T E C H N I C A L

R E P O R T

0250

ON THE IDENTIFIABILITY AND ESTIMABILITY OF LATENT CLASS MODELS: A BAYESIAN ANALYSIS

SAN MARTIN, E. and P. DE BOECK

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

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

Chapter 1

On the Identifiability and Estimability of Latent Class Models: A Bayesian Analysis

Ernesto San Martín[1] and Paul De Boeck[2]

1.1 Introduction

Latent Class Models are introduced by Lazarsfeld; see, for example, Lazarsfeld and Henry (1968). These modes assume that the population, from which the observed sample is taken, is composed of m mutually exclusive latent classes $C_1, \ldots, C_i, \ldots, C_m$. The parameters of interest here are the probabilities with which a randomly chosen subject belongs to each of the latent classes. For each person, l dichotomous measurements are made; let these be the item scores (incorrect, correct), or the item reactions (negative, positive). In LCM it is assumed that, for each item, every class C_i has a specific probability of positive responses: these are the parameters of interest in the conditional model. For details, see, e.g. Clogg (1995).

To ensure a coherent inference on the structural parameters, their identifiability is needed. Different solutions to this problem can be found in the

[1]Department of Statistics, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile.

[2]Department of Psychology, K. U. Leuven, Tiensestraat 102, B-3000 Leuven, Belgium.

psychometric literature. A first solution can be found in the early works of Andersen (1954), McHugh (1956) and Madansky (1960). Essentially these contributions provide algebraic methods to deduce the proportions of the population in each latent class and the probabilities of positive responses to each item from the proportions of response parameters in the population as a whole. These methods are also applied to estimate the parameters of interest.

Another group of contributions is represented by Goodman (1974) and Clogg and Goodman (1984). The main concern of these papers is to obtain parameter estimations from the likelihood function. These authors establish the identifiability of the parameters of interest from a local identification point of view. Broadly speaking, it is a matter of inverting the likelihood function in a neighborhood of the "true" parameter and thus to solve the likelihood equations.

The two approaches are related since both are based on linear equation systems defined on the marginal probabilities of response patterns of the individuals from the population as a whole.

A third approach to analyze this identification problem is motivated by the hierarchical structure underlying LCMs. As a relevant example, let us consider the identification analysis proposed by Maris (1999) in the context of Multiple Classification Latent Class Models (MCLCM). MCLCM consists of two component submodels: (a) a model for the latent class memberships and (b) a model for the item responses conditional on the latent class memberships. Maris considers two type of non-identifiability, one with respect to the conditional model, the other one with respect to the statistical model (obtained after averaging on the latent class memberships). In a particular case, Maris obtains identification restrictions to identify the conditional model. Furthermore, these restrictions together supplementary identification restrictions are used to identify the statistical model. It should be noticed that, in this heuristic argument, the marginal model generating the latent class memberships does not play any essential role.

This paper is motivated by the last approach. At the identification level, we establish that, under prior independence of the parameters of interest, the identification of the marginal model is a *necessary* condition for the identification of the statistical model. A second result we explore in this paper deals with the consistency of a Bayesian estimator of the probabilities of correct responses given the latent class, denoted as P. To establish this result we use the statistical tools developed by Florens, Mouchart and Rolin (1990) in the context of Bayesian models. We use the estimability theory developed by these authors in order to show that the conditional expectation of the P given the observations and the latent class probabilities converges to P. A similar result is established for the probabilities of latent classes.

This chapter is organized as follows: model specification is detailed in section 1.2. Next, in section 1.3, the identifiability of the marginal model generating the latent class memberships is explained to be a necessary condition for the identifiability of the statistical model. Finally, in section 1.4 we establish some results dealing with the estimability.

1.2 Specification of the LCM

Following Novick (1979), an efficient and interpretable tool for structural modelling is the conditional independence. As a matter of fact, in the context of psychometrics or econometrics the substantive theory leads to formalize dependencies between measurements and constructs. Recent works in graphical models exemplify these statement (Wermuth and Lauritzen, 1990). Moreover, taking conditional independence as basic, a unification of many statistical concepts is achieved. Thus, not only concepts as sufficient and consistent statistics can be expressed in terms of conditional independence, but also parametric identifiability and estimability. Since conditional independence is expressed with respect to a probability distribution defined on the product space "parameters $\times$ observations", the Bayesian approach becomes adequate to develop structural models and their respective statistical analysis. For a textbook discussion of conditional independence, see Florens et al. (1990, chapter 2).

