Chilean Journal of Statistics
Vol. 1, No. 2, September 2010, 69–91

Special Issue "In Memory of Pilar Loreto Iglesias Zuazola"
Research Paper

Bayesian identifiability: Contributions to an inconclusive debate

Ernesto San Martín[1,2,*] and Jorge González[2]

[1]Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Measurement Center MIDE UC, Pontificia Universidad Católica de Chile, Santiago, Chile

(Received: 12 April 2010 · Accepted in final form: 01 September 2010)

Abstract

Using the concept of reduction by sufficiency of a Bayesian model, the issue of Bayesian identifiability is discussed. Various statements given in the literature on Bayesian identifiability are revised. Particular attention is put on the possibility of updating unidentified parameters. This issue is discussed under a general framework and also carefully illustrated in a fully discrete Bayesian model.

Keywords: Minimal Sufficient Parameter · Statistical Model · Sufficient Parameter · Unidentified Parameter · Updating Process.

Mathematics Subject Classification: Primary 62B05 · Secondary 62B15.

1. Introduction

Identifiability is often treated as a necessary condition for a rigorously well specified statistical model. When structural modeling is being considered, that is, when the model formalizes a certain phenomenon, the identification condition is more than a simple technical assumption but covers a more fundamental aspect, namely, the adequacy of a theoretical statistical model for an observed process.

In the sampling approach, a statistical model is defined as an indexed family of distributions on the sample space, whereas the Bayesian approach considers a unique probability measure on the product space "parameters × observations"; see Caillot and Martin (1972), Martín et al. (1973), Florens et al. (1990) and Gourieroux and Monfort (1995). The sampling theory framework of identifiability has been extensively discussed in the literature; see, e.g., Prakasa Rao (1992), Manski (1995), Manski (2007) and references therein. Under

Personal account: Pilar was an active supporter and developer of Bayesian statistics. She knew the debate whether identifiability is of interest or not in a Bayesian approach. The first author remembers interesting discussions with Pilar about this debate. This paper represents a response to our late discussions. It is couched following Florens, Mouchart and Rolin's Elements of Bayesian Statistics (1990) perspective, which in turn develops the concept of reducing information formalized by Basu. It seems appropriate on this occasion to mention that Pilar's advisor was Carlos Pereira (Universidade de Sao Paulo, Brazil). Motivated by certain works of Michel Mouchart and Jean-Marie Rolin, Pereira wrote his Ph. D. thesis under the supervision of Basu. Basu met several times with Michel; and Michel was my advisor. Their contributions are the basis of this paper that we write in memory of Pilar.

*Corresponding author. E. San Martín. Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile. Email: esanmart@mat.puc.cl

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

the Bayesian approach, however, the concept of identifiability and, particularly, nonidentifiability (Dawid, 1979) has not been free from controversies, polemics and confusion. The famous sentence "In passing it might be noted that unidentifiability causes no real difficulty in the Bayesian approach" quoted by Lindley (1971) is a remarkable example.

Depending on whether the problem is in the prior, the likelihood or the posterior distribution, different views on the issue of identifiability have been given in the literature. Poirier (1998) argued that "A Bayesian analysis of a nonidentified model is always possible if a proper prior on all the parameters is specified". In the same line, Eberly and Carlin (2000) pointed out that "in some sense identifiability is a non-issue for Bayesian analyses, since given proper prior distributions the corresponding posteriors must be proper as well, hence every parameter can be well estimated". Gelfand and Sahu (1999) stated that too informative priors will dominate the inference and priors near to be improper will produce ill-behaved posteriors, yet on the other hand they argued that nonidentifiability would not depend on the nature of the prior specification but on lack of identifiability in the likelihood. In this sense, Dawid's (classical) definition of Bayesian nonidentifiability is equivalent with what a Bayesian would call likelihood identifiability; see Eberly and Carlin (2000, p. 2280) and Ghosh et al. (2000). An unified thought is given by Kadane (1975) who stated that "identification is a property of the likelihood function and is the same whether considered classically or from the Bayesian approach." The issue has also been discussed when simulation-based techniques are used for model fitting and inferences. It is commonly argued that nonidentifiability does not preclude Bayesian inference as long as a suitable informative prior is specified. Kass et al. (1998) pointed out that provided the posterior is proper, there is no problem for MCMC methods, assuming that one has determined that the nonidentifiability "isn't due to a bug".

In addition to these views, there exist a perspective in which not only the advantage of the Bayesian approach over the sampling approach in making inferences on unidentified parameters is emphasized (Neath and Samaniego, 1997), but also in this perspective it is acknowledged that "the lack of identifiability cannot be minimized" (Paulino, 1994, p.149), specially in the prior elicitation process; see Paulino et al. (2003), Swartz et al. (2004), Jiang and Dickey (2008) and Poleto et al. (2010).

The above-mentioned perspectives motivate the following questions: (i) How can a genuinely Bayesian identification concept be defined? (ii) To what extent is the unidentified parameter updated by the data? (iii) In which sense a precise prior elicitation ensures a correct Bayesian inference on unidentified parameters? This paper develops a conceptual discussion leading to some answers to these questions.

The paper is organized as follows. In Section 2, we introduce the Bayesian model together with the definitions of some basic concepts necessary to define Bayesian identifiability. In Section 3, we explain the construction of a fully discrete Bayesian model and characterize the Bayesian identifiability in this context. Also, in this section, we discuss the updating process of unidentified parameters. Finally, we conclude the paper with a discussion.

2. Bayesian Model

2.1 General construction of a Bayesian model

In a sampling theory framework, a statistical model is formally defined as

$$\mathcal{E} = \{(S, \mathcal{S}),\ P^a\colon\ a \in A\}, \tag{2.1}$$

where $(S, \mathcal{S})$ is a measurable space, the sample space and $\{P^a\colon a \in A\}$ is a family of probability measures on the sample space indexed by a parameter a belonging to a pa-

rameter space A; see Fisher (1922), Basu (1975), Barra (1981) and McCullagh (2002). The probabilities $\{P^a\colon a \in A\}$ are called sampling probabilities as they describe the sampling or data generating process. The parameter space A might be either a Euclidean space, a functional space, or a product of both as it is the case in parametric, non-parametric and semi-parametric models, respectively. Note that the statistical model given in (2.1) can be considered as an extension of a probability space $(S, \mathcal{S}, P)$ in the sense that a unique probability measure P is replaced by a family of probability measures P^a, $a \in A$.

A Bayesian model is defined as a unique probability measure Q on the product space "parameters $\times$ observations", denoted as $A \times S$. Taking as a starting point the statistical model given in (2.1), a probability measure Q on $A \times S$ is constructed by endowing the parameter space A with a probability measure μ on $(A, \mathcal{A})$, where the σ-field $\mathcal{A}$ of subsets of A makes $P^a(X)$ measurable for all $X \in \mathcal{S}$ and by extending to $\mathcal{A} \otimes \mathcal{S}$ (in a unique way) the function Q defined on $\mathcal{A} \times \mathcal{S}$ as

$$Q(E \times X) \;=\; \int_E P^a(X)\, d\,\mu, \quad E \in \mathcal{A}, \quad X \in \mathcal{S}. \tag{2.2}$$

The measure constructed from Equation (2.2) is denoted as the Markovian product $Q = \mu \otimes P^{\mathcal{A}}$. Thus, a Bayesian model is defined by the following probability space:

$$\mathcal{E} = (A \times S, \mathcal{A} \vee \mathcal{S}, Q = \mu \otimes P^{\mathcal{A}}). \tag{2.3}$$

Remark 2.1 In this paper, we shall systematically relate the sub-σ-field $\mathcal{B} \subset \mathcal{A}$ (resp., $\mathcal{T} \subset \mathcal{S}$) to the sub-$\sigma$-field[1] of the corresponding cylinders $\mathcal{B} \times S$ (resp. $A \times \mathcal{T}$). Thus, in Equation (2.3), we relate the product $\mathcal{A} \otimes \mathcal{S}$ to $\mathcal{A} \vee \mathcal{S}$, the σ-field generated by $(\mathcal{A} \times S) \cup (A \times \mathcal{S})$. This is to alleviate the notation.

By construction, P^a in Equation (2.2) becomes a probability transition representing a regular version of $P^{\mathcal{A}}$, the restriction to $\mathcal{S}$ of the conditional probability Q given $\mathcal{A}$ and this is so for whatever probability μ on $(A, \mathcal{A})$. Moreover, the so-called prior probability μ corresponds to the marginal probability of Q on $(A, \mathcal{A})$, namely $\mu(E) = Q(E \times S)$ for $E \in \mathcal{A}$. Similarly, the marginal probability P on the sample space $(S, \mathcal{S})$ given by $P(X) = Q(A \times X)$ for $X \in \mathcal{S}$ is called the predictive probability.

In addition to the decomposition $Q = \mu \otimes P^{\mathcal{A}}$, the probability Q is decomposed, under the usual hypotheses (Rao, 1993), into a marginal probability P, and a regular conditional probability given $\mathcal{S}$, represented by a probability transition denoted as $\mu^{\mathcal{S}}$; this is the so-called posterior distribution. When Q is decomposed as $Q = \mu \otimes P^{\mathcal{A}} = P \otimes \mu^{\mathcal{S}}$, the Bayesian model given in (2.3) is said to be regular. For more details, see Martín et al. (1973) and Florens et al. (1990). In the remaining of this paper we assume that the Bayesian model given in (2.3) is regular.

