is more fundamental for examining identifiability. For such a purpose, the consideration of the following relation between each pair (π_1, π_2) of elements of the set of partitions of Θ is thought of being useful:

Definition 2.3. π_1 is said to be a reduction of π_2 (or that π_1 is thicker than π_2), and one writes $\pi_1 < \pi_2$, if every part of π_1 is a union of parts of π_2.

It is easy to show that $<$ is the wide order relation (reflexive, wide antisymmetric and transitive) associated to the partial order relation «π_1 is a strict reduction of π_2», which in turn expresses the idea that at least one part of π_1 is a union of more than one part of π_2.

In an unidentifiable parametric space, the finest partition of Θ, $\{\{\vartheta\}, \vartheta \in \Theta\}$, cannot be discriminated by the data. By construction, the finest partition which is detectable by the data is $\Theta/\sim$. If a given partition π is a reduction of $\Theta/\sim$, it can be identified by the data inasmuch as distinct parts of π include distinct observational equivalence classes. Every function inducing this kind of partition will be constant in each equivalence class. In short:

Definition 2.4. i) A partition π is said identifiable if π is a reduction of $\Theta/\sim$. ii) A function $\varphi(\vartheta)$ is said identifiable if the corresponding induced partition is identifiable, i.e., $\forall \vartheta_1, \vartheta_0 \in \Theta$, $\vartheta_1 \in [\vartheta_0] \Rightarrow \varphi(\vartheta_1) = \varphi(\vartheta_0)$.

This definition obviously includes, as a particular case, Definition 2.2 ii) and shows that any singular partition (or any constant function) is trivially identifiable. The existence, in unidentifiable models, of an identifiable non-singular partition (non-constant function) makes it possible to define further identifiable partitions (functions), on the basis of the following easily demonstrable theorem:

Theorem 2.1. i) If π_2 is an identifiable partition and $\pi_1 < \pi_2$, then partition

π_1 also is identifiable. ii) If φ is an identifiable function and ψ is a function of φ, then ψ also is identifiable.

This theorem shows that an unidentifiable partition, π_1, cannot be a reduction of an identifiable partition, say π_2, which does not mean that π_1 necessarily is strictly finer than π_2 (recall that the partitions of Θ are only partially ordered by the reduction relation). Under an unidentifiable model, this theorem implies that any identifiable function cannot be one-to-one.

The definition of an identifiable function does not imply that it assumes different values in distinct equivalence classes. Barankin (1960) names every function with such a property as sufficient in the sense that it is «sufficient» to identify identical measures into $\mathcal{P}$. In the same direction, we will say that a partitition is sufficient if each observational equivalence class is a union of parts of such partition. That is to say,

Definition 2.5. i) A partition π is said sufficient if $\Theta/\sim$ is a reduction of π. ii) A function $\varphi(\vartheta)$ is said sufficient if the partition induced by it is sufficient, i.e.,

$$\forall \vartheta_1,\ \vartheta_0 \in \Theta : \varphi(\vartheta_1) = \varphi(\vartheta_0) \Rightarrow \vartheta_1 \in [\vartheta_0].$$

Examination of the parametric sufficiency can be advantageously made through the following characterization (Barankin [1960], Picci [1977]):

Theorem 2.2. A function $\varphi(\vartheta)$ is sufficient iff there exists a function $P^*(\cdot)$ defined in $\varphi(\Theta) \times \mathcal{A}$ such that $P_\vartheta(A) = P^*_{\varphi(\vartheta)}(A)$, $\forall A \in \mathcal{A}$, $\forall \vartheta \in \Theta$.

This result states that the distribution function depends on ϑ only through the sets of a sufficient partition. As a consequence, all values of ϑ belonging to the partition part containing ϑ_0 necessarily are observationally equivalent to ϑ_0. It must be emphasized that this does not mean that the above mentioned element of the sufficient partition contains all of the points of $[\vartheta_0]$. It suffices to think of the partition $\{\{\vartheta\},\ \vartheta \in \Theta\}$ that is always trivially sufficient.

Taking account of Definition 2.4 and 2.5 and properties of the reduction relation, one proves easily that

Theorem 2.3. i) If a partition π_1 is identifiable, then π_1 is a reduction of every sufficient partition. ii) If a function φ is identifiable, then φ is a function of every sufficient function.

Because of the wide antisymmetry of the reduction relation, $\Theta/\sim$ is the only partition which is simultaneously identifiable and sufficient. Using Kadane's (1974) terminology, we shall call it identifying partition. So

Definition 2.6. A function $\psi(\vartheta)$ is said identifying if it induces the identifying partition.

Therefore an identifying function is:

– both a sufficient function that is function of every sufficient function – hence, the designation minimal sufficient used by Barankin (1960);
– and an identifiable function that corresponds to the greatest discrimination of Θ allowed by the data – hence, the designation maximal identifiable by Picci (1977).

Putting Theorems 2.1 and 2.3 together, Kadane's (1974) identifiability characterization is obtained and the designation identifying is justified:

Theorem 2.4. Let $\psi(\vartheta)$ be an identifying function. A function $\varphi(\vartheta)$ is identifiable iff it is function of $\psi(\vartheta)$.

Every identifying function is one-to-one correspondence to any other identifying function. In this way, the elements of $\Theta/\sim$ may be defined through any identifying function $\psi(\cdot)$ by $[\vartheta_0] = \{\vartheta \in \Theta : \psi(\vartheta) = \psi(\vartheta_0)\}$. A general example can be given by the complex-valued function derived from the characteristic function of y, $\psi_t(\vartheta) = E(e^{it'y}|\vartheta)$, where t is a real vector with the same dimension as y.

If Θ is restricted by $\lambda(\vartheta) = 0$, the model is identifiable iff any identifying function $\psi(\vartheta)$ is one-to-one when its domain is the restricted parametric space, say Θ_*. In other words, iff ϑ_0 is the unique solution to the system $\psi(\vartheta) = \psi(\vartheta_0)$, $\lambda(\vartheta) = 0$ for all $\vartheta_0 \in \Theta$. Hence, investigation of identifiability becomes a problem pertaining to uniqueness of the solution to a system of possibly non-linear equations. When Θ is a Euclidean space, the linearization of $\psi(\cdot)$, assumed to be well-behaved functions, enables us, in the light of this argument, to obtain a necessary and sufficient condition for local identifiability. This condition, expressed through the rank of the matrix of the derivatives of $\psi(\cdot)$ and $\lambda(\cdot)$, becomes one for global identifiability if both functions are linear. These results, often used in econometric models (see, e.g. Rothenberg [1971], theors. 5 and 6) may thus be viewed as a consequence of Theorem 2.4 above.

3. Illustrative examples

The identifiability problem has been amply illustrated in the literature by models of interest for several applied areas-factor analysis models, linear models with errors in variables, simultaneous equations models and time series models (see, e.g., Koopmans and Reiersol [1950], Rothenberg [1971], Drèze [1974], Hannan [1971]). Our purpose, here, is to illustrate the notions and results of the proceeding section through statistical models structurally less complex than many econometric models. We will restrict ourselves to situations with complete and/or incomplete data generated by Normal and discrete distributions.

Example 1 (Normal linear model). Let us consider the univariate linear model, $y = X\beta + u$, where X is a $n \times p$ design matrix of rank $r < p < n$, β is a vector of unknown parameters, and u, the unobservable vector of errors, has a Normal distribution with parameters $E(u) = 0$ and $E(uu') = \sigma^2 I_n$ (I_n denotes the identity matrix of dimension n). Define $\vartheta = (\beta\ \sigma^2) \in \Theta = \{\vartheta \in \mathbb{R}^{p+1}: \beta \in \mathbb{R}^p,\ \sigma^2 \in \mathbb{R}_+\}$. Since the distribution of y depends on ϑ through $X\beta$ and σ^2, and X is not of full rank, there exist several distinct values of β compatible with the same mean vector of y. So, β (and a fortiori ϑ) is unidentifiable.