Before reviewing the standard specification of LCM, let us introduce the concept of *parametric sufficiency*. The relevance of this concept not only deals with the statistical interpretability of the model, but also with parameter identifiability and its Bayesian link with estimability (or Bayesian consistency).

1.2.1 Sufficiency on the parameter space

A characteristic feature of the Bayesian approach is to define a statistical model as a unique probability measure on the product space "parameters $\times$ observations", say $Q_{(X,\theta)}$, where X represents the observations and θ the parameters. In this context, parameters and observations have a symmetric role in the sense that $Q_{(X,\theta)} = Q_{(\theta,X)}$ and, therefore, X can be considered as a "parameter" and θ as an "observation"; for details, see Mouchart (1976) and Florens et al. (1990, chapter 1).

Let us consider a Bayesian experiment given characterized by a probability Q. According to the sampling definition of sufficiency, a statistic S

is sufficient if the conditional distribution of X given $S(X)$ does not depend on θ. When θ is a random variable such a definition is equivalent to $X \perp\!\!\!\perp \theta \mid S(X)$. Taking advantage of the symmetric role of parameters and observations, the definition of a sufficient parameter is similar to that of a sufficient statistic:

Definition 1.1 *A function $g(\theta)$ of the parameter is said to be sufficient in the Bayesian sense if the conditional distribution of X given θ is equal to that of X given $g(\theta)$, that is,*

$$X \perp\!\!\!\perp \theta \mid g(\theta). \tag{1.1}$$

Condition (1.1) says that the distribution of X is completely determined by $g(\theta)$, θ being redundant once $g(\theta)$ is known. By the symmetry of a conditional independence relation, it can also be concluded that $g(\theta)$ is a sufficient parameter if the conditional distribution of the redundant part θ, given the sufficient parameter $g(\theta)$, is the same in the posterior distribution as in the prior distribution, namely $p(\theta \mid X, g(\theta)) = p(\theta \mid g(\theta))$. Thus, once we have learned about $g(\theta)$ from the data, we can learn nothing more about θ, over and above what we knew already; see Dawid (1979) and Florens et al. (1990, chapter 2).

1.2.2 Model specification

Let us consider n persons indexed by k who belong to m mutually exclusive classes indexed by i. These persons are asked to answer l items indexed by j. Let $X_k = (X_{k1}, \ldots, X_{kj}, \ldots, X_{kl})' \in \{0,1\}^l$ be a discrete random vector representing the responses of a person k to the l items. Let $\eta_k \in \{0,1\}^m$ be a random vector such that $\sum_{i=1}^m \eta_{ki} = 1$. Thus, the possible values of η_k are $\{e_1, \ldots, e_i, \ldots, e_m\}$, where e_i is a vector containing a 1 in the i-th coordinate, whereas the other ones are equal to 0. The random vector η_k represents the class membership of person k to one of the m mutually exclusive classes. Let $\pi \in \mathcal{S}_{m-1}$, where $\mathcal{S}_{m-1} = \{r \in [0,1]^m : \sum_{i=1}^m r_i = 1\}$. Let $P = (p_{ij})$ be a $m \times l$ matrix, where $p_{ij} = \mathbb{P}[X_{kj} = 1 \mid \eta_k = e_i]$ for all persons k. The rows of matrix P are denoted by $P_{(i)}$, with $i = 1, \ldots, m$, whereas the columns of P are denoted by $P^{(j)}$, with $j = 1, \ldots, l$. Thus, all the parameters of interest are denoted as $\theta = (\pi, P)$. Finally, we denote X_1^n the matrix $(X_1, X_2, \ldots, X_n)$. Similarly, η_1^n denotes the matrix $(\eta_1, \ldots, \eta_n)$.

The hypothesis necessary to specify the process generating $(X_1^n, \eta_1^n, P, \pi)$ are the following:

- C1. $\eta_1^n \perp\!\!\!\perp \theta \mid \pi \quad \forall\, n \geq 1.$

This condition means that the process generating $(\eta_1^n \mid \pi, P)$, also called the *marginal model*, depends on π only. This condition defines, therefore, π as a sufficient parameter for η_1^n.