2.2 Family of Bayesian models

The main difference between models given in (2.1) and (2.3) is that the first is a family of sampling distributions, whereas the second is a unique probability measure defined on the product space "parameters $\times$ observations". It should be emphasized that in a Bayesian model, the prior distribution μ is fixed. However, when the interest is focused on the sensitivity of Bayesian procedures with respect to changes on the prior distribution (Macci and Polettini, 2001), or on the Bayesian inference using inter-subjective models

[1] A sub-σ-field is a subset of a σ-field which in turn is a σ-field.

(Dawid, 1979, Section 9), we are dealing with a family of Bayesian models indexed by prior distributions defined on the parameter space $(A, \mathcal{A})$, that is,

$$\mathcal{E} = \{(A \times S, \mathcal{A} \vee \mathcal{S}),\ Q^{\mu} = \mu \otimes P^{\mathcal{A}},\ \ \mu \in \mathcal{P}(A, \mathcal{A})\}, \tag{2.4}$$

where $\mathcal{P}(A, \mathcal{A})$ denotes the space of probability measures defined on the parameter space. It can be seen therefore that Equations (2.1) and (2.4) share a common mathematical structure.

2.3 σ-algebraic perspective of this paper

In the following subsections we define the concepts of statistic, parameter, sufficient parameter, minimal sufficient parameter, sampling information and Bayesian identification, based on the Bayesian model given in (2.3). These definitions are developed using a σ-algebraic perspective, because this is a powerful tool to define a genuinely Bayesian concept of identification (see later). In a sampling theory framework, the data generating process can be reduced either by marginalizing or by conditioning on a statistic; a sufficient reduction means, therefore, that conditionally on a sufficient statistic the sampling process does not provide any useful information on the parameters indexing the sampling probabilities; see Basu (1975).

In the Bayesian model given in (2.3), parameters and statistics play a symmetric role in the sense that A could be interpreted as a statistic and S as a parameter. This is due to the fact that a Bayesian model is fully characterized by a unique probability measure defined on the product space "parameters $\times$ observations". Taking advantage on that, sufficiency can also be defined at the parameter space level. Bayesian identifiability will, consequently, be defined through minimal sufficiency; see Section 2.7. However, in a sampling theory framework, minimal sufficiency can be defined in a dominated statistical model only, that is, when the sampling process can be described in terms of densities; see Pitcher (1957, 1965) and Barra (1981, Chapter II, Section 5). Therefore, it should be asked whether a concept of Bayesian identifiability based on the idea of minimal sufficiency is as general as possible.

A σ-algebraic approach provides us with the necessary tools leading to conclude that, in a Bayesian model, minimal sufficiency is a well defined concept either in dominated or undominated structures. This ensures a Bayesian concept of identifiability useful in parametric, semi-parametric and non-parametric contexts. Furthermore, as shown in Section 2.7, this perspective allows us to understand the role of the prior distribution in Bayesian identifiability.

2.4 Statistics and parameters in a Bayesian model

Consider the Bayesian model given in (2.3). A function T: $(S, \mathcal{S}) \mapsto (U, \mathcal{U})$ is called a statistic if and only if $\mathcal{T} \doteq T^{-1}(\mathcal{U}) \subset \mathcal{S}$. Here, $(U, \mathcal{U})$ is a measurable space (typically, a Borel space), and $\mathcal{T}$, also denoted as $\sigma(T)$ –the σ-field generated by T–, is the smallest σ-field which makes measurable the function T. Therefore, a statistic should be viewed not only as a function with values in U, but also as the set of events that may be described in terms of that random variable; see (Florens and Mouchart, 1982, p. 588). Consequently, $\mathcal{T}$ represents the information provided by the random variable T. This information does not depend on the coordinate system chosen to represent T (namely, the measurable space $(U, \mathcal{U})$) since $\sigma(T) = \sigma[h(T)]$ for all bi-measurable[1] and bijective function.

[1] A bijective function h: $(N, \mathcal{N}) \mapsto (U, \mathcal{U})$ is bi-measurable if, and only if, h is $\mathcal{U}$-measurable, and h^{-1} is $\mathcal{N}$-measurable.

Taking into account the previous considerations, a statistic is a sub-σ-field $\mathcal{T}$ of $\mathcal{S}$. The sub-σ-field $\mathcal{T}$ can heuristically be read as a function T and all bijective and bi-measurable functions of it: they represent the same information.

EXAMPLE 2.1 Let $T \sim \mathrm{N}(0,1)$ and $Y = \exp(T)$. It is known that $\ln(Y)$ is distributed according to a log-normal distribution. Although the probability distributions of T and Y are different, the information provided by them is the same because the transformation $\exp(\cdot)$ is a bijective and bi-measurable function. Such an information is represented by the σ-field $\mathcal{T} = \sigma(T) = \sigma(Y)$.

Similarly, a subparameter, or more simply a parameter, is a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$. The sub-σ-algebra $\mathcal{B}$ can heuristically be read as a function b and all bijective and bi-measurable functions of it.

2.5 Sufficiency in a Bayesian model

2.5.1 Sufficient statistics

Let $\mathcal{T} \subset \mathcal{S}$ be a statistic and $\mathcal{B} \subset \mathcal{A}$ be a parameter. A statistic $\mathcal{T}$ is sufficient if, conditionally on it, the sampling process is independent of the parameter. Using properties of conditional independence (see Appendix A), the last statement can be written as $\mathcal{S} \perp\!\!\!\perp \mathcal{A}|\mathcal{T}$ which is read as "$\mathcal{S}$ is independent of $\mathcal{A}$ given $\mathcal{T}$". Thus, a statistic $\mathcal{T}$ is sufficient for the parameter $\mathcal{A}$ if, for all $\mathcal{S}$-measurable function s, $\mathrm{E}(s|\mathcal{A} \vee \mathcal{T}) = \mathrm{E}(s|\mathcal{T})$; that is, the process generating the observations conditionally on $\mathcal{A} \vee \mathcal{T}$ only depends on the statistic $\mathcal{T}$. Equivalently, for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{S}) = \mathrm{E}(a|\mathcal{T})$, so that the posterior process is fully characterized by the sufficient statistic, the observations $\mathcal{S}$ being redundant once $\mathcal{T}$ is "given". Thus, the original Bayesian model given in (2.3) can be replaced, without any loss of information, by

$$\widetilde{\mathcal{E}} = (A \times S, \mathcal{A} \vee \mathcal{T}, Q_{\mathcal{T}}), \tag{2.5}$$

where $Q_{\mathcal{T}}$ is the restriction of Q on $\mathcal{T}$, that is, $Q_{\mathcal{T}}(E \times X) \doteq Q(E \times X)$ for $E \in \mathcal{A}$ and $X \in \mathcal{T}$. Following Basu (1975)'s and Florens et al. (1990)'s terminology, $\widetilde{\mathcal{E}}$ corresponds to a reduction by sufficiency of the Bayesian model given in (2.3). The Bayesian learning process on the parameter $\mathcal{A}$ is unaffected by this reduction.

2.5.2 Sufficient parameter

Taking advantage of the symmetric role of observations and parameters in a Bayesian model, it is possible to define a sufficient parameter in a way similar to that of a sufficient statistic. As a matter of fact, a parameter $\mathcal{B}$ is sufficient with respect to $\mathcal{S}$ if and only if $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}$, which means that, for all $\mathcal{S}$-measurable function s, $\mathrm{E}(s|\mathcal{A}) = \mathrm{E}(s|\mathcal{B})$; that is, the sufficiency of $\mathcal{B}$ means that $\mathcal{B}$ is "sufficient" to describe the sampling process. Equivalently, $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}$ means that for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{S} \vee \mathcal{B}) = \mathrm{E}(a|\mathcal{B})$; that is, conditionally on $\mathcal{B}$, the observation $\mathcal{S}$ does not bring information on $\mathcal{A}$.

Using the properties of conditional independence, it can be verified that if $\mathcal{B} \subset \mathcal{A}$ is a sufficient parameter for $\mathcal{S}$ and $\mathcal{C} \subset \mathcal{A}$ is a parameter, then $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B} \vee \mathcal{C}$, that is, the parameter $\mathcal{B} \vee \mathcal{C}$ is also sufficient. This means that the sampling process is described not only by $\mathcal{B}$, but also by $\mathcal{B} \vee \mathcal{C}$ for all $\mathcal{C} \subset \mathcal{A}$. In other words, once we have a sufficient description of the sampling process, then such a process can be described using redundant information at the parameter level, which in turn is sufficient. Consequently, it makes sense to look for the minimal sufficient parameter describing the sampling process.