Typical examples of this king of model include the linear regression model with extreme multicollinearity and the ANOVA models in their usual formulation. For instance, let us briefly consider the no-interaction ANOVA II model for a situation with one observation per cell, $y_{ij} = u_{ij}$, $i = 1, \ldots, a$; $j = 1, \ldots, b$; $(a, b > 1)$, where $\mu_{ij} = \eta + \alpha_i + \beta_j$. The vector of cell means $\mu = (\mu_{ij}) = X\beta$, *where* $\beta = (\eta\ \alpha_1 \cdots \alpha_a\ \gamma_1 \ldots \gamma_b)'$, belongs to the subspace of $\mathbb{R}^q$ $(q = ab)$ spanned by the p $(p = a + b + 1)$ columns of X. This subspace, denoted by $\mathcal{M}(X)$, has dimension $r = p - 2$, which implies that the observational equivalence class represented by $\vartheta_0 = (\beta_0\ \sigma_0^2)'$ where $\beta_0 = (\eta^0\ \alpha_1^0 \cdots \alpha_a^0\ \gamma_1^0 \cdots \gamma_b^0)'$ can be defined by

$$
\begin{aligned}
[\vartheta_0] &= \{\vartheta \in \Theta : X_0\beta = X_0\beta_0,\ \sigma^2 = \sigma_0^2\} \qquad (1)\\
&= \{\vartheta \in \Theta : \eta = \eta^0 + c + d,\ \alpha_i = \alpha_i^0 - c,\ \forall i,\\
&\qquad \gamma_j = \gamma_j^0 - d,\ \forall j,\ \sigma^2 = \sigma_0^2,\ c, d \in \mathbb{R}\}.
\end{aligned}
$$

Taking the definition of $\Theta/\sim$ into account, one may conclude that $((X\beta)'\ \sigma^2)'$ and $((X'X\beta)'\ \sigma^2)'$ are identifying functions whereas $X\beta$ and $X'X\beta$ are identifiable but not sufficient (they are so if σ^2 is known). The parameter ϑ and any bijective function of it are examples of (trivial) sufficient and unidentifi-

able functions. The function $(\sum_{i=1}^{a} \alpha_i, \sigma^2)'$ illustrates an unidentifiable function which is not sufficient.

The lack of identifiability caused by overparametrization of the freedom equation formulation of the model may be eliminated through reparametrization to the identifying function μ. Denoting by W' any $(q - r) \times q$ matrix of full rank such that $W'X = 0$ and by $\mathcal{N}(W')$ its null space, the known identity between $\mathcal{M}(X)$ and $\mathcal{N}(W')$ allows us to define alternatively the linear model as $y = \mu + u$ with $\mu \in \mathcal{N}(W')$.

The identifiability of the cell means restricted model in designs with at least one observation per cell may disappear in unbalanced designs with missing cells. As a matter of fact, once the restricted model is redefined in terms of freedom equations by reduction of the parametric dimensionality (see, e.g., Murray and Smith [1985]), its identifiability is not ensured because this is highly dependent on both the kind of restrictions and the number and location of the unobserved cells.

Example 2 (Log-linear model). In essence, the above discussion can be applied to the log-linear model for categorical data. Let y be the vector of counts in the q cells of a contingency table, with mean vector $\mu \in \mathbb{R}^q_+$. The ordinary log-linear model is defined by $ln\ \mu = X\beta$, where $\beta \in \mathbb{R}^p$ is a vector of unknown paramters and the matrix X, of rank assumed equal to $r < p \leq q$, is such that $\mathcal{M}(X)$ includes the vectors indicating the possible partition of the set Q of the q cells induced by the experimental design.

In the case where y represents the simultaneous realization of q independent Poisson processes, the probabilistic model is defined by the probability function

$$f(y|\mu) = \prod_{i \in Q} (y_i!)^{-1} \exp[y'\ ln\ \mu - 1'\mu],\ y \in \mathbb{N}_0^q,\ ln\ \mu \in \mathcal{M}(X). \qquad (2)$$

Suppose now that Q is partioned as $\{Q_k,\ k = 1, \ldots, s\}$ and let $C = (c_1, \cdots, c_s)$ be a $q \times s$ matrix such that c_k stands for the vector indicating the cells of Q_k, $n = 1, \ldots, s$, in a way that the experimental design is defined conditionally to $\{c_k' y = N_k,\ k = 1, \ldots, s\}$. Denoting by $\{y_k\}$ and $\{\mu_k\}$ the partitions of y and μ, it is known that the resulting probabilistic model is defined by

$$f(\{y_k\} \mid \{N_k, \mu_k\}) = \left(\prod_{k=1}^{s} \frac{N_k!}{\prod_{i \in Q_k} y_i!} \right) \exp\left(\sum_{k=1}^{s} y_k'\ ln\ \nu_k - \sum_{k=1}^{s} N_k\ ln\ N_k \right) \qquad (3)$$

where $\nu_k = (N_k|c_k'\mu)\mu_k$ is the conditional mean vector of y_k. Since

$$ln\ \nu \equiv ln(\nu_1', \cdots, \nu_s')' = ln\ \mu + \sum_{k=1}^{s} \left(ln\ \frac{N_k}{c_k'\mu}\right) c_k \tag{4}$$

and $\mathcal{M}(C) \subset \mathcal{M}(X)$ by assumption, the same log linear model for μ can be applied to ν. The only difference lies on the fact that the range of the $ln\ \nu$ is the set $\mathcal{M}(X)$ restricted by $c_k'\nu = N_k$, $k = 1, \ldots, s$.

In any case, the distribution of y depends on β through the function $X\beta$, which highlights the unidentifiability of the statistical model and the identifying nature of the functions $X\beta$ and $exp(X\beta)$, as well. This model constitutes a typical example of an identifiability problem which does not prevent some relevant inferences. In fact, many questions of interest can be alternatively expressed only in terms of the identifiable functions μ (under (2)) or ν (under (3)), which allow us to apply the maximum likelihood methodology without any trouble. Nevertheless, under the weighted least squares approach one has to deal with the same problems that arise in the Normal linear model analysis, when unrestricted to identifiable functions.

Example 3 (Models for discrete data under truncation or censoring). Even without introducing log-linear constraints, expression (3) has an underlying unidentifiability feature deleted by reparametrization from μ to ν. For simplicity, let us consider the case where $q = 2$ and $s = 1$. The parametric space $\mathcal{U}$, associated to model (2), defined by the first quadrant of the plane (μ_1, μ_2), is clearly identifiable. However, the same no longer occurs in model (3) characterized by the distribution $Bi(N, \mu_1/(\mu_1 + \mu_2))$, for y_1. It is easy to see that, for $\mu_0 = (\mu_1^0\ \mu_2^2) \in \mathcal{U}$,

$$[\mu_0] = \{\mu \in \mathcal{U} : \mu_1/\mu_2 = \mu_1^0/\mu_2^0\} \neq \{\mu_0\}. \tag{5}$$

The identifying partition $\mathcal{U}/\sim$ is the set of the half-lines of the first quadrant proceeding from the origin, and so, each observational equivalence class could be indexed by the slope of the respective half-line. One example of an identifying function is given by $\vartheta(\mu) = (\vartheta_1(\mu)\ \vartheta_2(\mu))'$, where $\vartheta_i(\mu) = \mu_i/(\mu_1 + \mu_2)$, $i = 1, 2$. The graphical representation of its range allows us to see how each class of $\mathcal{U}/\sim$ becomes singular. It is an identifying reparametrization of this kind that expression (3) embodies. Now it is easy to get other examples of identifiable functions, e.g., $E(y_1|N, \mu)$ and $var(y_1|N, \mu)$ of which only the firt one is sufficient (apply Definition 2.5 or Theorem 2.2).

A similar situation occurs in conditional multinomial models reflecting truncation or absence of reporting in certain categories. Consider, for instance, a 2×2 table of counts $y = (y_{ij}, i, j = 1,2)$ and let n be a vector repre-

senting the observable counts $\{y_{ij}, (i, j) \in S\}$, where $\sum_{(i,j)\in S} y_{ij} = N$. The resulting multinomial model for n, with probabilities $\{\vartheta_{ij} / \sum_{(i,j)\in S} \vartheta_{ij}\}$ is clearly unidentifiable. Let us now introduce restrictions on the parametric space $\Theta = \{(\vartheta_{11}, \vartheta_{12}, \vartheta_{21}) : \vartheta_{ij} > 0, \vartheta_{11} + \vartheta_{12} + \vartheta_{21} < 1\}$ defined by the independence structure in the full table. It is easy to check the identifiability of the marginal probabilities (and hence of $\vartheta = (\vartheta_{ij}, i, j = 1, 2)$), when $S = \{(1, 1), (1, 2), (2, 1)\}$. Nevertheless this is not true if $S = \{(1, 1), (2, 1)\}$. This result is a concrete indication, now in a set-up of categorical data, that the identifiability of the parameter indexing the complete data distribution, in problems with missing observations, depends on the number and location of these observations, as well as on the structural model at hand.