• C2. $X_1^n \perp\!\!\!\perp \theta \mid \eta_1^n, P \quad \forall\, n \geq 1$
This condition means that the process generating $(X_1^n \mid \eta_1^n, P, \pi)$, also called the *conditional model*, depends on (η_1^n, P) only. In other words, P is a sufficient parameter for the conditional process generating X_1^n given η_1^n.

In most applications, the persons to be asked to answer a test are randomly chosen. It seems reasonable to assume that persons' class memberships η_k's are mutually independent. More precisely,

• C3. $\underset{1\leq k\leq n}{\perp\!\!\!\perp} \eta_k \mid \pi \qquad \forall\, n \geq 1.$

• C4. For all $n \geq 1$

$$\text{(i)} \underset{1\leq k\leq n}{\perp\!\!\!\perp} X_k \mid \eta_1^n, P, \qquad \text{(ii)}\ X_k \perp\!\!\!\perp \eta_1^n \mid \eta_k, P \quad \forall\, k = 1, \ldots, n.$$

Property (i) means that the X_k's are mutually independent conditionally on (η_1^n, P). In other words, this hypothesis means that, conditionally on the class memberships of a sample of n persons, their responses are not mutually "contaminated". Note that this hypothesis does not deal with the choice of the persons since it is a condition *given their membership class*, that is, once the n persons were chosen.
Property (ii) means that the process generating $(X_k \mid \eta_1^n, P)$ depends on (η_k, P) only and not on any other $\eta_{k'}$ with $k \neq k'$. In other words, property (ii) says that the measurement X_k is explained by the class membership η_k only, and not by another one $\eta_{k'}$ with $k \neq k'$.

Applying successively properties (i) and (ii), condition C4 implies the following decomposition:

$$p(X_1^n \mid \eta_1^n, P) = \prod_{k=1}^{n} p(X_k \mid \eta_1^n, P) = \prod_{k=1}^{n} p(X_k \mid \eta_k, P) \qquad \forall\, n \geq 1.$$

• C5. For all $k = 1, \ldots, n$ and for all $n \geq 1$

$$\text{(i)} \underset{1\leq j\leq l}{\perp\!\!\!\perp} X_{kj} \mid \eta_k, P, \qquad \text{(ii)}\ X_{kj} \perp\!\!\!\perp P \mid \eta_k, P^{(j)} \quad \forall\, j = 1, \ldots, l.$$

Note that this condition can be interpreted as a Bayesian version of local independence. In particular, property (ii) says that $P^{(j)}$ is a sufficient parameter for X_{kj} given η_k. Applying successively properties (i) and (ii), it follows that

$$p(X_k \mid \eta_k, P) = \prod_{j=1}^{l} p(X_{kj} \mid \eta_k, P) = \prod_{j=1}^{l} p(X_{kj} \mid \eta_k, P^{(j)}).$$

Combining conditions C3 and C5, we obtain that

$$p(X_1^n \mid \eta_1^n, P) = \prod_{k=1}^{n} \prod_{j=1}^{l} p(X_{kj} \mid \eta_k, P^{(j)}).$$

This is the typical expression of the likelihood in the Latent Structural Model literature; see, e.g., Andersen (1980).

• C6. $(\eta_k \mid \pi) \sim \mathrm{Bern}_{(m)}(\pi) \quad \forall\, k = 1, \ldots, n \quad \forall\, n \geq 1$, where $\mathrm{Bern}_{(m)}$ denotes a m-dimensional multi-Bernoulli.
Thus, the probability that a person k belongs to the latent class c is given by $P[\eta_k = e_c \mid \pi] = \prod_{i=1}^{m} \pi_i^{e_c} = \pi_c$ for all $c = 1, \ldots, m$. Note also that conditions C3 and C6 can equivalently be rewritten as $(\eta_k \mid \pi) \sim$ iid.$\mathrm{Bern}_m(\pi)$.

• C7. $(X_{kj} \mid \eta_k, P^{(j)}) \sim \mathrm{Bern}\,(\eta_k' P^{(j)})\, \forall\, j = 1, \ldots, m\, \forall\, k = 1, \ldots, n\, \forall\, n \geq 1$.
This condition makes explicit that the realization of an observable random variable depends on an *unobservable* marginal Bernoulli process. Thus, for instance, $(X_{kj} \mid P, \eta_k = e_i) \sim \mathrm{Bern}\,(p_{ij})$.

