Identification Problems in Mixture Models: Some General Results*

Michel Mouchart [a,b], Ernesto San Martín [b]

[a] *CORE and* [b] *Institut de statistique, Université catholique de Louvain,*

20 Voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium

Abstract

A structural model generating jointly a manifest variable y and a latent variable η is considered. The object of this note is to examine, on the one hand, the relationships between the identification of the structural model and the component submodels, namely the conditional model generating y given η and the marginal model generating η, and, on the other hand, the relationships between the identification of the statistical model, *i.e.* a marginal model generating the manifest variable, and the component submodels. These relationships are obtained by using two concepts of identification, a weak one and a strong one.

Keywords: Completeness; Cut; Latent variable; Marginal-conditional reduction; Minimal Sufficiency; Sampling identification; Strong and weak identification.

1 Introduction

EM estimators have been introduced as one possible way of handling a large class of statistical models, namely models for partially observable processes; see *e.g.* Goodman (1974) or Dempster *et al.* (1977). The basic idea is

*This research was supported by the contract *Project d'Actions de Recherche Concertées* ARC 98/03-217 of the Belgian Government.

that completing data y by a latent, or unobservable, component η leads to specify the statistical model in the form of a mixture of, say, $p(y \mid \eta, a)$ through $p(\eta \mid a)$. Thus the data density may be symbolically written as $p(y \mid a) = \mathbb{E}_a [p(y \mid \eta, a)]$, where $\mathbb{E}_a$ stands the expectation with respect to a specified probability measure on η. In many cases the dimension of η increases with the sample size, *i.e.* η has an incidental nature, but this will not be an issue in this note. The object of this note is rather to examine the relationships between the identification of the statistical model, viewed as a marginal model of the joint structural model generating y and η, and the component submodels, namely the conditional model generating y given η and the other marginal model generating η, a particular interest being to obtain sufficient conditions for the identification of the statistical model. Thus, the main objective of this note is to obtain general results for a wide class of models currently used in actual practice.

The results obtained in this note rely on two concepts of identification, a weak one and a strong one. The weak one corresponds to the standard concept of the injectivity of the mapping from the parameter space to the space of sampling distributions (in the sequel called *sampling identification*), and the strong one is an extension of the concept of completeness (*i.e.* of a complete statistics) which is a concept of injectivity of conditional expectation (in L^p-spaces). Thus both concepts are standard in the statistical litterature but, to the best of own knowledge, their relationships have not been analyzed extensively (a first step in this direction being however Florens *et al.*, 1990, chapter 5). A secondary objective of this note is therefore to develop some relationships between these two concepts of identification deemed to be useful for the identification analysis of a large class of mixture models. Before presenting these results, this note spends some space to illustrate these concepts in order to facilitate a proper understanding of the main results.

1.1 General Framework and Summary of the Main Results

From a formal point of view, the framework may be considered as Bayesian as far as we work on a unique probability measure involving all the variables: the observables and the non observables (parmeters and/or latent variables); thus, no formal difference between an incidental parameter or a latent variable is introduced. Note however that the objectives of this note are not of a computational nature, and consequently the specification of the distributions on the observables and on the unobservables does not go further than (simple) structural properties without relying on particular analytical forms. In particular, the results of this note are relevant for non-parametric, semi-parametric or purely parametric models.

Let us therfeore consider a *structural experiment* $\mathcal{E} = (A \times X, \mathcal{A} \vee \mathcal{X}, \Pi)$, where $(A, \mathcal{A})$, the *parameter space*, and $(X, \mathcal{X})$ are measurable spaces and Π is a unique probability measure defined on the σ-field $\mathcal{A} \vee \mathcal{X}$. By decomposing $\mathcal{X}$ into $\mathcal{X} = \mathcal{Y} \vee \mathcal{H}$ (with $\mathcal{Y} = \sigma(y)$ and $\mathcal{H} = \sigma(\eta)$), the structural experiment $\mathcal{E}$ is accordingly decomposed into its complementary pair of marginal and conditional reductions with respect to $\mathcal{H}$, that is (i) a marginal experiment is defined by a marginal probability measure $\Pi_{\mathcal{H}}$ obtained as the trace of Π on $\mathcal{A} \vee \mathcal{H}$, and (ii) a conditional experiment is defined by a conditional probability measure $\Pi^{\mathcal{H}}$ obtained by conditioning Π with respect to $\mathcal{H}$.