2.6 Minimal sufficient parameter

Let $\Sigma_{\mathcal{A}}$ be the class of sufficient parameters $\mathcal{B} \subset \mathcal{A}$ for $\mathcal{S}$, namely $\Sigma_{\mathcal{A}} = \{\mathcal{B} \subset \mathcal{A}\colon \mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}\}$. It is clear that $\Sigma_{\mathcal{A}}$ is not empty since $\mathcal{A} \in \Sigma_{\mathcal{A}}$. Take therefore $\mathcal{B}_1$, $\mathcal{B}_2 \in \Sigma_{\mathcal{A}}$. Using both the definition of parametric sufficiency and the characterization of conditional independence in terms of a measurability condition (see Appendix A), it follows that the sufficiency of $\mathcal{B}_1$ implies that, for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{A})$ is $\overline{\mathcal{B}}_1$-measurable; and the sufficiency of $\mathcal{B}_2$ implies that, for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{A})$ is $\overline{\mathcal{B}}_2$-measurable. Here, $\overline{\mathcal{B}}_j$ $(j = 1, 2)$ denotes the measurable completion $\overline{\mathcal{B}}_j = \mathcal{B}_j \vee \{E \in \mathcal{A}\colon \mu(E)^2 = \mu(E)\}$, where $\{E \in \mathcal{A}\colon \mu(E)^2 = \mu(E)\}$ is the set of the prior null sets, i.e., the set of all the events whose prior probability is 0 or 1. Then, for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{A})$ is $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$-measurable, that is, $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$; thus, $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2 \in \Sigma_{\mathcal{A}}$. This argument is used to define a minimal sufficient parameter $\mathcal{B}_{\min}$, as follows

$$\mathcal{B}_{\min} = \bigcap_{\mathcal{B} \in \Sigma_{\mathcal{A}}} \overline{\mathcal{B}}. \tag{2.6}$$

$\mathcal{B}_{\min} \in \Sigma_{\mathcal{A}}$ always exists. By construction, the minimal sufficient parameter $\mathcal{B}_{\min}$ contains all the prior null sets. Once the minimal sufficient parameter has been constructed, the original Bayesian model given in (2.3) should be replaced by

$$\mathcal{E}_{\min} = (A \times S, \mathcal{B}_{\min} \vee \mathcal{S}, Q_{\mathcal{B}_{\min}}), \tag{2.7}$$

where $Q_{\mathcal{B}_{\min}}$ is the restriction of Q on $\mathcal{B}_{\min}$, that is, $Q_{\mathcal{B}_{\min}}(E \times X) \doteq Q(E \times X)$ for $E \in \mathcal{B}_{\min}$ and $X \in \mathcal{S}$.

The Bayesian model given in (2.7) does not contain redundant information at the parameter level because there is not a sufficient description of the sampling process (i.e. a $\mathcal{B} \in \Sigma_{\mathcal{A}}$) better than the description provided by the minimal sufficient parameter $\mathcal{B}_{\min}$. Thus, the minimal sufficient parameter corresponds to the greatest possible parameter reduction for which the prior information is updated by the sample, that is, for all $\mathcal{A}$-measurable function a, $\mathrm{E}(a|\mathcal{B}_{\min} \vee \mathcal{S}) = \mathrm{E}(a|\mathcal{B}_{\min})$. Consequently, the learning process underlying a Bayesian model is fully concentrated on the minimal sufficient parameter.

2.7 Minimal sufficient parameter, sampling information and Bayesian identification

It can be proved that the minimal sufficient parameter $\mathcal{B}_{\min}$ is equal to the σ-field generated by every version of the sampling expectations, namely

$$\mathcal{B}_{\min} = \sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\}, \tag{2.8}$$

where $[\mathcal{S}]^+$ denotes the set of non-negative $\mathcal{S}$-measurable functions; see Appendix B for a proof of this result. Thus, the minimal sufficient parameter $\mathcal{B}_{\min}$ coincides with all the information provided by the sampling process. Additionally, this information contains all the prior null sets.

In summary, we have described the sampling process in terms of the information it provides. Using the concept of sufficiency, it has also been clarified how this information is updated by the learning-by-observing process. These considerations allow us to define Bayesian identifiability.

DEFINITION 2.2 In the context of the Bayesian model given in (2.3), the parameter $\mathcal{A}$ is Bayesian identified by $\mathcal{S}$, which is denoted by $\mathcal{A} \prec \mathcal{S}$, if $\mathcal{A}$ is a minimal sufficient parameter; that is, $\mathcal{A} = \mathcal{B}_{\min}$. More generally, let $\mathcal{M}_i \subset \mathcal{A} \vee \mathcal{S}$, for $i = 1, 2, 3$, a function of either the parameters, the observations or both. It is said that $\mathcal{M}_1$ is Bayesian identified by $\mathcal{M}_2$ conditionally on $\mathcal{M}_3$ if $\mathcal{M}_1 \vee \mathcal{M}_3$ is Bayesian identified by $\mathcal{M}_2 \vee \mathcal{M}_3$, i.e.,

$$\sigma\{\mathrm{E}(f|\mathcal{M}_1 \vee \mathcal{M}_3)\colon f \in [\mathcal{M}_2 \vee \mathcal{M}_3]^+\} \;=\; \mathcal{M}_1 \vee \mathcal{M}_3.$$

This relationship is denoted as $\mathcal{M}_1 \prec \mathcal{M}_2|\mathcal{M}_3$, which by definition is equivalent to $\mathcal{M}_1 \vee \mathcal{M}_3 \prec \mathcal{M}_2 \vee \mathcal{M}_3$.

In Bayesian statistics, this concept was introduced by Florens and Mouchart (1977) and further developed by Florens and Rolin (1984), Mouchart and Rolin (1984) and, mainly, by Florens et al. (1990, chapter 4); see also Van Putten and Van Schuppen (1985) (for a table of the correspondence between their results and those contained in Mouchart and Rolin (1984), see Mouchart and Rolin (1985)) and Gourieroux and Monfort (1995). In any case, this concept can be traced back to McKean (1963), although there it is defined using the Lebesgue completion instead of the measurable completion.

From Definition 2.2, there are five features that are worth explaining in more detail. First, what steps should be considered in a Bayesian identification analysis; second, the learning-by-observing process is only based on the Bayesian identified parameter; third, the fact that we have a genuinely Bayesian concept of identifiability; fourth, the given definition yields a relation with Bayesian consistency; and fifth, the concept defined can be related to the sampling identification concept. We summarize these five issues in the following subsections.

2.7.1 A Bayesian identification analysis scheme

The way in which Bayesian identification has been defined allow us to propose a Bayesian identification scheme composed with two main steps: (i) the sufficient parameters describing the data generating process should be made explicit. (ii) Check whether the sufficient parameter found in (i) is the minimal sufficient parameter.

For expository purposes, suppose that a Bayesian model is specified as

$$p(X, a) = p(X|a)\, m(a),$$

where a is a sufficient parameter. The question now is whether a is the minimal sufficient parameter; that is, if a is a sufficient description of the data generating process, and if there exists other sufficient description, namely a sufficient parameter b, such that a is a function of b. By the Dynkin-Doob Lemma (see Appendix A), this is equivalent to $\sigma(a) \subset \sigma(b)$.

Suppose now that a is Bayesian identified by X and that we are interested in defining the Bayesian identifiability of a parameter c, which is a function of the Bayesian identified parameter a. According to the strategy mentioned above, the first step is to check whether c is a sufficient parameter. However, c is a function of the minimal sufficient parameter a. Therefore, if the function is not one-to-one then c does not describe the data generating process and, consequently, its Bayesian identifiability is not well defined. Evidently, it is possible to obtain the Bayesian model $p(X, c)$ from the Bayesian model $p(X, a)$ as

$$p(X|c) = \int p(X|a)\, d\, m(a|c),$$

where $m(a|c)$ is the conditional prior density of a given c. By so doing, c is a sufficient parameter of the data generating process characterized by $p(X|c)$, so the Bayesian identifiability of c should be studied with respect to the "new" Bayesian model $p(X, c)$, and not with respect to the "old" Bayesian model $p(X, a)$. In particular, this means that the Bayesian identifiability of c is not inherited from the Bayesian identifiability of a, as the following example (taken from San Martín et al. (2009)) shows.

EXAMPLE 2.3 Let $\boldsymbol{X}_i = (X_{i1}, X_{i2})^\top$, Σ be an unknown positive definite 2×2 symmetric matrix and $\alpha \neq 0$. Suppose that, for each $i = 1, \ldots, n$,

$$(\boldsymbol{X}_i|\theta_i, \alpha, \Sigma) \overset{\text{i.i.d.}}{\sim} \mathrm{N}_2\left(\begin{pmatrix} \theta_i \\ \alpha\theta_i \end{pmatrix}, \Sigma\right), \tag{2.9}$$

where the prior distribution is specified by the following two conditions:

(i) $(\theta_i|\alpha, \Sigma) \overset{\text{ind.}}{\sim} \mathrm{N}(0, 1)$;
(ii) (α, Σ) have a prior distribution absolutely continuous with respect to the Lebesgue measure.

Taking into account that

$$\theta_i = \mathrm{E}(X_{i1}|\theta_i, \alpha, \Sigma), \quad \alpha\,\theta_i = \mathrm{E}(X_{i2}|\theta_i, \alpha, \sigma) \quad \text{and} \quad \Sigma = \mathrm{V}(\boldsymbol{X}_i|\theta_i, \alpha, \Sigma),$$

it can be verified that, for each $i = 1, \ldots, n$, the parameters $(\alpha, \theta_i, \Sigma)$ are Bayesian identified by $\boldsymbol{X}_i$. Suppose now that we are interested in analyzing the data generating process conditionally on (α, Σ). This process is obtained from Equation (2.9) after marginalizing with respect to θ_i, namely

$$(\boldsymbol{X}_i|\alpha, \Sigma) \overset{\text{i.i.d.}}{\sim} \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} + 1 & \sigma_{12} + \alpha \\ \sigma_{12} + \alpha & \sigma_{22} + \alpha^2 \end{pmatrix}\right), \tag{2.10}$$

It can be verified that, although (α, Σ) is a sufficient parameter for $\boldsymbol{X}_i$, it is not the minimal sufficient parameter; i.e., (α, Σ) is not identified by $\boldsymbol{X}_i$, even though $(\alpha, \theta_i, \Sigma)$ is identified by $\boldsymbol{X}_i$. This is due to the marginalization of the original Bayesian model defined on $(\boldsymbol{X}_i, \theta_i, \alpha, \Sigma)$ into the Bayesian model defined on $(\boldsymbol{X}_i, \alpha, \Sigma)$.