• C8. $\pi \perp\!\!\!\perp P$.
Bayesian specification of LCM needs to be completed with a prior distribution. We assume that the distribution on (π, P) is the product of the prior distributions on π and on P; other aspects of the prior distributions *remain unspecified* in the present specification. There may exist situations in which this hypothesis is not necessarily realistic, but some identification results we present in this paper will depend on this condition.

1.3 A necessary identification relationship

Before analyzing some identification problems in LCM, a Bayesian definition of identification will be discussed.

1.3.1 Minimal sufficiency and identifiability

In the context of a Bayesian experiment, consider a sufficient parameter $\phi = g(\theta)$, namely $X \perp\!\!\!\perp \theta \mid \phi$, that is, the data X do not increase our prior knowledge about θ given ϕ. It follows that a part of the prior information is not revised by the observations. Therefore, the parametrization θ is not "identified" by the data X; see, e.g., Dawid (1979), Florens and Mouchart (1986), Poirier (1998) and Gelfand and Sahu (1999).

This situation can be avoided if the parameter ϕ is a *minimal sufficient parameter*, that is, if ϕ is a sufficient parameter and it is a function of any other sufficient parameter. In such a case, there does not exist a function of ϕ, conditionally on which the prior and the posterior distributions are the same. It follows that if the parametrization of a statistical model is minimally sufficient, then the parametrization does not contain redundant information. These considerations motivate the following definition (Florens and Mouchart, 1984):

Definition 1.2 *A parameter $\theta^* = h(\theta)$ is said to be Bayesian identified (or, b-identified) if θ^* is a minimal sufficient parameter.*

It can be shown that a minimal sufficient parameter can equivalently be represented as a function of countably many sampling expectations of statistics $\mathbb{E}\,[g(X) \mid \theta]$; see Florens et al. (1990, chapter 4.3). In the case of discrete random variables, the verification that a parameter, say θ, is b-identified by an observation X reduces to express θ as a function of some sampling expectation of X. This type of argument will be used in the rest of this paper.

Remark 1 In a classical set-up, a statistical experiment is defined as a family of sampling distributions P_X^θ indexed by a parameter $\theta \in \Theta$, where Θ denotes the parameter space; see, e.g., Barra (1981). The parametrization θ is said to be identified by the observation if for two different parametrizations θ_1 and θ_2, there are two different probability distributions $P_X^{\theta_1}$ and $P_X^{\theta_2}$, that is, if the mapping $\theta \longmapsto P_X^\theta$ is injective; see, e.g., Koopmans and Reirsøl (1950). This is the standard identification concept typically used in psychometrics and econometrics. Under some technical conditions, it can be established that identification implies b-identification for all prior distribution μ defined on Θ; for details, see Florens et al. (1985) and Florens et al. (1990, chapter 4).

1.3.2 A Minimal sufficient parameter in the statistical model

In LCM, the statistical model bearing on the observable variables only is given by

$$\mathbb{P}[X_1^n = x_1^n \mid \pi, P] \;=\; \prod_{k=1}^{n} \left\{ \sum_{i=1}^{m} \pi_i \prod_{j=1}^{l} p_{ij}^{x_{kj}} \,(1-p_{ij})^{1-x_{kj}} \right\}, \tag{1.2}$$

where $x_k \in \{0,1\}^l$ for all $k = 1,\ldots,n$. The mutual independence of the X_k's given (π, P) is implied by conditions C3 and C4. Since $X_k \in \{0,1\}^l$ and given that the X_k's are mutually independent conditionally on (π, P), the minimal sufficient parameter, and therefore the b-identified parameter of the statistical model, is given by

$$\mathbb{E}\,[X_k \mid \pi, P] \;=\; \sum_{i=1}^{m} e_i{}' P\, p_i \;=\; \begin{pmatrix} \sum_{i=1}^{m} p_{i1}\pi_i \\ \vdots \\ \sum_{i=1}^{m} p_{il}\pi_i \end{pmatrix} \in [0,1]^l. \tag{1.3}$$

These are actually the marginal expected values. It is clear that these minimally sufficient parameters are a function of (π, P) but not vice versa, and therefore such a function is not injective. The identification of (π, P) by the observations means to introduce additional restrictions in such a way that (π, P) becomes the minimal sufficient parameter and, consequently, that (π, P) becomes a function of (1.3) indeed. Since the right hand of (1.3) *does not depend on the index* k, the number of persons in the sample does not will provide any additional information to obtain such an injective function. Finally note that the classical identification problem, namely the injectivity of the mapping $(\pi, P) \longmapsto \mathbb{P}(\cdot \mid \pi, P)$, where $\mathbb{P}(\cdot \mid \pi, P)$ is given by (1.2), is exactly the same. It should be said that, in spite of the simplicity of the LCM, the identifiability of (π, P) is a difficult problem.