In some cases, certain cells are considered as having probabilities known beforehand and so they are excluded from the analysis. This situation fits in the above framework but now the sum of probabilities of the observed cells is known. As a consequence, the saturated model describing the so-called incomplete tables (with structural zeros) is identiable. However, under reduced structures (e.g., quasi-log-linear models) problems of unidentifiability can appear. Examples of this are given by the $2^2 \times 3$ tables, where $S = \{(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 3)\}$, and by the table 2×3^2, where $S = \{(1, 1, 1,), (1, 2, 2), (1, 3, 2), (2, 1, 2), (2, 2, 3), (2, 3, 3)\}$, both under quasi-independence structure.

An identical argument can be applied to situations where categorical data are censored either deterministically (by design) or randomly due to an uncontrolled deficient reporting process. The genetic problem of the ABO blood system is an example of the first case, where the unidentifiability of the saturated Multinomial model vanishes under, for instance, the Hardy-Weinberg equilibrium assumption. The second case, when every sample unit is only partially categorized, can be analysed similarly under appropriate assumptions concerning ignorability for the reporting process (see, e.g., Paulino [1991]). Example 8, described in Section 8, is a simple illustration of this case.

4. Parametric sufficiency versus sample sufficiency

The concept of parametric sufficiency has a close analogy to that of sample sufficiency. To show it in the context of a classical statistical model, let us consider shortly the Fisherian concept of sufficiency in specific situations of the dominated case.

Let $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be defined on the measurable sets of subset $\mathcal{Y}$ of a euclidean space and dominated by a σ-finite measure m, with Θ being also euclidean. Denote by $f(y|\vartheta)$ a density function of P_ϑ relative to m. Suppose that for every $\vartheta \in \Theta$, $f(y|\vartheta) > 0$, $\forall y \in \mathcal{Y}$ and, in the case of $\mathcal{Y}$ being uncountable, that $f(y|\vartheta)$ is continuous in $\mathcal{Y}$.

Let us define in $\mathcal{Y}$ the equivalence relation

$$y_1 * y_0 \Leftrightarrow f(y_1|\vartheta) = c(y_1, y_0)\, f(y_0|\vartheta), \quad \forall \vartheta \in \Theta \tag{6}$$

where $c(y_1, y_0) > 0$ is independent of ϑ. The class $[y_0] \in \mathcal{Y}/*$ is formed by the sample points whose likelihood functions, $l(\vartheta|y)$, are proportional to $l(\vartheta|y_0)$. Fix $\vartheta_0 \in \Theta$ and consider $ln[l(\vartheta|y)/l(\vartheta_0|y)]$ as a function of y indexed by ϑ, denoted by $T_\vartheta(y)$. It is easy to see that

$$[y_0] = \{y \in \mathcal{Y} : T_\vartheta(y) = T_\vartheta(y_0), \quad \forall \vartheta \in \Theta\}. \tag{7}$$

That is, the minimum sufficient partition $\mathcal{Y}/*$ (recall the Lehmann-Scheffé method) is induced by the family $\{T_\vartheta(\cdot) : \vartheta \in \Theta\}$. So, any sample partition finer than $\mathcal{Y}/*$ is sufficient. Let us call necessary any partition thicker (or coarser) than $\mathcal{Y}/*$.

Now returning to the parametric space and keeping the assumptions about $f(y|\vartheta)$ in mind, the elements of $\Theta/\sim$ can be defined by

$$[\vartheta_0] = \{\vartheta \in \Theta : f(y|\vartheta) = f(y|\vartheta_0),\ y \in \mathcal{Y}\}. \tag{8}$$

Fixing $y_0 \in \mathcal{Y}$ and defining $\psi_y(\vartheta) = ln[f(y|\vartheta)/f(y_0|\vartheta)]$ one has that

$$[\vartheta_0] = \{\vartheta \in \Theta : \psi_y(\vartheta) = \psi_y(\vartheta_0), \quad \forall y \in \mathcal{Y}\}. \tag{9}$$

Comparing the concepts of sufficiency and necessity for a sample partition and the expression (8) with the definitions 2.5 i), 2.4 i) and the expression (10), a close parallelism is displayed, which particularly helps to understand the reason of the designation parametric sufficiency. This analogy can be developed and made rigorous in the set-up of a Bayesian model (see Picci, 1977; Paulino, 1988). In the specific situation above, and using a prior density, $q(\vartheta)$, strictly positive in Θ, the definition of $[y_0]$ is equivalent to

$$[y_0] = \{y \in \mathcal{Y} : q(\vartheta|y) = q(\vartheta|y_0), \quad \forall \vartheta \in \Theta\} \tag{10}$$

where $q(\vartheta|y)$ stands for the posterior density. The comparison of (11) and (9) already shows the duality between the identifying partition, $\Theta/\sim$, and the

minimum sufficient partition, $\mathcal{Y}/*$, within the framework of a dominated Bayesian model.

5. Identifiability versus sample information

Inspection of identifiability of many parametrically complex models is not an easy and straightforward task. Often it is the confrontation of the researcher with anomalous results when one is making inferences that leads him/her to reanalyse at length the model structure.

On the other hand, most of the early results on the unidentifiability problem were related to specific models, which tended to hide the general nature of the problem, as properly pointed out by Koopmans and Reiersol (1950). It was the acknowledgement of this fact that led Rothenberg (1971) to attempting to derive general criteria of identifiability. Realizing that unidentifiability is consequence of the lack of enough information to discriminate among alternative parametric values in the model specification, Rothenberg approached the problem in terms of the Fisher information measure. Bowden (1973) generalized this approach through the Kullback-Leibler information measure which we shall discuss now.

Let ϑ_0 be a given value of the parameter $\vartheta \in \Theta$ indexing the family of distribution $F(y|\vartheta)$. The Kullback-Leibler information function for discriminating $F(y|\vartheta_0)$ against $F(y|\vartheta)$ per observation of $F(y|\vartheta_0)$ is the real-valued function defined on $\Theta \times \Theta$ by

$$I(\vartheta_0, \vartheta) = E\left[ln \frac{dF(y|\vartheta_0)}{dF(y|\vartheta)} \;\middle|\; \vartheta_0 \right] . \tag{11}$$

Its range is $[0, +\infty]$ (see Kullback, 1959, pag. 11) and takes only finite values if the measures P_ϑ and P_{ϑ_0} are mutually absolutely continuous (Wilks, 1962, p. 347). This condition, assumed henceforth and denoted by $P_\vartheta \approx P_{\vartheta_0}$, is satisfied for the dominated case models with support independent of ϑ.

Following the reasoning of Bowden (1973) it can be shown that

Theorem 5.1. Letting $\vartheta_1 \neq \vartheta_0$ be two points of Θ for which $P_{\vartheta_1} \approx P_{\vartheta_0}$, then $\vartheta_1 \in [\vartheta_0]$ iff $I(\vartheta_0, \vartheta_1) = 0$.

Relying on this result, Bowden gets as corollary the identifiability criterion:

Criterion I. Assuming that $P_\vartheta \approx P_{\vartheta_0}$, $\forall \vartheta \in \Theta$, ϑ_0 is globally (locally) identifiable iff ϑ_0 is the unique solution of the equation $I(\vartheta_0, \vartheta) = 0$ in Θ (in an open neighbourhood of ϑ_0).

Note that if this criterion is satisfied for an arbitrary $\vartheta_0 \in \Theta$, a necessary and sufficient condition for identifiability of Θ is obtained when the family of distributions is self-dominated by any of its members. This criterion, when applicable, has the advantage of putting together examination of both local and global identifiability and allowing us to define the identifying partition since $[\vartheta_0] = \{\vartheta \in \Theta : I(\vartheta_0, \vartheta) = 0\}$.