2.7.2 Bayesian identifiability and the learning-by-observing process

The Bayesian identified parameter fully concentrates the learning-by-observing process. To show the validity of this statement, let us consider a parameter $\mathcal{C} \subset \mathcal{A}$ in which we are interested to make inferences. By definition of the identified parameter $\mathcal{B}_{\min}$, $\mathcal{A} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}_{\min}$; this relationship implies that

$$\mathcal{C} \perp\!\!\!\perp \mathcal{S}|\mathcal{B}_{\min}, \quad \forall\, \mathcal{C} \subset \mathcal{A}. \tag{2.11}$$

Here the parameter $\mathcal{C}$, although non-identified by $\mathcal{S}$, might be of interest and, consequently, it should be updated by the observations. Since the posterior expectation $\mathrm{E}(c|\mathcal{S})$ exists for all $\mathcal{C}$-measurable function c, this updating is possible, however, the question is whether we are actually learning about the unidentified parameter $\mathcal{C}$. The answer is not: although we compute the posterior distribution of $\mathcal{C}$, what we are actually updating is always the

identified parameter, nothing more. Indeed, for all $\mathcal{C}$-measurable function c,

$$\begin{aligned}\mathrm{E}(c|\mathcal{S}) &= \mathrm{E}[\mathrm{E}(c|\mathcal{S}\vee\mathcal{B}_{\min})|\mathcal{S}] \qquad \text{(by iterating the conditional expectation)}\\ &= \mathrm{E}[\mathrm{E}(c|\mathcal{B}_{\min})|\mathcal{S}] \qquad \text{(by using Equation (2.11)).}\end{aligned}$$

By definition of conditional expectation, $\mathrm{E}(c|\mathcal{B}_{\min})$ only depends on the identified parameter $\mathcal{B}_{\min}$; this and only this is updated by the observations.

From a practical point of view, this means that if an unidentified parameter is estimated by its posterior distribution, the users should be warned that this estimates does not provide any updating of the unidentified parameter, but only of the identified parameter. This is more relevant when the unidentified parameter is a parameter of interest. In this case, if such a warning is not explicit, erroneous conclusions can be drawn from the analysis, namely, it can be concluded that information about the unidentified parameter has been obtained, when in reality the only information that is obtained is about the identified parameter. We illustrate this point in the following example.

Example 2.4 Consider the problem of estimating the prevalence, the sensitivity and the specificity of a diagnostic test in the absence of a gold-standard. Let $Z_i \in \{0,1\}$ be a binary random variable indicating the true state of a subject i, that is, $Z_i = 1$ if subject i is diseased, and $Z_i = 0$ otherwise. Let $Y_i \in \{0,1\}$ be a binary random variable indicating the classification of a subject i through a diagnostic test, that is, $Y_i = 1$ if subject i is classified as diseased by the diagnostic test, and $Y_i = 0$ otherwise. Note that Y_i is an observed variable, whereas Z_i is unobservable. Taking into account that this is a problem of misclassification, the parameters of interest are the sensitivity of the test, $\alpha = P(Y_i = 1|Z_i = 1)$; the specificity of the test, $\beta = P(Y_i = 0|Z_i = 0)$; and the true prevalence, $\omega = P(Z_i = 1)$. The sampling process is given by a sequence of mutually independent random variables Y_i conditionally on $p(\alpha,\beta,\omega)$, where

$$p(\alpha,\beta,\omega) = P(Y_i = 1|\alpha,\beta,\omega) = \alpha\omega + (1-\beta)(1-\omega). \tag{2.12}$$

If it is assumed that $\alpha + \beta > 1$, then $p(\alpha,\beta,\omega)$ is an increasing function of ω. The model is completed by specifying a prior probability distribution on (α,β,ω). The σ-field of the sample space is given by $\mathcal{S} = \sigma(Y_1,\ldots,Y_n)$, and the one of the parameters is given by $\mathcal{A} = \sigma(\alpha)\vee\sigma(\beta)\vee\sigma(\omega)$. Now, $p(\alpha,\beta,\omega) = \mathrm{E}(Y_i|\mathcal{A})$, so

$$\sigma\{p(\alpha,\beta,\omega)\} \subset \sigma\{\mathrm{E}(f|\mathcal{A})\colon f \in [\mathcal{S}]^+\} \subset \mathcal{A}.$$

Moreover, $\sigma(\alpha) \not\subset \sigma\{p(\alpha,\beta,\omega)\}$, $\sigma(\beta) \not\subset \sigma\{p(\alpha,\beta,\omega)\}$, and $\sigma(\omega) \not\subset \sigma\{p(\alpha,\beta,\omega)\}$, because there not exist measurable functions such that α is a function of $p(\alpha,\beta,\omega)$, β is a function of $p(\alpha,\beta,\omega)$, and ω is a function of $p(\alpha,\beta,\omega)$: it is enough to consider relationship given in Equation (2.12), in which there is one equation and three unknowns.

Therefore $\mathcal{A}$ is not Bayesian identified, and neither are α, β and ω. Furthermore, from the equality $\mathrm{E}(Y_i|\mathcal{A}) = \mathrm{E}(Y_i|p(\alpha,\beta,\omega))$, it follows that the sampling process is fully described by $p(\alpha,\beta,\omega)$. In other words, $p(\alpha,\beta,\omega)$ is Bayesian identified by Y_i. Now, in this example, it is always possible to compute the posterior expectation of the parameters of interest. However, these expectations does not provide any information about them, but only on the apparent prevalence $p(\alpha,\beta,\omega)$.

2.7.3 The role of the prior distribution in Bayesian identifiability

The concept of identification as introduced in Definition 2.2 is genuinely Bayesian because it depends on the prior distribution through the prior null events. This means that if, additionally to the Bayesian model given in (2.3), a second model is specified in such a way that

$$\mathcal{E}' = (A \times S, \mathcal{A} \vee \mathcal{S}, Q' = \mu' \otimes P^{\mathcal{A}}),$$

then $\mathcal{B}_{\min}$ is still the Bayesian identified parameter in the context of $\mathcal{E}'$ if and only if μ and μ' are equivalent probability measures (i.e., they have the same null sets. Thus, the Bayesian identifiability property (in the context of the Bayesian model given in (2.3)) can be lost if the prior null sets change; for more details, see Florens et al. (1990, Proposition 4.6.8). This means that an unidentified parameter does not become identified once a proper prior distribution is specified on it, even if it is "well concentrated" on the unidentified parameter, unless prior null sets are defined –which is equivalent to introduce dogmatic constraints. The point is illustrated in the following example.

Example 2.5 Let $A = \mathbb{R}$ and $\mathcal{A}$ be the Borel sets of $\mathbb{R}$. Assume that the sampling process is specified as $(X_i|a) \sim \mathrm{N}(|a|, 1)$, where $X_1, \ldots . X_n$ are mutually independent conditionally on $\mathcal{A}$. Let μ be a prior probability distribution. Since $|a| = \mathrm{E}(X_i|\mathcal{A})$, it follows that the corresponding minimal sufficient parameter is given by

$$\sigma\{|a|\} \vee \{E \in \mathcal{A}\colon \mu(E) = 0 \text{ or } 1\},$$

where $\sigma\{|a|\} = \{E \in \mathcal{A}\colon -E = E\}$. If μ is equivalent to the Lebesgue measure (i.e., μ has the same null sets as the Lebesgue measure on $\mathbb{R}$), then the parameter a is not identified because $\sigma\{a\} \not\subset \sigma\{|a|\}$, that is, by the Dynkin-Doob Lemma (see Appendix A), there exists no measurable function h such that $a = h(|a|)$. However, if μ is equivalent to the Lebesgue measure on $\mathbb{R}^+$ and $\mu(\mathbb{R}^-) = 0$, then the parameter a is identified.

2.7.4 Bayesian identifiability and convergence of posterior expectations

In an i.i.d. process conditionally on a parameter $\mathcal{A}$, the corresponding Bayesian identified parameter is not only Bayesian consistent (in the sense that it is the limit of its posterior expectation as $n \to \infty$), but also concentrates, in the limit, the posterior expectation of any other parameter. In this sense, Bayesian identification is related to consistency, just as (in a pure sampling theory framework) identification is related to consistency; see San Martín and Quintana (2002). For details, see Appendix C.

2.7.5 Bayesian identification and sampling identification

In a pure sampling theory framework, parameter identifiability is defined as the injectivity of the mapping $a \mapsto P^a$, where $\{P^a\colon a \in A\}$ is a family of sampling distributions. This identification concept (which we call sampling identification or s-identification) is related with Bayesian identification as introduced in Definition 2.2. More precisely, if $\mathcal{A}$ is a Blackwell σ-field[1] and $\mathcal{S}$ is separable[2], then s-identification implies Bayesian identification for

[1] A σ-field $\mathcal{M}$ is a Blackwell σ-field if $\mathcal{M}$ is separable and if, for all $\mathcal{M}$-measurable function m and for all $A \in \mathcal{M}$, $m(A)$ is an analytic set of $\mathbb{R}$; see Blackwell (1956) and Florens et al. (1990, Chapter 0). An example of a Blackwell σ-field is the Borel σ-field defined on the real line.

[2] A σ-field $\mathcal{S}$ is separable if there exist a countable family of events $\mathcal{C}$ such that $\mathcal{S} = \sigma(\mathcal{C})$.

all prior probability measure on $(A, \mathcal{A})$; for details and proofs of these results, see Florens et al. (1985) and Florens et al. (1990, chapter 4).

It should be emphasized that this relationship between s-identification and Bayesian identification depends on the separability of both the sample space and the parameter space. Since we use measurable completion (see Appendix A), we avoid the danger of losing the separability and, consequently, we may use identification results established in a pure sampling theory approach.