When we don't know if a model is identified, a possible approach is to investigate what can be implied if we *assume* that the model is identified? The implied, typically called *necessary identification restrictions*, are then imposed on the model. In what follows we establish a condition under which the identifiability of the marginal model generating η_1^n is a necessary condition for the identifiability of the statistical model generating X_1^n only.

1.3.3 Identification relationship between the marginal and the statistical models

The following theorem is not only true for the LCM, but also for all Latent Structural Model characterized by conditions C1, C2 and C8.

Theorem 1.1 *The b-identification of the marginal model* $(\eta_1^n \mid \pi)$ *is a necessary condition for the b-identification of the statistical model* $(X_1^n \mid \pi, P)$.

Proof: Conditions C1 and C8 jointly imply that $P \perp\!\!\!\perp \pi \mid \eta_1^n$. This last condition along with C2 imply that

$$\text{(i) } X_1^n \perp\!\!\!\perp_r \pi \mid \eta_1^n, \qquad \text{(ii) } P \perp\!\!\!\perp \pi \mid X_1^n, \eta_1^n. \tag{1.4}$$

Assume that the statistical model is identified, that is, (π, P) is a minimal sufficient parameter for X_1^n. It follows that (π, P) is also minimal sufficient for (X_1^n, η_1^n). By the basic properties of the b-identifiability (see Florens et al., 1990, Proposition 4.5.2), it follows that π is b-identified by (X_1^n, η_1^n, P), and therefore also by η_1^n given (1.4.i) and (1.4.ii).

□

Given the independence between of π and P, i.e. condition C8, it is clear that the statistical model is not identified if the marginal model is not identified. This shows the role of condition C8 under which the identification of the marginal model becomes a *necessary* condition for the identification of the statistical model.

In the classical LCM, the marginal model is trivially identified. As a matter of fact, under condition C6, $\pi = \mathbb{E}\,[\eta_k \mid \pi]$ for all $k \geq 1$. It follows that π is b-identified by η_k (see the comments immediately after Definition 1.2). Therefore, *under condition C8*, the identified marginal model is a necessary condition for the identifiability of the statistical model.

Finally, let us say that Theorem 1.1 is relevant for other Latent Structural Models in the sense that the specification of the corresponding marginal model generating the η_k's should be made in such a way that *it be identified.* As a simple example, let us consider a LCM in which P and π are constrained by means of linear restrictions; see Forman (1985). In this case, the marginal parameter π is specified as $\pi = V\lambda + d$, where V is a $m \times t$ matrix and λ is the parameter of interest. In this case, the parameter λ is b-identified by η_k if $r(V) = t$. Thus, under this restriction and C8, Theorem 1.1 applies.

1.4 Exact Estimability of P

The concern of this section is to study the Bayesian consistency (or estimability) of P given the data, as an approach to identifiability of P. Let us start by reviewing some concepts about estimability.

1.4.1 Estimability and convergence of posterior expectations

Let θ be a parameter to be estimated. Since in a Bayesian set-up the main concern deals with the "learning by observing" process, the posterior distribution of θ given the data is of interest. Moreover, the information contained in the posterior distribution can be uncovered from the posterior expectation of *all* functions of the parameter given the data, say $\mathbb{E}\,[f(\theta) \mid X_1^n]$. Furthermore, the posterior expectation of $f(\theta)$ is the usual Bayesian estimator corresponding to a quadratic loss function. This posterior expectation *always* converges almost surely to $\mathbb{E}\,[f(\theta) \mid X_1^\infty]$, the posterior expectation of θ given the infinite sequence of observations, provided that $f(\theta)$ be an integrable function [this is due to the Martingale Theorem]. By analogy to sampling consistency, it seems natural to say that the Bayesian estimator $\mathbb{E}\,[f(\theta) \mid X_1^n]$ is consistent if its limit is a.s. equal to $f(\theta)$, that is, $\mathbb{E}\,[f(\theta) \mid X_1^\infty] = f(\theta)$ a.s. These facts motivate the following definition:

Definition 1.3 *Let f be any (measurable) function. A parameter $f(\theta)$ is said to be estimable if the posterior expectation $\mathbb{E}[f(\theta) \mid X_1^n]$ converges a.s. to $f(\theta)$.*

An alternative characterization of estimability can be obtained by using a basic property of conditional expectation, namely $\mathbb{E}\,[V \mid W] = V$ if and only if there exists a (measurable) function h such that $V = h(W)$. As a matter of fact, a parameter $f(\theta)$ is estimable if and only if $\mathbb{E}[f(\theta) \mid X_1^\infty] = f(\theta)$ a.s., which is equivalent to say that $f(\theta)$ is a.s. a function of X_1^∞. This means that the infinite sequence of observations X_1^∞ contains the relevant information necessary to *construct* estimable parameters. Note that this provides a Bayesian interpretation about the meaning of a parameter.

Finally, let us say that if there exists a statistic T_n such that it is a consistent estimator of θ (i.e. $T_n \longrightarrow \theta$ a.s.), then b is estimable for all prior distributions on θ. This link between sampling-consistency and Bayesian estimability helps to consider Definition 1.3 as "Bayesian consistency". For details and proofs, see Florens et al. (1990, section 7.4)

Definition 1.3 leads us to consider asymptotic arguments in which the so-called tail-statistics plays an essential role. Broadly speaking, the tail-statistic of a sequence of statistics, say $\{Z_n\}$, depends on the last coordinates Z_n but not on the first m ones for any finite m. For a formal definition and properties, see, e.g., Ellis and Junker (1997, Appendix). The motivation to introduce tail-statistics is that to prove estimability results we need the following general results: for a sequence of random objects Z_k iid given θ, it follows that Z_T is almost surely (a.s.) the minimal sufficient statistics, which is also a.s. an injective function of the corresponding minimal sufficient parameter; for a proof, see Theorem 9.3.12 in Florens et al. (1990). Moreover, if θ is b-identified by Z_k, then Z_T is a.s. an injective function of θ.

1.4.2 Main results

Let us establish the following facts implied by the specification of LCM.

• F1. Conditions C3 and C6 imply that $\{\eta_k\}$ is an iid sequence given π. It follows that π is a.s. a function of the tail-latent variable η_T [apply Theorem 9.3.12 of Florens et al., 1990]

• F2. Conditions C1, C2, C3 and C4 jointly imply that

$$\perp\!\!\!\perp_{1 \leq k \leq n} (X_k, \eta_k) \mid \pi, P \qquad \forall\, n \geq 1. \tag{1.5}$$

Moreover, $\mathbb{E}\,[(X_k, \eta_k) \mid \pi, P] = \mathbb{E}\,[(X_1, \eta_1) \mid \pi, P]$ for all $k \in \mathbb{N}$. Therefore, $\{(X_k, \eta_k)\}$ is an iid sequence given (π, P).

• F3. P is b-identified by X_1^∞ conditionally on η_1^∞, where X_1^∞ denotes the infinite matrix (sequence) $(X_1, X_2, \ldots, X_k, \ldots)$, and η_1^∞ the infinite matrix (sequence) $(\eta_1, \eta_2, \ldots, \eta_k, \ldots)$.

Proof of Fact 3: condition C7 implies that $\mathbb{E}\,[X_{kj} \mid P^{(j)}, \eta_k = e_i] = p_{ij}$ for all $i = 1, \ldots, m$ and for all $j = 1, \ldots, m$. It follows that $\eta_k{}'P^{(j)}$ is b-identified by X_{kj} conditionally on η_k for all (j, k). Using condition C5, this last identification relationship implies, by Theorem 2 of Mouchart and San Martín (2003), that $\eta_k{}'P = (\eta_k{}'P^{(1)}, \ldots, \eta_k{}'P^{(j)}, \ldots, \eta_k{}'P^{(l)})$ is b-identified by X_k conditionally on η_k for all $k \geq 1$. Similarly, this last identification restriction along with condition C4 implies that $(\eta_1{}'P, \ldots, \eta_n{}'P)$ is b-identified by X_1^n conditionally on η_1^n for all $n \geq 1$ (apply Theorem 2 of Mouchart and San Martín, 2003).
Fact 3 follows if we prove that there exists an injective mapping between $(\eta_1{}'P, \ldots, \eta_n{}'P, \eta_1^n)$ and (P, η_1^n). It is clear that if there exists a full rank $m \times m$ matrix (with probability one), say $\tilde{\eta}$, then the desired injectivity can be established. As a matter of fact, the probability that, in a sample of n persons, there exists at least one latent class to which any person belongs is equal to $\sum_{i=1}^{m} [1 - \pi_i]^n$. Since this probability converges to 0 as $n \to \infty$, it follows that, in an infinite population of persons, we have, with probability one, at least one person from each latent class. Let us mention that this asymptotic argument leads to state that P is b-identified by X_1^∞ given η_1^∞.
□