However, for many models it is not easy to determine all the solutions of the equations $I(\vartheta_0, \vartheta) = 0$ in a direct way. When $I(\vartheta_0, \vartheta)$ is a continuous function of ϑ whose domain is an open set of euclidean space, this problem can be solved through mathematical optimization methods. For illustration, let us consider the following example:

Example 4. Let us consider the distribution (2) under the log-linear model for the mean vector of the q cells. Letting $\mu \equiv \mu(\beta) = exp(X\beta)$ and denoting the image through $\mu(\cdot)$ of a given $\beta_0 \in \mathbb{R}^p$ by μ_0, one can check that

$$I(\beta_0, \beta) = \mu_0'[ln\ \mu_0 - ln\ \mu(\beta)] - 1_q'[\mu_0 - \mu(\beta)]. \tag{12}$$

The critical points, β^*, of $I \equiv I(\beta_0, \beta)$ satisfy the equation

$$\frac{\partial I}{\partial \beta}\bigg|_{\beta^*} = 0 \Leftrightarrow X'[\mu(\beta^*) - \mu_0] = 0 \tag{13}$$

as a consequence of $\dfrac{\partial I}{\partial \mu} = -D^{-1}_{\mu(\beta)}\,\mu_0 + 1_q$ and $\dfrac{\partial \mu}{\partial \beta'} = D_{\mu(\beta)}\,X$, where D_a denotes a diagonal matrix with the elements of the vector a along the main diagonal. In the usual log-linear models, $1_q \in \mathcal{M}(X)$ implying that $1_q'[\mu_0 - \mu(\beta^*)] = 0$. Thus

$$\begin{aligned}[\beta_0] &= \{\beta^* \in \mathbb{R}^p : \mu_0'[ln\ \mu_0 - ln\ \mu(\beta^*)] = 0\} \\ &= \{\beta^* \in \mathbb{R}^p : X\beta^* = X\beta_0\},\end{aligned} \tag{14}$$

showing that the domain of β is decomposed in classes defined by the $(p - r)$ dimensional linear manifolds $\{\beta_0 + \mathcal{N}(X)\}$ (see, e.g., Nering, 1964, p. 183-4). This is a result analogous to the one obtained for the Normal linear model. a formally identical result is obtained for the Multinomial model (3). □

One must emphasize that the absolute continuity assumption in criterion I excludes the nonregular cases where $\mathcal{Y}$ depends on ϑ. However, by analysing the proof of the Theorem 2.5, it follows that:

- Condition $I(\vartheta_0, \vartheta) = 0$ is always necessary to the observational equivalence of ϑ and ϑ_0;

– Condition $I(\vartheta_0, \vartheta) > 0$ is necessary for ϑ and ϑ_0 not being observationally equivalent, provided that $I(\vartheta_0, \vartheta)$ is finite.

For illustration of the applicability of these conclusions in the nonregular cases referred to, consider the following example:

Example 5. Let $\mathcal{Y}$ be the interval of $\mathbb{R}$, $L(\vartheta) = [\varphi_1(\vartheta), \varphi_2(\vartheta))$ where $\varphi_i(\vartheta)$, $i = 1, 2$, are continuous functions of $\vartheta \in \mathbb{R}$, with $\varphi_1(\vartheta) < \varphi_2(\vartheta)$, $\forall \vartheta \in \mathbb{R}$, and let $\mathcal{P}$ be the family of uniform distributions in $L(\vartheta)$.

From the expression of $F(y|\vartheta)$ it is easy to conclude that, for every $\vartheta_0 \in \mathbb{R}$, $[\vartheta_0] = \{\vartheta \in \mathbb{R} : \varphi_i(\vartheta), i = 1, 2\}$. So, the model is identifiable iff both ends of $L(\vartheta)$ are one-to-one functions. It is clear from (12) that $I(\vartheta_0, \vartheta) = 0$, $\forall \vartheta \in [\vartheta_0]$, as would be expected.

On the other hand, for every pair (ϑ_0, ϑ) such that $L(\vartheta_0)$ is not included in $L(\vartheta)$, $I(\vartheta_0, \vartheta) = \infty$. That is, a finite value of $I(\vartheta_0, \vartheta)$ implies that $L(\vartheta_0) \subset L(\vartheta)$, assuring that P_{ϑ_0} is dominated by P_ϑ, and in these circumstances

$$I(\vartheta_0, \vartheta) = \ln \frac{\varphi_2(\vartheta) - \varphi_1(\vartheta)}{\varphi_2(\vartheta_0) - \varphi_1(\vartheta_0)} .$$

That is what happens, e.g., when $\vartheta < \vartheta_0$ if $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ are increasing and decreasing functions, respectively. Hence, in these cases, $I(\vartheta_0, \vartheta) = 0$ implies that the lengths of $L(\vartheta)$ and $L(\vartheta_0)$ are equal, whereby $L(\vartheta) = L(\vartheta_0)$ (note that $L(\vartheta_0) \subset L(\vartheta)$). It is thus illustrated that, under finiteness of $I(\vartheta_0, \vartheta)$, $I(\vartheta_0, \vartheta) = 0$ implies $\vartheta \in [\vartheta_0]$. A concrete example is provided by $\varphi_2(\vartheta) = -\varphi_1(\vartheta) = \vartheta^2$, $\vartheta \in \mathbb{R}$, where $[\vartheta_0] = \{-\vartheta_0, \vartheta_0\}$, $\forall \vartheta_0 \in \mathbb{R}$.

The models referred to in Example 4 illustrate cases of continuity of $I(\vartheta_0, \cdot)$, which is satisfied under the usual regularity conditions on the family $\{F(y|\vartheta)\}$. In these cases, the gradient vector and the Hessian matrix of $I(\vartheta_0, \cdot)$ evaluated at ϑ_0 are respectively equal to the null vector and the Fisher information matrix per observation of $F(y|\vartheta_0)$, $\mathcal{I}(\vartheta_0)$. As a consequence, the positive definiteness of $\mathcal{I}(\vartheta_0)$ is a sufficient condition for existence of an unique local minimum of $I(\vartheta_0, \cdot)$.

On the other hand, a Taylor expansion about ϑ_0 yields

$$I(\vartheta_0, \vartheta) = \frac{1}{2} (\vartheta - \vartheta_0)' \mathcal{I}(\vartheta^*)(\vartheta - \vartheta_0), \quad \vartheta^* = t\vartheta_0 + (1 - t)\vartheta, \quad 0 < t < 1$$

so that $\mathcal{I}(\vartheta^*)$ is positive definite if $I(\vartheta_0, \vartheta) > 0$ in a neighbourhood of ϑ_0. If one assumes that the rank of $\mathcal{I}(\vartheta)$ does not change in an open neighbourhood of ϑ_0 (the Rothenberg's regulary condition of $\mathcal{I}(\vartheta)$ in ϑ_0), then one can

conclude that $\mathcal{I}(\vartheta_0)$ is positive definite. Taking criterion I at its local version into account, these conclusions set up the criterion of local identifiability of Rothenberg (1971), as follows:

Criterion II. Under the mentioned regularity conditions, nonsingularity of $\mathcal{I}(\vartheta_0)$ implies local identifiability of ϑ_0 which, in turn, implies the nonsingularity of $\mathcal{I}(\vartheta_0)$ if $\mathcal{I}(\cdot)$ is regular in an open neighbourhood of ϑ_0.

Thus the singularity of $\mathcal{I}(\vartheta_0)$, only by itself, does not necessarily imply the local unidentifiability of ϑ_0. This fact can be understood from Taylor series expansion of $I(\vartheta_0, \vartheta)$ at ϑ near ϑ_0,

$$I(\vartheta_0, \vartheta) = \frac{1}{2}(\vartheta - \vartheta_0)' \mathcal{I}(\vartheta_0)(\vartheta - \vartheta_0) + O(\|\vartheta - \vartheta_0\|^3).$$

The higher order terms can ensure that $I(\vartheta_0, \vartheta) > 0$ for every $\vartheta \neq \vartheta_0$ in the neighbourhood of ϑ_0, even though the quadratic form in the above expression be null.

Rothenberg (1971) still extends this criterion to the case of a constrained parametric space and proves that nonsingularity of $\mathcal{I}(\vartheta)$ is sufficient to global idenfitiability of the natural parameter of the multiparametric exponential family.

To finish this section, we illustrate the application of the above criteria with the first case discussed along Example 3.

Example 6. Taking the conditional distribution $Bi(N, \mu_1/(\mu_1 + \mu_2))$ into account it is easy to check that the information matrix

$$\mathcal{I}(\mu) = \frac{N}{\mu_1+\mu_2}\begin{pmatrix} \mu_1^{-1}-(\mu_1+\mu_2)^{-1} & -(\mu_1+\mu_2)^{-1} \\ -(\mu_1+\mu_2)^{-1} & \mu_2^{-1}-(\mu_1+\mu_2)^{-1} \end{pmatrix}$$

is singular. Then μ is not identifiable locally, and hence, globally, by Criterion II. Furthermore, the observational equivalence classes can be determined from the Criterion I. Denoting $\mu_0 = (\mu_1^0, \mu_2^0)'$ a value of μ, it is easy to verify that

$$I(\mu_0, \mu) = N \sum_{i=1}^{2} \frac{\mu_i^0}{\mu_1^0+\mu_2^0} \ln\left(\frac{\mu_i^0 \mu_1+\mu_2}{\mu_i \mu_1^0+\mu_2^0}\right).$$

This function is identically null iff $\mu_i/(\mu_1 + \mu_2) = \mu_i^0/(\mu_1^0 + \mu_2^0)$, $i = 1, 2$. The point μ_0 is not identifiable and the set of minimum points of $I(\mu_0, \cdot)$ coincides with the class $[\mu_0]$ defined by (5).