3. Identification in a Fully Discrete Bayesian Model

The preceding sections explain, in a general way, the concepts of Bayesian model and Bayesian identifiability. The σ-fields $\mathcal{A}$ and $\mathcal{S}$ could have been either uncountable or finite, or a mixture of both. In this section we illustrate these concepts for the particular case in which both $\mathcal{A}$ and $\mathcal{S}$ are σ-fields of finite subsets, yielding a fully discrete Bayesian model.

3.1 Model construction

Let $S = \{s_1, \ldots, s_n\}$ be a finite set of observations and $A = \{a_1, \ldots, a_m\}$ be a finite set of parameters. The corresponding σ-fields are the power sets of S and of A, respectively. Let μ be a prior probability defined on $(A, \mathcal{A})$ and let $A_\mu = \{a \in A\colon \mu(a) > 0\}$. The parametric support of an observation $s \in S$ is defined as $A_s = \{a \in A\colon p(s|a) > 0\}$. The sampling probabilities are defined as

$$\begin{aligned} p(s_i|a) &= \sum_{a_j \in A_\mu} p(s_i|a_j) \mathbb{1}_{\{a=a_j\}} \\ &= \sum_{a_j \in A_\mu \cap A_{s_i}} p(s_i|a_j) \mathbb{1}_{\{a=a_j\}}, \qquad i = 1, \ldots, n. \end{aligned}$$

When $a \notin A_\mu$, the sampling probabilities are arbitrarily defined as $p(s|a) = c$ (for $s \in S$), with $c \neq 0$. Therefore, $A_\mu^c \subset A_s$ for each $s \in S$, and

$$q(s,a) = \begin{cases} p(s|a)\mu(a),\ s \in S, \ \ a \in A_\mu \cap A_s, \\ \\ \qquad 0, \qquad s \in S, \ \ a \in A_\mu^c \cup (A_\mu \cap A_s^c). \end{cases}$$

The predictive distribution is defined as

$$p(s) = \sum_{a_j \in A_\mu \cap A_s} p(s|a_j)\, \mu(a_j) = \sum_{a_j \in A} p(s_i|a_j)\mu(a_j), \qquad s \in S.$$

It should be remarked that for $s \in S$

$$p(s) > 0 \quad \Longleftrightarrow \quad A_\mu \cap A_s \neq \emptyset.$$

Therefore the joint probability $q(s,a)$ can be decomposed as

$$\begin{pmatrix} q(a_1,s_1) & \cdots & q(a_1,s_n) \\ \vdots & \ddots & \vdots \\ q(a_m,s_1) & \cdots & q(a_m,s_n) \end{pmatrix} = \begin{pmatrix} \mu(a_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu(a_m) \end{pmatrix} \begin{pmatrix} p(s_1|a_1) & \cdots & p(s_n|a_1) \\ \vdots & \ddots & \vdots \\ p(s_1|a_m) & \cdots & p(s_n|a_m) \end{pmatrix}$$

$$= \begin{pmatrix} \mu(a_1|s_1) & \cdots & \mu(a_1|s_n) \\ \vdots & \ddots & \vdots \\ \mu(a_m|s_1) & \cdots & \mu(a_m|s_n) \end{pmatrix} \begin{pmatrix} p(s_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p(s_n) \end{pmatrix}. \quad (3.1)$$

These equalities illustrate the difference between a probability measure (the prior and the predictive) and a probability transition (the sampling and the posterior): the first one corresponds to a diagonal matrix, whereas the second corresponds to a rectangular matrix. It can also be verified the dominance property of the posterior transition with respect to the prior distribution; see Mouchart (1976). In other words, the prior null events are still posterior null events. More precisely,

(i) If $\mu(a_j) = 0$, then $\mu(a_j|s) = 0$ for all $s \in S$;

(ii) If $\mu(a_j) = 1$, then $\mu(a_k) = 0$ for each $k \neq j$ and, therefore, $\mu(a_k|s) = 0$ for each $k \neq j$ and for $s \in S$. Hence $\mu(a_j|s) = 1$ for each $s \in S$.

It can also be grasped that the converse relationship (namely, the dominance of the prior distribution with respect to the prior distribution) is not necessarily true. In fact, it could exists a $s \in S$ such that $\mu(a|s) = 0$ for $a \in A_s^c \cap A_\mu$; note that this intersection is equal to A_s^c since $A_\mu^c \subset A_s$.

3.2 Statistics and parameters in a discrete Bayesian model

In the discrete Bayesian model described in the previous section, a parameter $b(a)$ is characterized by the partition it induces on the parameter space A, namely

$$A_b = \{a \in A\colon b(a) = b\} \subset A.$$

Thus, the information provided by the parameter $b(a)$ corresponds to A_b. Similarly, a statistic $t(s)$ is characterized by the partition it induces on the sample space, namely

$$S_t = \{s \in S\colon t(s) = t\} \subset S,$$

and the information provided by the statistic $t(s)$ corresponds to S_t.

3.3 Bayesian identification in the discrete case

In this section, Bayesian identification is characterized for a fully discrete Bayesian model. More specifically, the idea is to characterize the relationship $X_1 \prec X_2|X_3$, where X_1, X_2, X_3 are discrete random variables defined on a common probability space $(M, \mathcal{M}, P)$ such that $X_r\colon M \mapsto N_r$, where N_r, for $r = 1,2,3$, are finite sets. These random variables can be interpreted differently. For example, X_1 can be interpreted as a parameter, X_2 as an observation, and X_3 as a latent variable or as a parameter. X_3 can also be assumed to be a constant a.s. In this case, the generated σ-field is given by $\{E \in \mathcal{M}\colon P(E) = P^2(E)\}$.

Let us define

(i) $K = \{k \in N_3\colon P(X_3 = k) > 0\}$.

(ii) $N_1^k = \{i \in N_1\colon P(X_1 = i|X_3 = k) > 0\}$ for $k \in K$.

(iii) $N_2^k = \{j \in N_2\colon P(X_2 = j|X_3 = k) > 0\}$ for $k \in K$.

(iv) $\left(P^k\right)_{ij} = P(X_1 = i, X_2 = j|X_3 = k)$, a $|N_1^k| \times |N_2^k|$ matrix for $k \in K$.

By Definition 2.2, X_1 is Bayesian identified by X_2 conditionally on X_3 (i.e. $X_1 \prec X_2|X_3$) if and only if

$$\sigma\left\{P(X_2 = j, X_3 = k|X_1, X_3)\colon k \in K,\ j \in N_2^k\right\} = \tag{3.2}$$

$$\sigma\left\{\{X_1 = i\} \cap \{X_3 = k\}\colon k \in K,\ i \in N_1^k\right\} \vee \sigma\left\{\{X_1 = i\} \cap \{X_3 = k\}\colon k \in N_3,\ i \in (N_1^k)^c\right\}.$$

As a matter of fact, recalling the definition of Bayesian identification, the σ-field of the left-side of Equation (3.2) contains all the null sets of (X_1, X_3). This class corresponds to the σ-field

$$\sigma\left\{\{X_1 = i\} \cap \{X_3 = k\}\colon k \in N_3,\ i \in (N_1^k)^c\right\}$$

because

$$P(X_1 = i, X_3 = k) \begin{cases} = P(X_1 = i|X_3 = k)\, P(X_3 = k),\ i \in (N_1^k)^c,\ k \in K; \\ \\ \leq P(X_3 = k),\ k \in K^c. \end{cases}$$

Thus, $P(X_1 = i|X_3 = k) = 0$ since $i \in (N_1^k)^c$ and $k \in K$, whereas $P(X_3 = k) = 0$ since $k \in K^c$. Therefore, $P(X_1 = i, X_3 = k) = 0$ for $i \in (N_1^k)^c$ and $k \in N_3$.

By this same fact –that all the null sets of (X_1, X_3) are contained into the σ-field of the left-side of Equation (3.2)–, equality in Equation (3.2) is equivalent to the following relationship:

$$\sigma\left\{\{X_1 = i\} \cap \{X_3 = k\}\colon k \in K,\ i \in N_1^k\right\} \subset \sigma\left\{P(X_2 = j, X_3 = k|X_1, X_3)\colon k \in K,\ j \in N_2^k\right\}. \tag{3.3}$$

Now, let us characterize the generators of

$$\sigma\left\{P(X_2 = j, X_3 = k|X_1, X_3)\colon k \in K,\ j \in N_2^k\right\}.$$

For each $k \in K$,

$$P(X_2 = j, X_3 = k|X_1, X_3) = \sum_{i \in N_1^k} p_{j|ik}\, \mathbb{1}_{\{X_1 = i, X_3 = k\}} \doteq Y_j.$$

Then, for each $j \in N_2^k$,

$$Y_j^{-1}\left[\{p_{j|ik}\}\right] = \{X_1 \in I_{ij},\ X_3 = k\},$$

where $I_{ij} \subset N_1^k$ is a set which depends on $(i,j) \in N_1^k \times N_2^k$. Moreover, $Y_j^{-1}\left[\{p_{j|ik}\}\right]$ contains $\{X_1 = i, X_3 = k\}$. It follows that

$$\{X_1 = i, X_3 = k\} \subset \bigcap_{j \in N_2^k} Y_j^{-1}\left[\{p_{j|ik}\}\right] = \{X_1 \in I_i, X_3 = k\},$$

where

$$I_i = \bigcap_{j \in N_2^k} I_{ij}.$$

Thus, $\{X_1 \in I_i, X_3 = k\}$ is the smallest set in

$$\sigma\left\{P(X_2 = j, X_3 = k|X_1, X_3)\colon k \in K,\ j \in N_2^k\right\}$$

containing $\{X_1 = i, X_3 = k\}$. Therefore, X_1 is Bayesian identified by X_2 conditionally on X_3 –which is equivalent to relation given in Equation (3.3)– if and only if, for each $k \in K$,

$$I_i = \{i\} \quad \forall\, i \in N_1^k,$$

which is equivalent to

$$\bigcap_{j \in N_2^k} \left\{\omega\colon \sum_{l \in N_1^k} p_{j|lk} \mathbb{1}^{(\omega)}_{\{X_1 = l, X_3 = k\}} = p_{j|ik}\right\} = \{X_1 = i,\ X_3 = k\} \quad \forall\, i \in N_1^k.$$

This last condition is equivalent to the following one: for each $k \in K$, $\nexists\ i, i' \in N_1^k$ such that $p_{j|ik} = p_{j|i'k}\ \forall\, j \in N_2^k$, which in turn can be written as $\exists\ i, i' \in N_1^k$ such that $p_{ij|k} = c_{ii'}\, p_{i'j|k}\ \forall\, j \in N_2^k$. Summarizing, we obtain the following theorem.