The b-identifiability of the conditional model seems an immediate consequence. However, as can be viewed in the proof above, the η_k's are *not observable*. The identifiability of P by the observations is obtained after evaluating the conditional probability in C7 on the realizations of the η_k's. The identification follows after noticing that, in an infinite sampling process, all the latent classes are visited at least one time *with probability one*.

These facts are sufficient to proof the main result of this paper, namely

Theorem 1.2 *Assume conditions C1, ..., C7. It follows that*

$$\mathbb{E}\left[f(P) \mid X_1^n, \pi\right] \longrightarrow f(P) \quad a.s. \qquad \text{for all integrable function } f. \tag{1.6}$$

This theorem deals with the asymptotic behavior of the posterior expectation of P given the observations and π, the parameter of the marginal model. Let us sketch its proof:

• R1. Fact F2 implies, by Theorem 9.3.12 in Florens et al. (1990), that the minimal sufficient statistics for (π, P) is a.s. a function of the tail-statistics (X_T, η_T), and viceversa. Moreover, the same theorem implies that (X_T, η_T) is a.s. a function of the minimal sufficient parameter for $(X_1^\infty, \eta_1^\infty)$, and viceversa. It is important to note that corresponding minimal sufficient parameter is a function of (π, P).

• R2. Since η_T is, by definition, a function of η_1^∞, fact F1 implies that π is a.s. a function of η_1^∞.

• R3. Conclusion R2 along with fact F3 implies that the minimal sufficient parameter for $(X_1^\infty, \eta_1^\infty)$ is (π, P) (apply Proposition 4.6.6. of Florens et al., 1990). Note that this relationship means that (π, P) is identified by the complete process $(\eta_1^\infty, X_1^\infty)$; this identifiability is a *necessary* condition for the identifiability of the statistical model. Condition C8 is not used in this case.

• R4. Conclusions R1 and R3 jointly imply that the minimal sufficient statistics (X_T, η_T) is a.s. a function of (π, P), and vice versa. Moreover, under fact F1, it follows that (X_T, π) is a.s. a function of (π, P).

• R5. Finally, since (X_T, π) is a function of (X_1^∞, π), it follows that P is a function of (X_1^∞, π). This last conclusion leads, by Definition 1.3, to statement (1.6).

Theorem 1.2 essentially depends on the actual specification of a LCM, which leaves unspecified the prior distribution on (π, P). Therefore, it is robust with respect to an arbitrary choices of the prior. Therefore, Theorem 1.2 suggests an algorithm to estimate P, namely to take a fixed realization of π and then to compute the conditional expectation of P given X_1^n. This step can be repeated to compare different Bayesian estimators of P. The behavior of these estimators, but this issue is outside the scope of the present paper.

1.4.3 A secondary result on π

It seems interesting to add some secondary results. As a matter of fact, conditions C3 and C4 imply that X_k's are mutually independent given (π, P). Moreover, by (1.3), it follows that the X_k's are iid given (π, P). Therefore, the tail-statistic X_T is a.s. equal to the minimal sufficient parameter of the statistical model given by (1.3). Note that it is equal to $\pi' P$. It follows that (a) (X_T, P) is a.s. a function of $(\pi' P, P)$, and vice versa. But (b) π is identified by X_1 given P if $r(P) = m$ (with probability 1) since $\pi' = \pi'\, PP^{-1}$. Let us mention that $\mathbb{P}[r(P) = m] = 1$ because the elements of the matrix P are belong to $[0, 1]$; hence $m \leq l$. From (a) and (b) it can be concluded that (X_T, P) is a.s. a function of (π, P), and viceversa. We establish, therefore, the following Proposition:

Proposition 1.1 *Assume condition C1, ..., C7. It follows that*

$$\mathbb{E}\,[h(\pi) \mid X_1^\infty, P] \longrightarrow h(\pi) \qquad \textit{for all integrable function } h.$$

Acknowledgement: This manuscript reflects a bilateral collaboration between the Department of Statistics, Pontificia Universidad de Chile, and the Department of Psychology, K. U. Leuven, Belgium, in the context of the bilateral project *Nonlinear Mixed Models for Educational Measurement and Standard Setting* sponsored by the BIL01/01 grant and ..... The first author acknowledges the financial support of the FONDECYT Project N$^{\mathrm{O}}$ 3010069, Corporación Nacional de Ciencia y Tecnología, Chile. The authors have a particular debt to M. Mouchart and J. -M. Rolin for many discussions during the elaboration of this paper.

References

Anderson, T. W. (1954), On Estimation of Parameters in Latent Structure Analysis. *Psychometrika* **19**, 1–10.

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

Clogg, C. C. (1995), Latent Class Models. In: G. Arminger, C. C. Clogg and M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, chapter 6. Plenum Press, New York.

Clogg, C. C. and Goodman, L. A. (1984), Latent Structure Analysis of a Set of Multidimensional Contingency Tables. *Journal of the American Statistical Association* **79**, 762–771.

Dawid, A. P. (1979), Conditional Independence in Statistical Theory (with Discussion). *Journal of the Royal Statistical Society, Series B* **41**, 1–31.

Ellis, J. L. and Junker, B. W. (1997). Tail-Measurability in Monotone Latent Variable Models. *Psychometrika* **62**, 495–523.

Evans, M. J., Gilula, Z. and Guttman, I. (1989), Latent class analysis of two-way contingency tables by Bayesian methods. *Biometrika* **76**, 557–563.

Florens, J. -P. and Rolin, J. -M. (1984), Asymptotic sufficiency and exact estimability in Bayesian experiment. In: J. -P. Florens, M. Mouchart, J. -P. Raoult and L. Simar (eds.), *Alternatives Approaches to Time Series Analysis.* Publication des Facultés Universitaires Saint-Louis, Bruxelles.

Florens, J. -P. and Mouchart, M. (1986), Exhaustivité, ancilarité et identification en statistique bayésienne. *Annles d'économie et de statistique* **4**, 63–93.

Florens, J. -P., Mouchart, M. and Rolin, J. -M. (1985), On Two Definitions of Identification. *Statistics* **16**, 213–218.

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

Formann, A. K. (1985), Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology* **38**, 87–111.

Gelfand, A. E. and Sahu, S. K. (1999), Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association* **94**, 247–253.

Goodman, L. A. (1974), Exploratory latent structural analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.

Koopmans, T. C. and Reiersøl, O. (1950), The Identification of Structural Characteristics. *The Annals of Mathematical Statistics* **21**, 165–181.

Lazarsfeld, P. F. and Henry, N. W. (1968), *Latent structure analysis.* Houghton Mifflin Company, Boston.

Madansky, A. (1960), Determinantal Methods in Latent Class Analysis. *Psychometrika* **25**, 183–198.

Maris, E. (1999), Estimating Multiple Classification Latent Class Models. *Psychometrika* **64**, 187–212.

McHugh, R. B. (1956), Efficient Estimation and Local Identification in Latent Class Models. *Psychometrika* **21**, 331–347.

Mouchart, M. (1976), A note on Bayes theorem. *Statistica* **36**, 349–357.

Mouchart, M. and San Martín, E. (2003), Specification and identification issues in models involving a latent hierarchical structure. *Journal of Statistical Planning and Inference* **111**, 143–163.

Novick, M. R. (1979), Comments on "Conditional Independence in Statistical Theory". *Journal of the Royal Statistical Society, Series B* **41**, 27.

Poirier, D. J. (1998), Revising beliefs in nonidentified models. *Econometric Theory* **14**, 482–509.

San Martín, E. (2000), *Latent Structural Models: Specification and Identification Problems.* Ph. D. Dissertation, Institute of Statistics, Université Catholique de Louvain, Belgium.

Wermuth, N. and Lauritzen, S. L. (1990), On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Models (with Discussion). *Journal of the Royal Statistical Society, Series B* **52**, 21–50.