The parameter $\mathcal{A}$ is accordingly decomposed into $\mathcal{A} = \mathcal{A}_1 \vee \mathcal{A}_2$, where $\mathcal{A}_1$ and $\mathcal{A}_2$ are *defined* as follows:

$$\text{(i) } \mathcal{H} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{A}_1, \qquad \text{(ii) } \mathcal{Y} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{H} \vee \mathcal{A}_2. \tag{1}$$

Conditions (1) mean that $\mathcal{A}_1$ and $\mathcal{A}_2$ are defined as sufficient parameters for the marginal model generating $\mathcal{H}$ and for the conditional model generating $(\mathcal{Y} \mid \mathcal{H})$, respectively.

Once $\mathcal{H}$ corresponds to non observable events, the structural experiment $\mathcal{E}$ leads to the *statistical experiment* $\mathcal{E}_{\mathcal{Y}} = (A \times Y, \mathcal{A} \vee \mathcal{Y}, \Pi_{\mathcal{Y}})$, to be used for the inference based on the observable data only, where $\Pi_{\mathcal{Y}}$ is the trace of Π on $\mathcal{A} \vee \mathcal{Y}$.

Thus, this note establishes general relationships between the identifiability of these submodels; three main theorems, given in section 4, make them precise: Theorem 1 establishes that the identifiability of the marginal model is a necessary condition for the identifiability of the complete model; Theorem 2 gives an answer to the following question: are the identification of both models, the marginal and the conditional ones, sufficient conditions for the identification of the complete model? Finally, Theorem 3 gives sufficient conditions for the identifiability of the statistical model.

The two concepts of identification, along with some comments, are recalled in section 2. Section 3 illustrates these concepts in two cases: a discrete one and a normal one, and section 5 draws some conclusion. The final section gathers the proofs.

2 Weak and Strong Identification

Let $(M, \mathcal{M}, P)$ be an abstract probability space; sub-σ-fields of $\mathcal{M}$ are denoted by $\mathcal{X}_i$.

Definition 1 The sub-σ-field $\mathcal{X}_1$ *is weakly identified by the sub-σ-field* $\mathcal{X}_2$, denoted as $\mathcal{X}_1 \prec \mathcal{X}_2$, if $\sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_1) : f_2 \in [\mathcal{X}_2]^+\} = \mathcal{X}_1$, where $[\mathcal{X}_2]^+$ denotes the set of positive $\mathcal{X}_2$-measurable functions; and $\mathcal{X}_1$ *is weakly identified by* $\mathcal{X}_2$ *conditionally on* $\mathcal{X}_3$, denoted by $\mathcal{X}_1 \prec \mathcal{X}_2 \mid \mathcal{X}_3$, if $(\mathcal{X}_1 \vee \mathcal{X}_3)$ is weakly identified by $(\mathcal{X}_2 \vee \mathcal{X}_3)$.

Definition 2 Let $p \in [1, \infty]$; $\mathcal{X}_1$ is *p-strongly identified by* $\mathcal{X}_2$, denoted as $\mathcal{X}_1 \ll_p \mathcal{X}_2$, if for all f_1 positive p-integrable $\mathcal{X}_1$-measurable function, the following implication follows: $\mathbb{E}\,\{f_1 \mid \mathcal{X}_2\} = 0$ a.s. implies $f_1 = 0$ a.s.; and $\mathcal{X}_1$ *is p-strongly identified by* $\mathcal{X}_2$ *conditionally on* $\mathcal{X}_3$, denoted by $\mathcal{X}_1 \ll_p \mathcal{X}_2 \mid \mathcal{X}_3$, if $\mathcal{X}_1 \vee \mathcal{X}_3$ is p-strongly identified by $\mathcal{X}_2 \vee \mathcal{X}_3$.