THEOREM 3.1 Let $(\Omega, \mathcal{M}, P)$ be a probability space and $X_r\colon \Omega \mapsto N_r$, for $r = 1, 2, 3$, be random variables, where N_r $(r = 1, 2, 3)$ are finite sets. The following are equivalent:

(i) $X_1 \prec X_2|X_3$, that is, X_1 is Bayesian identified by X_2 conditionally on X_3.

(ii) For each $k \in K$, any two rows of P^k are linearly independent.

(iii) For each $k \in K$ and for each $i, i' \in N_1^k$, there not exists a $c_{ii'}$ such that

$$P(X_1 = i, X_2 = j|X_3 = k) = c_{ii'} P(X_1 = i', X_2 = j|X_3 = k) \quad \forall\, j \in N_2^k.$$

3.4 Relationships between Bayesian and sampling identification

As pointed out in Section 2.7.5, sampling identification implies Bayesian identification for all prior distribution provided the parameter space is Blackwell and the sample space is separable. This implication can be verified in the fully discrete Bayesian model. As a matter of fact, consider the fully discrete Bayesian model as specified in Section 3.1. By Theorem 3.1, the parameter a is Bayesian identified by s if and only if any two rows of

the matrix

$$\begin{pmatrix} q(a_1, s_1) & \cdots & q(a_1, s_n) \\ \vdots & \ddots & \vdots \\ q(a_m, s_1) & \cdots & q(a_m, s_n) \end{pmatrix}$$

are linearly independent. Considering decomposition given in Equation (3.1), this is equivalent to the following property: any two rows of the matrix representing the sampling transition are linearly independent; that is, the mapping $a \mapsto p(\cdot|a)$ is injective, for all $a \in A_\mu \subset A$. Now, if $A_\mu = A$, then Bayesian identification and sampling identification are equivalent concepts. The property $A_\mu = A$ means that the prior distribution μ puts a positive mass on each element of A. Let us summarize this relationship in the following theorem.

THEOREM 3.2 Consider the fully discrete Bayesian model as specified in Section 3.1. If $A_\mu = A$, then Bayesian identification and sampling identification are equivalent.

3.5 Updating unidentified parameters

As pointed out in Section 2.7.2, the posterior distribution of an unidentified parameter can be computed. However, from a modeling point of view, the statistical meaning of this posterior distribution is of interest. Let us consider, therefore, a fully discrete Bayesian model defined by the following components: $S = \{s_1, s_2\}$, $A = \{a_1, a_2, a_3\}$, the sampling process characterized by

$$p(s_1|a_1) = \frac{1}{2}, \quad p(s_1|a_2) = p(s_1|a_3) = \frac{1}{3},$$

and the prior distribution satisfying that $\mu(a_i) > 0$ for $i = 1, 2, 3$. All these components imply the following joint distribution of (a, s), with $a \in A$ and $s \in S$:

	s_1	s_2
a_1	$\frac{1}{2}\,\mu(a_1)$	$\frac{1}{2}\,\mu(a_1)$
a_2	$\frac{1}{3}\,\mu(a_2)$	$\frac{2}{3}\,\mu(a_2)$
a_3	$\frac{1}{3}\,\mu(a_3)$	$\frac{2}{3}\,\mu(a_3)$

By Theorem 3.1, the Bayesian identified parameter is characterized by the partition $\{\{a_1\}, \{a_2, a_3\}\}$. The posterior distribution of a_1 is given by

$$p(a_1|s) = \frac{\frac{1}{2}\mu(a_1)}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_1\}} + \frac{\frac{1}{2}\mu(a_1)}{\frac{1}{2}\mu(a_1) + \frac{2}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_2\}};$$

the posterior distribution of (a_2, a_3) is given by

$$p(a_2, a_3|s) = \frac{\frac{1}{3}[\mu(a_2) + \mu(a_3)]}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_1\}} + \frac{\frac{2}{3}[\mu(a_2) + \mu(a_3)]}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_2\}}.$$

The posterior distribution of a_2 and of a_3 –which are unidentified parameters– are respectively given by

$$p(a_2|s) = \frac{\frac{1}{3}\mu(a_2)}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_1\}} + \frac{\frac{2}{3}\mu(a_2)}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_2\}},$$

$$p(a_3|s) = \frac{\frac{1}{3}\mu(a_3)}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_1\}} + \frac{\frac{2}{3}\mu(a_3)}{\frac{1}{2}\mu(a_1) + \frac{1}{3}[\mu(a_2) + \mu(a_3)]}\mathbb{1}_{\{s=s_2\}}.$$

It follows that

$$p(a_2|s) = \frac{\mu(a_2)}{\mu(a_2) + \mu(a_3)}\, p(a_2, a_3|s), \qquad p(a_3|s) = \frac{\mu(a_3)}{\mu(a_2) + \mu(a_3)}\, p(a_2, a_3|s). \tag{3.4}$$

That is, the posterior distribution of the unidentified parameter a_2 coincides with a function of the posterior distribution of the identified parameter (a_2, a_3). It occurs similarly for the posterior distribution of a_3. It is also clear from (3.4) that the prior distribution does not have any impact in repairing the lack of identifiability of a_2 and of a_3. Therefore, it is not worth calculating the posterior probability of these unidentified parameters, but rather replace the original Bayesian model with the following:

	s_1	s_2
a_1	$\frac{1}{2}\,\mu(a_1)$	$\frac{1}{2}\,\mu(a_1)$
a_2, a_3	$\frac{1}{3}\,[\mu(a_2) + \mu(a_3)]$	$\frac{2}{3}\,[\mu(a_2) + \mu(a_3)]$

In this case, the σ-field $\mathcal{A}$ of the parameter space is given by $\{\emptyset, \{a_1\}, \{a_2, a_3\}, A\}$. It can be noted that the events $\{a_2\}$ and $\{a_3\}$ are not measurable, that is, they are not events of interest; their prior probabilities can not be computed, neither their posterior probabilities.

4. Discussion

The identification problem arose as a consequence of a reformulation of the specification problem as stated by Fisher (1922). Such a reformulation establishes that "the investigator's inquisitiveness is not just a population in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated"; see Koopmans and Reiersøl (1950, p.165) and also Hurwicz (1950). The identification problem consists, therefore, in investigating whether one and only one structure explains the observed phenomenon.

These considerations explain why identifiability was considered as a pre-statistical problem (McHugh, 1956) and, consequently, more related to statistical modeling than statistical inference. In spite of that, identifiability has been traditionally considered as a necessary condition to ensure a coherent inference –that is, the existence of unbiased estimators and consistent estimators; see San Martín and Quintana (2002) and the references therein. In a pure sampling theory approach, this constitutes a limitation in the sense that it is not possible to compute estimators of unidentified parameters. Bayesian statistics seems to be a way out to this limitation. As a matter of fact, it is always possible (if not, some hypotheses can be introduced) to compute the posterior distribution of unidentified parameters. However, from a modeling point of view, the problem is to know what is the statistical meaning of these estimators. In order to provide an answer to this question,

an identifiability analysis is unavoidable: once a parameter of interest has been identified, its statistical meaning becomes explicit. In practice, this is the issue which matters; for an example in psychometrics, see San Martín et al. (2009). Therefore, in problems where parameters have a substantive meaning, Bayesian statistics will be useful if identifiability is warranted. Otherwise, the analysis will only provide illusory solutions.

Let us mention another debate, related to the identifiability of complex hierarchical models. This type of models is specified by means of a marginal latent model generating the unobservable variables (or random effects), and a conditional model generating the observable variables conditionally on the unobservable ones. For instance, in the class of generalized linear mixed models, the marginal latent model corresponds to the process generating the random effects, whereas the conditional model is specified as a generalized linear model; see Verbekee and Molenberghs (2000). The statistical model, bearing on the observable variables only, is obtained after integrating out the latent variables or random effects; this model is indexed by two type of parameters: those indexing the marginal latent model, and those indexing the conditional model. In the statistical literature, there exists some rules relating the identifiability of the statistical model to the identifiability of the marginal latent model and/or the identifiability of the conditional model. For instance, in the psychometric literature dealing with structural equation models, it has been proposed (without formal proof) that the identification of both the marginal latent and the conditional models are sufficient to ensure the identification of the statistical model; see Jöreskog (1981, pp. 89-90), Bollen (1989, p. 328) and Maruyama (1998, p. 191). In the literature dealing with Generalized Linear Mixed Models, similar examples can be found. For instance, Chen and Dey (1998) point out that, when the distribution of the random effects is a normal $\mathrm{N}(0, \varphi^2)$, the variance φ^2 "is nearly not identified" (p. 325), although such a concept is not well defined neither in a Bayesian approach, nor in a sampling theory framework. Although this type of results may appear to be intuitive, it is possible to offer counterexamples to this type of rules. As an illustration, consider Example 2.3, where θ_i can now be viewed as a random effect; after integrating out θ_i, the identifiability of the parameters indexing the conditional model is lost when its identifiability is considered in the statistical model; for more details, see San Martín (2003), San Martín and Mouchart (2007) and San Martín et al. (2009). Other counter examples based on the fully discrete Bayesian model of Section 3 are being considered by the authors as future work.