6. Identifiability versus estimability

In much of the statistical literature there is some lack of clarity in the undistinct use of the terms identifiability and estimability, even though this latter concept is clearly inferential, independently of the various meanings in which is used. The purpose of this section is to expound the relationship between these two concepts in general and in the context of the linear model where they are often used.

As mentioned in Section 2, in an unidentifiable model a sample realization only gives information on an observational equivalence class, not operating a further selection within it. The definition of $\Theta/\sim$ in terms of an identifying function $\psi = \psi(\vartheta)$, whose estimation is not problematic, clarifies this point by highlighting that the estimation of an unidentifiable parameter is merely arbitrary. The lack of uniqueness in the construction of estimators in this case is independent not only of the data set but also of the estimation method. This problem is distinct from that occurring in various identifiable models, where nonuniqueness follows from the lack of global concavity of the likelihood function for some (eventually all) data sets. One example of this is given by any random sample of the uniform model in $(\vartheta - 1/2, \vartheta + 1/2)$, $\vartheta \in \mathbb{R}$, when analysed by the maximum likelihood method. Another example, where this estimation method does not yield unique estimates for certain data sets, is given by the genetic problem referred to in Example 3 under the Hardy-Weinberg equilibrium structure (see, e.g., Haberman, 1974).

A further undesirable aspect of the unidentifiable models, with serious implications on the asymptotic theor of tests of hypotheses, is the inexistence of consistent estimators. It is sufficient to think that, whatever be the sample size, the sampling distribution of any estimator is the same for all observationally equivalent points. The lack of consistency of the maximum likelihood estimator may in some cases be seen in the light of an argument around the infimum points of $I(\vartheta_0, \vartheta)$ (see Silvey, 1975). Thus, the existence of consistent estimators is also a sufficient condition for identifiability, as is the estimation uniqueness mentioned above. Nevertheless, it must be added that the reciprocal proposition is not true (see, e.g., Zacks [1971, Sec. 5.3] and Gabrielsen [1978]).

Although, sometimes, the term estimability is used in the sense of uniqueness of estimation, in a way that is independent of the data, or of existence of consistent estimators, it is more frequently applied in the following sense:

Definition 6.1. A function $\varphi(\vartheta)$ is said (linearly) estimable if it admits $a(n)$ (linear) unbiased estimator.

It is evident that this estimability concept is also a sufficient condition for identifiability, i.e.

Theorem 6.1. If $\varphi(\vartheta)$ is an estimable function, then $\varphi(\vartheta)$ is identifiable.

This result shows that every estimable function (for instance, the population moments, when they exist) induces an identifiable partition, but not necessarily coincident with the identifying partition. This latter aspect could be satisfied under some requirements on the corresponding estimator (Tuncer, 1985).

In the analysis of linear models $y = X\beta + u$, defined as in Example 1, the interest is centered on linear functions of ϑ, namely σ^2 and functions $A\beta$, where A is a $s \times p$ matrix of rank $s \leqslant p$. It is widely known that a necessary and sufficient condition for linear estimability of $A\beta$ is the existence of a $s \times n$ matrix, B, such that $A = BX$. The equivalence between the concepts of identifiability and linear estimability for linear functions, proved by Reiersol (1963), could be demonstrated from Theorem 2.4, as follows:

Theorem 6.2. A function $A\beta$ is identifiable iff it is linerly estimable.

Proof. Since $((X\beta)'\ \sigma^2)'$ is an identifying function and $A\beta$ is a function independent of σ^2, Theorem 2.4 tells us that $A\beta$ is identifiable iff there exists a function h such that $A\beta = h(X\beta)$, $\forall\beta \in \mathbb{R}^p$. So, the implication linearly estimable $\Rightarrow$ identifiable follows.

Now, let us decompose $\mathbb{R}^p$ in the direct sum of $\mathcal{N}(X)$ with the subspace $\mathcal{G}$ generated by a given genealized inverse of X, i.e., $\mathbb{R}^p = \mathcal{N}(X) + \mathcal{G}$. Since the transformation X restricted to $\mathcal{G}$, $\mathcal{G} \xrightarrow{X} \mathcal{M}(X)$, is an isomorfism, denote by $T\alpha$, for every $\alpha \in \mathcal{M}(X)$, the element of $\mathcal{G}$ such that $\alpha = XT\alpha$. Then, by hypothesis, $h(\alpha) \equiv h(XT\alpha) = AT\alpha$, showing the linearity of the restriction of h to $\mathcal{M}(X)$. Hence, $A\beta \equiv h(X\beta) = ATX\beta$, $\forall\beta \in \mathbb{R}^p$.

The application of this result to a restricted linear model under some unobserved cells, as referred to in Example 1, shows that the model identifiability is equivalent to the general concept of connectedness of Murray and Smith (1985). This result still allows us to realize the reason for restricting the estimability concept for linear functions to that of linear estimability. In effect, from Theorems 6.1 and 6.2, one has that:

Theorem 6.3. A function $A\beta$ is identifiable iff it is estimable.

This theorem shows that there cannot be any estimable linear function that is not also linearly estimable. However, consideration of nonlinear functions

permits illustrating that identifiability does not imply estimability, in general. On the other hand, this theorem permits deriving identifiability conditions from results on estimability (see Richmond [1974]).

In some situations, the domain of β is constrained to a subset of $\mathbb{R}^p$, often defined by the subspace $\mathcal{N}(H)$, where H is a $t \times p$ matrix of rank $t < p$. According to the analysis of the unrestricted model, the model identifiability is equivalent to the existence of an unique solution to the system $X\beta = \bar{\mu}$, with $\beta \in \mathcal{N}(H)$, for each value of the mean $\bar{\mu}$ of the observations. When this does not occur, the class of points observationally equivalent to $\beta_0 \in \mathcal{N}(H)$ is defined by

$$[\beta_0] = \{\beta \in \mathcal{N}(H) : \beta = \beta_0 + \alpha, \alpha \in \mathcal{N}(X)\}$$

where the null space of X is a $(p - r)$-dimensional subspace of $\mathbb{R}^p$. Taking $X\beta_0 = \bar{\mu}_0$, $[\beta_0]$ is the set of solutions to the nonhomogeneous linear system, $G\beta = (\bar{\mu}_0' \, 0')'$, where $G = (X' \; H')'$. So, $[\beta_0]$ is, for a fixed $\bar{\mu}_0$, a linear manifold of dimension $p - rank(G)$ lesser than its dimension $(p - r)$ in the unrestricted model. This definition of $[\beta_0]$ shows clearly that it contains only β_0 iff rank $(G) = p$. In this set-up, the functions $A\beta$ are to be viewed as linear transformations of $\mathbb{R}^{p-t}$ (isomorphic to $\mathcal{N}(H)$) and this yields modifications on the linear estimability criterion. Reiersol (1963) establishes this new criterion as follows:

Theorem 6.4. Under the assumption $\beta \in \mathcal{N}(H)$, the function $A\beta$ is linearly estimable iff the row vectors of A, $\{a_i', i = 1, \ldots, s\}$, are such that $a_i \in \mathcal{M}(G')$.

Due to Theorem 6.1 on one hand, and Theorems 2.4 and 6.4 on the other hand, the results expressed by Theorems 6.2 and 6.3 hold formally for the restricted model, that is

Theorem 6.5. Under the assumption $\beta \in \mathcal{N}(H)$, the function $A\beta$ is identifiable iff it is estimable.

Naturally, every function $A\beta$ identifiable in the unrestricted model is also identifiable in the restricted model. The reciprocal is not, however, true. The function itself which results from $A\beta$, assumed to be identifiable in $\mathbb{R}^{p-t}$, by embodying the constraints $H\beta = 0$, may no longer be identifiable in the unrestricted model. In effect, since the identifiable function $A\beta$, when $\beta \in \mathcal{N}(H)$, is linearly estimable, let By be a linear unbiased estimator of it. Then, $(A - BX)\beta = 0$, $\forall\beta \in \mathcal{N}(H)$, meaning that the rows of $A - BX$, viewed as column vectors, belong to the orthocomplement of $\mathcal{N}(H)$, which is $\mathcal{M}(H')$.