Assume that $\mathcal{M} = \mathcal{X}_1 \vee \mathcal{X}_2$, where $\mathcal{X}_1$ represents a parameter, $\mathcal{X}_2$ an observation and that $(X_1 \times X_2, \mathcal{X}_1 \vee \mathcal{X}_2, P)$ is regular, that is, there exists a regular conditional probability on $\mathcal{X}_1$ given $\mathcal{X}_2$ and on $\mathcal{X}_2$ given $\mathcal{X}_1$. Then:

• The weak identification is defined by means of the σ-field $\sigma\{\mathbb{E}\,(f_2 \mid \mathcal{X}_1) : f_2 \in [\mathcal{X}_2]^+\}$, which is the smallest sub-σ-field of $\mathcal{X}_1$ which makes $\mathcal{X}_1$ and $\mathcal{X}_2$ conditionally independent, *i.e.* it is the *minimal sufficient* parameter; also it contains all the null sets of $\mathcal{X}_1$. To say that $\mathcal{X}_1$ is weakly identified by $\mathcal{X}_2$ means therefore that any (positive) measurable function of the parameters may be represented as a function of countably many sampling expectations of statistics. The identification problem appears therefore as a problem of minimal sufficiency on the parameter space; for more details, see Florens *et al.* (1990), section 4.3. Up to some technical conditions, the (usual) concept of sampling identification of a statistical model (see *e.g.* Koopmans and Reiersøl, 1950 or Bunke and Bunke, 1974), is essentially equivalent to the weak identification of the parameter by the observation; for more precision, see Florens *et al.* (1990), section 4.6.2.

• The strong identification is defined in terms of the injectivity of conditional expectation, extending thus the concept of (bounded) completeness in the sense of Lehmann and Scheffé (1955). Thus, to say that $\mathcal{X}_1$ is p-strongly identified by $\mathcal{X}_2$ means that $\mathcal{X}_1$ is a *parameter p-complete* for the process generating $(\mathcal{X}_2 \mid \mathcal{X}_1)$; for more details, see Florens *et al.* (1990), chapter 5.
• The weak identification $\mathcal{X}_1 \prec \mathcal{X}_2$ essentially depends on the sampling distribution of $(\mathcal{X}_2 \mid \mathcal{X}_1)$, whereas the p-strong identification $\mathcal{X}_1 \ll_p \mathcal{X}_2$ depends on the posterior distribution of $(\mathcal{X}_1 \mid \mathcal{X}_2)$. Nevertheless, there is a relationships beetween them, namely if $\mathcal{X}_1$ is p-strongly identified by $\mathcal{X}_2$ (with $p > 0$), then $\mathcal{X}_1$ is weakly identified by $\mathcal{X}_2$; for a proof, see Florens *et al.* (1990), Theorem 5.4.12.

3 Illustration of the Identification Concepts

This section exemplifies the concepts of weak as well as of strong identification in two particular cases: a discrete one and a normal one. The first case shows that the weak identification does not necessarily imply the strong identification, whereas the second case is an example in which these two concepts are equivalent. Furthermore, in both cases it is shown that the strong identification between two random variables implies a dimension restriction between them. For the sake of simplicity, we identify the random vectors with the σ-fields they generate.