Last, but not least, Bayesian identification is always implied by sampling identification. This means that (some) identification results established in a pure sampling theory framework could be useful when models are specified under a Bayesian approach.

Appendix

Appendix A. Conditional Independence

A.1 General definition

The concept of conditional independence becomes relevant in statistical theory, where it can be used as a basic tool to express many of the important concepts of statistics (such as sufficiency, ancillarity, identifiability, etc.), unifying many areas that are, at first sight, different.

Let $(M, \mathcal{M}, P)$ be a probability space and N be a sub-σ-algebra of $\mathcal{M}$. The completed σ-field $\overline{\mathrm{N}}$ of N is defined as

$$\overline{\mathrm{N}} = \mathrm{N} \vee \{E \in \mathcal{M}\colon P(A) = P^2(A)\},$$

that is, the σ-algebra generated by the union between N and the completed trivial σ-field.

Following Florens et al. (1990) (see also Chow and Teicher, 1988), the σ-fields are only completed by measurable sets and not by subsets of measurable sets as is usually done in Lebesgue completion. In this way, it is avoided the danger of losing the separability of σ-field. As mentioned in the main text, the separability of σ-fields is essential to relate Bayesian and sampling identification.

DEFINITION A.1 Let $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$ be sub-σ-algebras of $\mathcal{M}$. It is said that $\mathcal{M}_1$ is independent of $\mathcal{M}_2$ conditionally on $\mathcal{M}_3$, denoted as $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 | \mathcal{M}_3$, if and only if one of the following equivalent conditions hold:

(i) $\mathrm{E}(f_1\, f_2|\mathcal{M}_3) = \mathrm{E}(f_1|\mathcal{M}_3)\,\mathrm{E}(f_2|\mathcal{M}_3)$ a.s. for all positive function f_i measurable with respect to $\mathcal{M}_i$, for $i = 1, 2$.

(ii) $\mathrm{E}(f_1|\mathcal{M}_2 \vee \mathcal{M}_3) = \mathrm{E}(f_1|\mathcal{M}_3)$ a.s. for all positive function f_1 measurable with respect to $\mathcal{M}_1$.

For a proof on the equivalence between (i) and (ii), see Florens et al. (1990, Theorem 2.2.1). When $\mathcal{M}_3$ is equal to the trivial σ-field $\{\emptyset, M\}$, this definition reduces to the usual independence between σ-fields; in such a case, we write $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2$.

It is clear from condition (i) that the concept of conditional independence is symmetric in $\mathcal{M}_1$ and $\mathcal{M}_2$, namely $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 | \mathcal{M}_3$ is equivalent to $\mathcal{M}_2 \perp\!\!\!\perp \mathcal{M}_1 | \mathcal{M}_3$. Condition (ii) provides an heuristic meaning of conditional independence: $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 | \mathcal{M}_3$ means that the process generating $\mathcal{M}_1$ conditionally on $\mathcal{M}_2 \vee \mathcal{M}_3$ depends on $\mathcal{M}_3$ only; or, equivalently, the process generating $\mathcal{M}_2$ conditionally on $\mathcal{M}_1 \vee \mathcal{M}_3$ depends on $\mathcal{M}_3$ only. This heuristic meaning actually corresponds to a measurability property of conditional independence: $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 | \mathcal{M}_3$ if and only if, for all $\mathcal{M}_2$-measurable function f_2, $\mathrm{E}(f_2|\mathcal{M}_1 \vee \mathcal{M}_3)$ is measurable with respect to $\overline{\mathcal{M}}_3$; for a proof of this result, see Florens et al. (1990, Theorem 2.2.6). For details and properties of conditional independence, the reader is referred to Martín et al. (1973), Dawid (1979), Döhler (1980) and Florens et al. (1990), among many others.

A.2 Conditional independence: its meaning in terms of random variables

For a better understanding of the abstract concept of conditional independence we now present its definition in terms of random variables. this presentation is based on the Lemma of Dynkin-Doob. This lemma establishes the following: let $(M, \mathcal{M})$, (N, N) and $(O, \mathcal{O})$ be three measurable spaces and let X: $M \mapsto N$ be a measurable function (that is, $\sigma(X) \doteq X^{-1}(\mathrm{N}) \subset \mathcal{M}$) and let Z: $M \mapsto O$ be measurable function with respect to $\sigma(X)$ (that is, $\sigma(Z) \doteq Z^{-1}(\mathcal{O}) \subset \sigma(X)$). Then there exists a function Y: $N \mapsto O$ measurable with respect to N (that is, $\sigma(Y) \doteq Y^{-1}(\mathcal{O}) \subset \mathrm{N}$) such that $Z = Y \circ X$; for a proof of this result, see Dellacherie and Meyer (1975) and Rao (1984). Figure A1 summarizes these relationships.

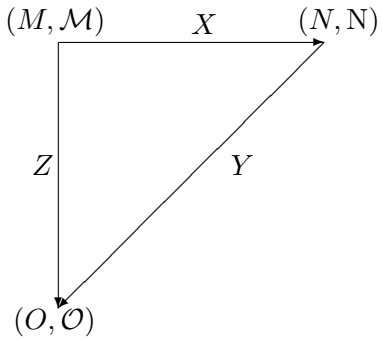

Figure A1. Measurability relationships between X and Z.

As pointed out in the main text (see Section 2.4), the σ-field generated by a random variable corresponds to the set of events that may be described in terms of that random variable. In this sense, the Lemma of Dynkin-Doob establishes that when the information provided by a random variable Z is strictly contained in the information provided by X (that is, the events described by Z are contained into the events described by X), then Z is a measurable transformation of X –i.e., Z is a reduction of X.

Let m_i: $(M, \mathcal{M}) \mapsto (\mathbb{R}, \mathcal{B})$, $i = 1, 2, 3$, be three random variables and let $\mathcal{M}_i = \sigma(m_i)$. Now, $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 | \mathcal{M}_3$ if and only if, for all $\mathcal{M}_2$-measurable function f_2, $\mathrm{E}(f_2|\mathcal{M}_2 \vee \mathcal{M}_3)$ is $\overline{\mathcal{M}}_3$-measurable. Using the Lemma of Dynkin-Doob, this is equivalent to say that for all measurable function h, there exists a measurable function g such that

$$\mathrm{E}(h(m_1)|\mathcal{M}_2 \vee \mathcal{M}_3) = g(m_3) \quad \text{a.s.}$$

That is, the conditional expectation of all measurable transformations of m_1 given (m_2, m_3) are a.s. a function of m_3.

Appendix B. Proof of Equality (2.8)

The σ-field generated by every version of the sampling expectations, namely $\sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\}$, is the smallest sub-$\sigma$-field of $\mathcal{A}$ that makes the sampling expectations measurable; here, $[\mathcal{S}]^+$ denotes the set of non-negative $\mathcal{S}$-measurable functions. Using the characterization of conditional independence in terms of a measurability condition (see Appendix A), the latter condition is equivalent to

$$\mathcal{A} \perp\!\!\!\perp \mathcal{S} | \sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\}.$$

Therefore, $\sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\} \in \Sigma_{\mathcal{A}}$ and, consequently,

$$\sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\} \supset \mathcal{B}_{\min}.$$

On the other hand, $\mathcal{B}_{\min}$ satisfies the condition $\mathcal{A} \perp\!\!\!\perp \mathcal{S} | \mathcal{B}_{\min}$, which is equivalent to say that, for all $\mathcal{S}$-measurable function s, $\mathrm{E}(s|\mathcal{A})$ is $\mathcal{B}_{\min}$-measurable. Therefore, by definition of $\sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\}$, it follows that

$$\sigma\{\mathrm{E}(s|\mathcal{A})\colon s \in [\mathcal{S}]^+\} \subset \mathcal{B}_{\min}.$$

Appendix C. Bayesian Identifiability and Consistency

Let $\{X_n\colon n \in \mathbb{N}\}$ be i.i.d. random variables conditionally on $\mathcal{A}$. In the context of the asymptotic Bayesian model, it can be shown that the minimal sufficient parameter $\mathcal{B}_{\min} \doteq \sigma\{\mathrm{E}(f|\mathcal{A})\colon f \in [\mathcal{X}_1^\infty]^+\}$ satisfies the following condition:

$$\mathcal{B}_{\min} \subset \overline{\mathcal{X}_1^\infty}, \tag{C1}$$

where $\mathcal{X}_1^\infty = \sigma(X_1, \ldots, X_n, \ldots)$; for a proof of this result, see Florens et al. (1990, Theorem 9.3.2). This means that, in an i.i.d. process, the Bayesian identified parameter is a.s. a

function of $(X_1, X_2, \dots)$. Furthermore, condition (C1) ensures the a.s.-convergence as well as the L^q-convergence of the following posterior expectations:

(i) $\mathrm{E}(b|\mathcal{X}_1^n)$ for all $\mathcal{B}_{\min}$-measurable function b;

(ii) $\mathrm{E}(a|\mathcal{X}_1^n)$ for all $\mathcal{A}$-measurable function a.