Thus, there exists a $s \times t$ matrix C such that $A - BX = CH$, whereby $A\beta = BX\beta + CH\beta$. In this way, if $H\beta$ is not identifiable, $A\beta$ may not be identifiable in $\mathbb{R}^p$. As an illustration, consider the ANOVA II model referred to in Example 1 and the identifiable function $\varphi(\beta) = \eta + \alpha_l + \frac{1}{b}\sum_{j=1}^{b} \gamma_j$. On imposing the constraint $\sum_{j=1}^{b} \gamma_j = 0$, defined by an unidentifiable function, $\varphi(\beta)$ is converted to $\eta + \alpha_l$. This new function is identifiable in the restricted model but not in the unrestricted model.

7. The process of the model identification

In classical inferences the unidentifiability problems require a reformulation of the model so as to get its identification. In this section we intend to describe how the model can be identified, analysing in detail how the process of imposing exact restrictions can achieve that purpose. This question is crucial for a whole understanding that seemingly different analyses of, namely, the liner model with no full rank are actually equivalent.

Given the existence of identifying functions, the model reparametrization to a function of this kind establishes a way of getting identification of the model. The restricted model of the cell means referred to in Example 1 illustrates this form of identification. An identical procedure may be applied to the log linear model for the cell probabilities of a Multinomial table.

That identifying reparametrization appears to be quite convenient when the inferential purpose is centered on an identifying function or on functions of it. However, the identification process most applied in a classical context consists of imposing constraints on the parametric space so that each observational equivalence class becomes a singleton.

In order to clarify the conditions on the constrains required by the model identification, let $\psi(\vartheta)$ be an identifying function so that $[\vartheta_0] = \{\vartheta \in \Theta : \psi(\vartheta) = \psi_0\}$ where $\psi_0 = \psi(\vartheta_0)$. Let us denote the restricted parametric space by $\bar{\Theta} = \{\vartheta \in \Theta : \lambda(\vartheta) = 0\}$. As referred to in Section 2, $\bar{\Theta}$ is identifiable iff each $[\vartheta_0]$ is made singular by $\lambda(\vartheta) = 0$. It is easy to see at once that this goal cannot be attained with an identifiable function $\lambda(\vartheta)$. In fact, in this case there will exist classes $[\vartheta_0]$ incompatible with the value 0 for $\lambda(\vartheta)$ and so, the system of equations $\psi(\vartheta) = \psi_0$, $\lambda(\vartheta) = 0$ will not be consistent for all classes of $\Theta/\sim$. The compatible classes, in number of at least one (there will be only one if $\lambda(\vartheta)$ is further sufficient), will not have their dimensionality reduced and so the unidentifiability problem will remain in $\bar{\Theta}$.

The identification process of Θ based on imposing the consistent restrictions $\lambda(\vartheta) \equiv (\lambda_1(0), \cdots, \lambda_t(\vartheta))' = (0, \cdots, 0)'$, with $\{\lambda_i(\vartheta)\}$ unidentifiable, can be

elucidatively viewed in sequential terms. Suppose, with no loss of generality, that each equivalence class is represented by a point of $\bar{\Theta}$ and denote by Θ_j the parametric space restricted by $\lambda_i(\vartheta)$, $i = 1, \ldots, j$ ($\Theta_t \equiv \bar{\Theta}$). Imposing first $\lambda_1(\vartheta) = 0$ causes that in each $[\vartheta_0] \in \Theta/\sim$ there are only some points compatible with such a restriction. This means that the dimensionality of $[\vartheta_0]_1 = [\vartheta_0] \cap \Theta_1$ is less than that of $[\vartheta_0]$. Now imposing $\lambda_2(\vartheta) = 0$, $[\vartheta_0]_2$ is obtained from $[\vartheta_0]$ by removing other points, and so on. If the dimensionality of each $[\vartheta_0]_j \in \Theta/\sim$ is greater than one, a further restriction, $\lambda_{j+1}(\vartheta) = 0$, is imposed to reduce their dimensionality. This sequential process stops when each element of $\Theta/\sim$ is a singleton. If this occurs after adding $\lambda_t(\vartheta) = 0$, then $\Theta_t/\sim = \{\{\vartheta\} : \vartheta \in \Theta_t\}$ and the desired identifiability is achieved.

In many situations, the parts of $\Theta/\sim$ and the restrictions are defined by linear functions, say $C\vartheta$ and $\Lambda\vartheta$ respectively. The mathematical results on linear systems then allow us to define necessary and sufficient conditions for uniqueness of the solution to the equations $C\vartheta = \psi_0 \equiv C\vartheta_0$, $\Lambda\vartheta = 0$.

By the results in Section 6, $\Lambda\vartheta$ is unidentifiable iff the rows of Λ do not belong to $\mathcal{M}(C')$. That is, iff no non-null vector of $\mathcal{M}(\Lambda')$ is vector of $\mathcal{M}(C')$, i.e., $\mathcal{M}(\Lambda') \cap \mathcal{M}(C') = \{0\}$. This condition is necessary and sufficient for consistency of the linear system quoted above, $\forall \psi_0 \in \mathcal{M}(C)$ – see, e.g., the algebraic result contained in Scheffé (1959, Sect. 1.4). Thus, the condition that the rows of Λ and C are mutually linearly independent is not more than the unidentifiability condition of $\Lambda\vartheta$, as emphasized by Reiersol (1963).

Once its consistency is ensured, the system will have a unique solution if the rank of $(C', \Lambda')'$ is equal to the dimensionality of ϑ, say p. Assuming that C has rank $r < p$, which implies that it is enough to define $t = p - r$ restrictions, the system will be consistent and will have a unique solution iff $\Lambda\vartheta$ is unidentifiable and Λ is of full rank. In this case, ϑ is expressed through $\psi = C\vartheta$ by $\vartheta = (C'C + \Lambda'\Lambda)^{-1}C'\psi$, showing the identifiability of $\bar{\Theta}$.

For illustrative purposes, let us examine the following example:

Example 7. Consider tha ANOVA model with two factors without interaction for a situation where there is one observation in each of the $2 \times 2 = 4$ cells. The model is specified by a 4×5 matrix of rank 3, $X = (x_1 \cdots x_5)$, where $x_1 = (1\ 1\ 1\ 1)'$, $x_2 = (1\ 1\ 0\ 0)'$, $x_3 = (0\ 0\ 1\ 1)'$, $x_4 = (1\ 0\ 1\ 0)'$ and $x_5 = (0\ 1\ 0\ 1)'$. The parameter vector is $\beta = (\eta\ \alpha_1\ \alpha_2\ \gamma_1\ \gamma_2)' \in \mathbb{R}^5$. As seen in Example 1, for a given $\beta_0 = (\eta^0\ \alpha_1^0\ \alpha_2^0\ \gamma_1^0\ \gamma_2^0)' \in \mathbb{R}^5$, the class $[\beta_0]$ is a bidimensional linear manifold defined by

$$\begin{aligned}[\beta_0] &= \{\beta \in \mathbb{R}^5 : X\beta = X\beta_0\} \\ &= \{\beta \in \mathbb{R}^5 : \eta = \eta^0 + c + d,\ \alpha_i = \alpha_i^0 - c, \\ &\qquad \gamma_j = \gamma_j^0 - d;\ i, j = 1, 2,\ c, d \in \mathbb{R}\}.\end{aligned}$$

Impose the restriction $\lambda_1'\beta = 0$, with $\lambda_1' = (0\ 1\ 1\ 0\ 0)$, which is suitable for the purpose since $\lambda_1 \notin \mathcal{M}(X')$ (Theorem 2.6). The new classes, $[\beta_0]_1 = [\beta_0] \cap \mathcal{N}(\lambda_1')$ define each a unidimensional linear manifold since $G_1 = (X', \lambda_1)'$ has rank 4. This operation removes the points of each $[\beta_0]$ regarding to $c \neq 0$ and so, for $\beta_0 \in \mathcal{N}(\lambda_1')$

$$[\beta_0]_1 = \{\beta \in \mathbb{R}^5 : G_1\beta = G_1\beta_0\}$$
$$= \{\beta \in \mathbb{R}^5 : \eta = \eta^0 + d,\ \alpha_i = \alpha_i^0,\ \gamma^0{}_j - d;\ i, j = 1.2;\ d \in \mathbb{R}\}.$$

To achieve the identification of the model, it suffices to add an extra restriction, necessarily linearly independent of the proceeding one (otherwise, $\mathcal{N}(\lambda_1'(/\sim$ would not change). Note that the restriction $\lambda_3'\beta = 0$, where $\lambda_3' = (0\ 0\ 0\ 1\ -\ 1)$, for instance, is unsuitable because the fact that $\lambda_3 = G_1'\delta$, with $\delta = (1\ -\ 1\ 0\ 0\ 0)'$, shows that $\lambda_3'\beta$ is identifiable (Theorems 6.4 and 6.5). Adding, instead, $\lambda_2'\beta = 0$, where $\lambda_2' = (0\ 0\ 0\ 1\ 1)$ is already fruitful. As a matter of fact, $\lambda_2 \notin \mathcal{M}(G_1')$ and its unidentifiability removes from each $[\beta_0]_1$ the points corresponding to $d \neq 0$. Thus, denoting $G_2 = (G_1', \lambda_2)'$

$$[\beta_0]_2 = [\beta_0]_1 \cap \mathcal{N}(\lambda_2') = \{\beta \in \mathbb{R}^5 : G_2\beta = G_2\beta_0\} = \{\beta_0\}.$$

Naturally, the model identification above does not require those two linearly independent restrictions corresponding to $\Sigma_i\, \alpha i = \Sigma_j\, \gamma_j = 0$. Another set of restrictions is $\alpha_a = \gamma_b = 0$ defining the so-called reference cell parametrization.

Once an identifiable model is obtained, the arbitrariness in estimation obviously vanishes. At this point, it is important to stress that the classical inferences about functions $A\beta$, identifiable in the unrestricted normal linear model, do not change when working with the identified model, as can be drawn from Reiersol's (1963) paper. The discussion in Searle (1971, Ch. 5, Sec. 6), having Theorems 6.4 and 6.5 in mind, is quite elucidative in this respect.

In this way, it is evident the double role that the unidentifiable functions play, that of ensuring the construction of an identifiable model and that of not «disturbing» the inferences about the identifiable functions. These aspects facilitate the understanding of the two usual means of classical analysis of the Normal linear model under the freedom equations formulations. One of them is the previous conversion to an identifiable model. The other is the use of an estimate of the original parameters, obtained from one generalized inverse of $X'X$ or from imposition of non-identifiable restrictions enough for a unique solution to exist to the normal equations sytem.

8. Unidentifiability and the classical and Bayesian approaches

As we have seen earlier, classical estimation of an unidentifiable parameter is merely arbitrary. Moreover, certain asymptotic optimality properties of any estimator like consistency are not valid anymore. These aspects make asymptotic tests of hypotheses unfeasible due to problems related to the definition of the test statistics (observed values dependent on the estimate used and possible singularity of the Fisher information matrix) and to the determination of their asymptotic distribution. We must note that these problems may not disappear entirely when the model is identifiable under the null hypothesis. Think, for instance, of the Wald test when the constrained parametric space, $\bar{\Theta} = \{\vartheta \in \Theta : \lambda(\vartheta) \equiv (\lambda_1(\vartheta), \cdots, \lambda_t(\vartheta))' = (0, \cdots, 0)'\}$ is identifiable. The nonuniqueness of the unrestricted maximum likelihood estimator in addition to the necessary unidentifiability of $\lambda(\vartheta)$ imply that the Wald statistic is not well defined.

That kind of problems, from a classical point of view, justifies attempts to eliminating non-identifiability, generally based upon imposing exact restrictions. In the setting quoted above, if only $r < t$ restrictions $\lambda_i(\vartheta) = 0$ are enough to ensure identifiability, then these restrictions are sometimes used to define the new parametric space, under which the $t - r$ remaining restrictions are to be tested. The original testing problem is disregarded and replaced by a conditional test of the $t - r$ restrictions given the r identifying restrictions. The asymptotic tests of Silvey (1959) and Aitchison and Silvey (1960), for instance, are set in this context.

The need for identification of the original model to make inferences feasible may lead to a criticizable practice of adopting restrictions of an arbitrary nature. To paraphrase Koopmans and Reiersol (1950), «Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested».

Rather, one can take advantage of the existence of identifiable parametric functions to increase our knowledge about unknown aspects of the model. Even when the restrictions themselves come from an analysis based on theoretical considerations and/or earlier observations, the categorical and sharp type of prior information embodied in the exact restrictions raises further difficulties given the little precise nature of this type of information, in general.

That kind of difficulties will be mitigated with a stochastic specification of all or part of the prior information, given its more flexible (and so, more justifiable) nature. Note that this is not only admitted but also required by

the Bayesian inferential process, whether or not the model is identifiable. The Bayesian conjunction of stochastic prior information with sample information implies that the posterior distribution is proper whenever the prior distribution is, independently of the model identifiability. In this case, existence of posterior moments, in particular, does not call for identifiability. It is certainly this that is underlying to the concise phrase of Lindley (1971, p. 46), «In passing, it might be noted that unidentifiability causes no real difficulty in the Bayesian approach». In our opinion, what happens is a displacement of the difficulties emerging from the lack of identifiability, which become related to the necessity of defining a proper prior distribution more carefully, since some parametric functions are nor updated by data, as will be seen afterwards.

In spite of the distinct notions of identifiability which can be found in the Bayesian literature (Zellner [1971], Drèze [1972], Leamer [1978], Florens et al. [1990]), in the dominant sense, the notions of Bayesian identifiability and sufficiency are essentially identical to those defined in Section 2 (Kadane [1974], Drèze [1974], Picci [1977], Deistler and Seifert [1978]).

With that concept of sufficiency, the characterization set in Theorem 2.2 is equivalent to saying that within each equivalence class, specified by the identifying function $\psi(\vartheta)$, the posterior probability measure is essentially equal to the prior probability measure. By the duality between the concepts of parametric and sample sufficiency, one can prove (Picci [1977], Paulino [1988]) that, in the usual dominated Bayesian models, sufficiency of $\psi(\vartheta)$ is characterized by the factoring of the posterior density of ϑ into two factors, one dependent of y and of ϑ through $\psi(\vartheta)$ and the other independent of y. These results show clearly that, once the distribution of $\psi(\vartheta)$ is updated, the role of the data is exhausted. This means that any other function $\gamma(\vartheta)$, one may define in order to establish a one-to-one correspondence between (ψ, γ) and ϑ, is not distributionally modified. For illustration consider the folloing example:

Example 8. Suppose that a random sample of size N is drawn from a population partitioned in four categories with (positive) probabilities ϑ_i, $1 \leq i \leq 4$, $\Sigma_i \vartheta_i = 1$. Suppose further that, rather than the usual outcome of counts $\{x_i\}$ for all categories, the observed reports only indicated the number of elements classified or not in the categories 1 and 4 together, that is, the value of $y = x_1 + x_4$ (and, of course, $N - y = x_2 + x_3$). Let us assume that the reporting process is modelled in such a way that the relevant model for inferences on $\{\vartheta_i\}$ is defined by the family of distribution $B_i(N, \vartheta_1 + \vartheta_4)$ (see Paulino [1991]).

Assume further that the accumulated prior information about $\{\vartheta_i\}$ can be

quantified by a Dirichlet distribution with parameter $b = (b_1\, b_2\, b_3\, b_4)' \in \mathbb{R}^4_+$. It is easy to check that the posterior density of ϑ can be expressed as

$$q_y(\vartheta) = g(\psi(\vartheta);\, y)\, h(\vartheta) \tag{15}$$

where $\psi(\vartheta) = \vartheta_1 + \vartheta_4$,

$$g(\psi(\vartheta);\, y) = [\psi(\vartheta)]^{a_1-1}[1 - \psi(\vartheta)]^{a_2-1}/B(a_1,\, a_2)$$

with $a_1 = b_1 + b_4 + y$ and $a_2 = b_2 + b_3 + N - y$, and

$$h(\vartheta) = \frac{\left(\frac{\vartheta_1}{\vartheta_1+\vartheta_4}\right)^{b_4-1}\left(\frac{\vartheta_4}{\vartheta_1+\vartheta_4}\right)^{b_4-1}}{B(b_1,b_4)}\ \frac{1}{\vartheta_1} \times \frac{\left(\frac{\vartheta_2}{\vartheta_2+\vartheta_3}\right)^{b_2-1}\left(\frac{\vartheta_3}{\vartheta_2+\vartheta_3}\right)^{b_3-1}}{B(b_2,b_3)}\ \frac{1}{\vartheta_2}.$$

This factorization of $q_y(\vartheta)$ confirms the identifying feature of $\psi = \vartheta_1 + \vartheta_4$ and the lack of sufficiency of the non-identifiable functions $\varphi_1 = \vartheta_1/(\vartheta_1 + \vartheta_4)$ and $\varphi_2 = \vartheta_2/(\vartheta_2 + \vartheta_3)$.