Here $\mathcal{X}_1^n = \sigma(X_1, \dots, X_n)$ and, for $q \in [1, \infty]$, $L^q(A, \mathcal{A}, \mu)$ is the set of the q-integrable $\mathcal{A}$-measurable functions. As a matter of fact, by the Martingale Theorem, $\mathrm{E}(b|\mathcal{X}_1^n)$ converges a.s. Moreover, it converges to $\mathrm{E}(b|\mathcal{X}_1^\infty)$ in L^q for all $b \in L^q(A, \mathcal{B}_{\min}, \mu)$. Using condition (C1), it follows that $\mathrm{E}(b|\mathcal{X}_1^\infty) = b$ a.s.

Similarly, let $a \in L^q(A, \mathcal{A}, \mu)$. Taking into account that $\mathcal{A} \perp\!\!\!\perp \mathcal{X}_1^n | \mathcal{B}_{\min}$ for all $n \in \mathbb{N}$, it follows that

$$\begin{aligned} \mathrm{E}(a|\mathcal{X}_1^n) &= \mathrm{E}[\mathrm{E}(a|\mathcal{X}_1^n \vee \mathcal{B}_{\min})|\mathcal{X}_1^n] \\ &= \mathrm{E}[\mathrm{E}(a|\mathcal{B}_{\min})|\mathcal{X}_1^n] \\ &\underset{n\to\infty}{\to} \mathrm{E}[\mathrm{E}(a|\mathcal{B}_{\min})|\mathcal{X}_1^\infty] = \mathrm{E}(a|\mathcal{B}_{\min}) \quad \text{a.s. and in } L^q, \end{aligned}$$

since $\mathrm{E}(a|\mathcal{B}_{\min})$ is measurable w.r.t. the minimal sufficient parameter $\mathcal{B}_{\min}$.

Acknowledgements

The first author acknowledges partial financial support from PUENTE Grant 08/2009, Pontificia Universidad Católica de Chile. The authors are grateful to the Editors, an anonymous referee, and Guido del Pino for their valuable comments.

References

Barra, J.R., 1981. Mathematical Basis of Statistics. Academic Press, New York.

Basu, D., 1975. Statistical information and likelihood (with discussion). Sankhyā, 37, 1-71.

Blackwell, D., 1986. On a class of probability spaces. Proceeding of the Third Symposium of Mathematical Statistics and Probability, 2, 1-6.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Caillot, P., Martin, F., 1972. Le modèle Bayésien. Annales de l'Institut Henri Poincaré - Section B, 8, 183-210.

Chen, M.H., Dey, D.K., 1998. Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. Sankhyā, 60, 322-343.

Chow, Y.S., Teicher, H., 1988. Probability Theory. Independence, Interchangeability, Martingales. Second Edition. Springer, New York.

Dawid, A.P., 1979. Conditional independence in statistical theory (with discussion). Journal of The Royal Statistical Society Series B—Statistical Methodology, 41, 1-3.

Dellacherie, C., Meyer, P.A., 1975. Probabilité et Potentiel I. Heman, Paris.

Döhler, R., 1980. On the conditional independence of random variables. Theory of Probability and Applications, 25, 628-634.

Eberly, L., Carlin, B., 2000. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Statistics in Medicine, 19, 2279-2294.

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A, 222, 309-368.

Florens, J.-P., Mouchart, M., 1977. Reduction of Bayesian experiments. CORE Discussion Paper 7737. Université Catholique de Louvain, Belgium.

Florens, J.P., Mouchart, M., 1982. A note on non causality. Econometrica, 50, 583-591.

Florens, J.P., Mouchart, M., Rolin, J.M., 1985. On two definitions of identification. Statistics, 16, 213-218.

Florens, J.P., Rolin, J.M., 1984. Asymptotic sufficiency and exact estimability. In Florens, J.P., Mouchart, M., Raoult, J.P., Simar, L., (eds.). Alternative Approaches to Time Series Analysis. Publications des Facultés Universitaires Saint-Louis, Bruxelles, pp. 121-142.

Florens, J.P., Mouchart, M., Rolin, J.M., 1990. Elements of Bayesian Statistics. Marcel Dekker, New York.

Gelfand, A.E., Sahu, S.K., 1999. Identifiability, improper priors, and gibbs sampling for generalized linear models. Journal of the American Statistical Association, 94, 247-253.

Gourieroux, C., Monfort, A., 1995. Statistics and Econometric Models. University Press, Cambridge.

Ghosh, M., Ghosh, A., Chen, M.H., Agresti, A., 2000. Noninformative priors for one-parameter item response models. Journal of Statistical Planning and Inference, 88, 99-115.

Hurwicz, L., 1950. Generalization of the Concept of Identification. Statistical Inference in Dynamic Economic Models, Cowles Commision Research Monograph 11. 238-257. Wiley, New York.

Jiang, T.J., Dickey, J.M., 2008. Bayesian methods for categorical data under informative censoring. Bayesian Analysis, 3, 541-554.

Jöreskog, K.G., 1981. Analysis of covariances structures (with discussion). Scandinavian Journal of Statistics, 8, 65-92.

Kadane, J., 1975. The Role of Identification in Bayesian Theory. In Fienberg S.E., Zellner A., (eds.). Studies in Bayesian Econometrics and Statistics. North-Holland, Amsterdam, pp. 175-191.

Kass, R., Carlin, B., Gelman, A., Neal, R., 1998. Markov chain Monte Carlo in practice: A roundtable discussion. The American Statistician, 52, 93-100.

Koopmans, T.C., Reiersøl, O., 1950. The identification of structural characteristics. The Annals of Statistics, 21, 165-191.

Lindley, D.V., 1971. Bayesian Statistics: A Review. Society for Industrial and Applied Mathematics, Philadelphia.

Macci, C., Polettini, S., 2001. Bayes factor for non-dominated statistical models. Statistics and Probability Letters, 53, 79-89.

Manski, C.F., 1995. Identification Problems in the Social Sciences. University Press, Harvard.

Manski, C.F., 2007. Identification for Prediction and Decision. University Press, Harvard.

Martín, F., Petit, J.L., Littaye, M., 1973. Indépendance conditionelle dans le modèle statistique bayésien. Annales de l'Institut Henri Poincaré - Section B, 9, 19-40.

McHugh, R.B., 1956. Efficient estimation and local identification in latent class analysis. Psychometrika, 21, 331-347.

McKean, H.P.Jr., 1963. Brownian motion with a several-dimensional time. Theory of Probability and Applications, 8, 335-354.

McCullagh, P., 2002. What is a statistical model? (with discussion). The Annals of Statistics, 30, 1225-1310.

Mouchart, M., 1976. A note on Bayes theorem. Statistica, 36, 349-357.

Mouchart, M., Rolin, J.M., 1984. A note on conditional independence. Statistica, 44, 557-584.

Mouchart, M., Rolin, J.M., 1985. Letter to the Editor. Statistica, 45, 427-430.

Maruyama, G.M., 1998. Basic of Structural Equation Modeling. Sage Publications, Thousand Oaks.

Neath, A.A., Samaniego, F.J., 1997. On the efficacy of Bayesian inference for nonidentifiable models. The American Statistician, 51, 225-232.

Paulino, C.D., Pereira, C., 1994. On identifiability of parametric models. Journal of the Italian Statistical Society, 3, 125-151.

Paulino, C.D., Soares, P., Neuhaus, J., 2003. Binomial regression with misclassification. Biometrics, 59, 670-675.

Pitcher, T.S., 1957. Sets of measures not admitting necessary and sufficient statistics or subfield. The Annals of Mathematical Statistics, 23, 267-268.

Pitcher, T.S., 1965. A more general property than domination for sets of probability measures. Pacific Journal of Mathematics, 15, 597-611.

Poirier, D.J., 1998. Revising beliefs in nonidentified models. Econometric Theory, 14, 483-509.

Poleto, F., Paulino, C.D., Molenberghs, G., Singer, J., 2010. Inferential implications of over-parametrization: a case study in incomplete categorical data. Technical report RT-MAE 2010-04. Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil.

Prakasa Rao, B.L.S., 1992. Identifiability in Stochastic Models. Characterizations of Probability Distributions. Academic Press, Boston.

Rao, M.M., 1984. Probability Theory with Applications. Academic Press, New York.

Rao, M.M., 1993. Conditional Measures and Applications. Marcel Dekker, New York.

San Martín, E., 2003. Modeling problems motivated by the specification of latent linear structures. Journal of Mathematical Psychology, 47, 572-579.

San Martín, E., Quintana, F., 2002. Consistency and identifiability revisited. Brazilian Journal of Probability and Statistics, 16, 99-106.

San Martín, E., Mouchart, M., 2007. On joit completeness: sampling and Bayesian versions and their connections. Sankhyā, 69, 780-807.

San Martín, E., González, J., Tuerlinckx, F., 2009. Identified parameters, parameters of interest and their relationships. Measurement: Interdisciplinary Research and Perspective, 7, 97-105.

Swartz, T., Haitovsky, Y., Vexler, A., Yang, T., 2004. Bayesian identifiability and misclassification in multinomial data. The Canadian Journal of Statistics, 32, 1-18.

Van Putten, C., Van Schuppen, J.H., 1985. Invariance properties of the conditional independence relation. The Annals of Statistics, 13, 934-945.

Verbekee, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer, New York.