Considering the transformation $(\vartheta_1,\, \vartheta_2,\, \vartheta_4) \to (\vartheta_1,\, \vartheta_2,\, \psi)$ one can see that $g(\psi;\, y)$ and $h(\vartheta)$ stand for the posterior distribution of ψ and the conditional posterior distribution of $(\vartheta_1,\, \vartheta_2)$ given ψ, respectively. This latter factor reveals that, given ψ and y, the functions φ_1 and φ_2 are independently distributed as $B(b_1,\, b_4)$ and $B(b_2,\, b_3)$, respectively, thus coinciding with the corresponding prior distribution. Hence, the (degenerate) conditional prior distribution of ϑ given ψ cannot be updated by y.

Instead using directly the posterior distribution of ϑ, the posterior moments of ϑ, for instance, can be computed from the relationship $E_\vartheta[\Pi_i\, \vartheta_i^{r_i}|y] = E_\psi[E_\vartheta[\Pi_i\, \vartheta_i^{r_i}]|y,\, \psi]$, by considering

$$\prod_i \vartheta_i^{r_i} = \psi^{r_1+r_4}(1 - \psi)^{r_2+r_3} \prod_{i=1,4}\left(\frac{\vartheta_i}{\psi}\right)^{r_i} \prod_{i=2,3}\left(\frac{\vartheta_i}{1-\psi}\right)^{r_i}.$$

So, it follows that

$$\mathrm{E}_\vartheta\left[\prod_i \vartheta_i^{r_i}|y\right] = \frac{B(a_1+r_1+r_4,a_2+r_2+r_3)}{B(a_1,a_2)}\ \frac{B(b_1+r_1,b_4+r_4)}{B(b_1,b_4)}\ \frac{B(b_2+r_2,b_3+r_3)}{B(b_2,b_3)}$$

This procedure examplifies how Bayesian analyses can be made on the space of an identifying function as expressed by Kadane (1974). Furthermore, it shows how the specification of a proper prior distribution unblocks inferences with no necessity of imposing more rigid a priori structures like in the

classical approach. It is a procedure similar to this which underlies the Bayesian solution of the incomplete categorical data problem developed by Paulino and Pereira (1992, 1994), without resorting to the assumption of a noninformative reporting process.

The discussion made in this section shows that the inferential relevance of the identifiability present distinct degrees in the classical and Bayesian approaches. The latter presents a greater flexibility and naturality in making inferences about unidentifiable parameters practicable, though the lack of identifiability cannot be minimized.

In the way of conclusion, what seems important to emphasize is the judiciousness and the honesty which should guide the statistician when faced with a model suffering from unidentifiability. He/she must resist any unreasonable tendencies to discard the model or deform it through arbitrary esthetic operations. Even though many typical characteristics of the model are not «accessible», to analyse what is analysable is a more credible alternative. And from this viewpoint, choosing Bayesian ways could lead us farther as a consequence of less sharp obstacles caused by lack of identifiability.

Acknowledgements

The author gratefully acknowledge the support that was given by FAPESP, Brazil and JNICT, Portugal. In addition, we wish to express thanks to reviewers whose comments have improved the readability of this article.

REFERENCES

AITCHISON, J. and SILVEY, S. D. (1960). Maximum likelihood estimation procedures and tests of significance. *J. Royal Statist. Soc.*, B, 31, 154-171.

BARANKIN, E. W. (1960). Sufficient parameters: solution of the minimal dimensionality problem. *Ann. Inst. Statist. Math.*, 12, 91-118.

BASU, A. P. (1983). Identifiability. In *Encyclopedia of Statistical Sciences 4*, p. 2-6, N. L. Johnson and S. Kotz (eds.), Wiley, New York.

BOWDEN, R. (1973). The theory of parametric identification. *Econometrica*, 41, 1069-1074.

DEISTLER, M. and SEIFERT, H. G. (1978). Identifiability and consistent estimability in econometric models. *Econometrica*, 46, 969-980.

DRÈZE, J. (1972). Econometrics and decision theory. *Econometrica*, 40, 1-17.

DRÈZE, J. (1974). Bayesian theory of identification in simultaneous equations models. In *Studies in Bayesian Econometrics and Statistics*, p. 159-174, S. E. Fienberg and A. Zellner (eds.), North-Holland, Amsterdam.

EVERITT, B. S. (1985). Mixture distributions. In *Encyclopedia of Statistical Sciences 5*, p. 559-569, N. L. Johnson and S. Kotz (eds.), Wiley, New York.

FISHER, F. M. (1966). *The Identification Problem in Econometrics*. McGraw-Hill, New York.

FLORENS, J. P., MOUCHART, M. and ROLIN, J. M. (1990). *Elements of Bayesian Statistics*. Marcel Dekker, Inc., New York.

GABRIELSEN, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8, 261-263.

HABERMAN, S. J. (1974). Log-linear models for frequency tables derived by indirect observation: maximum likelihood equations. *Ann. Statist.*, 5, 911-924.

HANNAN, E. J. (1871). The identification problem for multiple equation systems with moving average errors. *Econometrica*, 39, 751-765.

HSIAO, C. (1983). Identification. In *Handbook of Econometrics, vol. I*, p. 223-283, Z. Griliches and M. D. Intriligator (eds.), North-Holland Publ. Co., Amsterdam.

KADANE, J. B. (1974). The role of identification in Bayesian theory. In *Studies in Bayesian Econometrics and Statistics*, p. 175-191, S. E. Fienberg and A. Zellner (eds.), North-Holland, Amsterdam.

KOOPMANS, T. C. and REIERSOL, O. (1950). The identification of structural characteristics. *Ann. Math. Statist.*, 21, 165-181.

KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.

LEAMER, E. E. (1978). *Specification Searches*. John Wiley & Sons, New York.

LINDLEY, D. V. (1971). *Bayesian Statistics: A Review*. SIAM, Philadelfia.

MURRAY, L. W. and SMITH, D. W. (1985). Estimability, testability and connectedness in the cell means model. *Comm. Statist. - Th. Meth.*, 14, 1889-1917.

NERING, E. D. (1964). *Linear Algebra and Matrix Theory*. John Wiley & Sons, Inc., New York.

PAULINO, C. D. M. (1988). *Analysis of Incomplete Categorical Data: Foundations, Methods and Applications*. Doctoral Thesis (in portuguese). University of São Paulo, Brasil.

PAULINO, C. D. M. (1991). Analysis of incomplete categorical data: a survey of the conditional maximum likelihood and weighted least squares approaches. *Brazilian Journal of Probability and Statistics*, 5(1), 1-42.

PAULINO, C. D. M. and PEREIRA, C. A. B. (1992). Bayesian analysis of categorical data informatively censored. *Comm. Statist. - Th. Meth.*, 21(9), 2689-2705.

PAULINO, C. D. M. and PEREIRA, C. A. B. (1994). Bayesian methods for categorical data under informative general censoring. Accepted for publication in Biometrika.

PICCI, G. (1977). Some connections between the theory of sufficient statistics and the identifiability problem. *SIAM J. Appl. Math.*, 33, 383-398.

REIERSOL, O. (1963). Identifiability, estimability, pheno-restricting specifications and zero Lagrange multipliers in the analysis of variance. *Skandinavisk Aktuarietids-krift*, 131-142.

RICHMOND, J. (1974). Identifiability in linear models. *Econometrica*, 42, 731-736.

ROTHENBERG, T. (1971). Identification in parametric models. *Econometrica*, 39, 576-591.

SEARLE, S. R. (1971). *Linear Models*. John Wiley & Sons, New York.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.*, 30, 389-407.

SILVEY, S. D. (1975). *Statistical Inference*. Chapman and Hall, London.

TUNCER, Y. (1985). Equivalence of parametric identifiability and estimability. *Comm. Statist. - Th. Meth.*, 14, 393-400.