3.1 Identifications in a Discrete Case

Let $(M, \mathcal{M}, P)$ be a probability space, N_r for $r = 1,2,3$ be finite sets, and $X_r : M \longrightarrow N_r$ for $r = 1,2,3$ be random variables. We define

$$K = \{k \in N_3 : P[X_3 = k] > 0\},$$

$$N_1^{(k)} = \{i \in N_1 : P[X_1 = i \mid X_3 = k] > 0\} \quad \text{for } k \in K,$$

$$N_2^{(k)} = \{j \in N_2 : P[X_2 = j \mid X_3 = k] > 0\} \quad \text{for } k \in K,$$

and, for $k \in K$, the $|N_1^{(k)}| \times |N_2^{(k)}|$ matrix $\mathbf{P}^{(k)}$ by

$$p_{ij|k} \equiv (\mathbf{P}^{(k)})_{ij} = P[X_1 = i, X_2 = j \mid X_3 = k] \quad \text{for } (i,j) \in N_1^{(k)} \times N_2^{(k)}.$$

For $k \in K$, let $\{p_{(i)}^{(k)} : i \in N_1^{(k)}\}$ be the rows of $\mathbf{P}^{(k)}$. The following proposition characterizes the weak and the strong identification in the discrete case; for a proof, see section 6.1.

Proposition 1

(i) $X_1 \prec X_2 \mid X_3$ *if and only if* $(\forall\, k \in K)$ $(\forall\, i, i' \in N_1^{(k)})$ $p_{(i)}^{(k)}$ *and* $p_{(i')}^{(k)}$ *are linearly independent.*

(ii) *For* $p > 0$, $X_1 \ll_p X_2 \mid X_3$ *if and only if* $(\forall\, k \in K)$ $\mathbf{P}^{(k)\prime}$ *is an injective linear transformation, i.e.* $r(\mathbf{P}^{(k)}) = |N_1^{(k)}|$.

Proposition 1 deserves the following comments:

1. The identification conditions given in this proposition are established $X_3 - a.s.$, that is, in the P-null set K^c they are not necessarily true.

2. As pointed out in section 2, the p-strong identification implies the weak one, and this is indeed trivial in the discrete case, and equally trivial is the fact that the inverse implication does not hold in general.

3. The condition $X_1 \ll_p X_2 \mid X_3$ implies ($X_3 - a.s.$) a dimension restriction between X_1 and X_2, namely that, for any $k \in K$, $r(\mathbf{P}^{(k)}) = |N_1^{(k)}| \leq |N_2^{(k)}|$, whereas $X_1 \prec X_2 \mid X_3$ does not imply any dimensional restriction.

4. If, for any $k \in K$, $\mathbf{P}^{(k)}$ is a bijective linear transformation (hence $|N_1^{(k)}| = |N_2^{(k)}|$), then $X_1 \ll_p X_2 \mid X_3$ as well as $X_2 \ll_p X_1 \mid X_3$, and reciprocally.

5. Suppose now that $\sigma(X_3)$ is trivial (*i.e.* $\sigma(X_3) = \{\emptyset, M\}$), and consider that X_1 is interpreted as a parameter and X_2 as an observation. In this discrete case, sampling identification is identical to the weak identification of the parameter by the observation as soon as the parameter space is defined as the set of points of (strictly) positive (prior) probability.

As a teacher's corner, we offer the following simple example: let $\Theta = \{\theta_1, \theta_2, \theta_3\}$ be the parametric space, $X = \{x_1, x_2, x_3\}$ be the sampling space and its joint probability distribution given by:

	x_1	x_2	x_3	
θ_1	.05	.05	.10	.20
θ_2	.10	.10	.10	.30
θ_3	.15	.15	.20	.50
	.30	.30	.40	1.00

It is easily verified that $\Theta \prec X$ is true, whereas neither $X \prec \Theta$ nor $\Theta \ll_p X$ is true. These results are related with the following immediate properties implied by the joint probability distribution of (Θ, X): that $\sigma(x_1 + x_2, x_3)$ is a sufficient statistics, and that $\sigma(\theta_1, \theta_2, \theta_3)$ is a minimal sufficient parameter.

3.2 Identifications in a Normal Case

Let $X = (X_1', X_2', X_3')' \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \mathbb{R}^{p_3}$ be a random vector. Let

$$Ker[V(X_1 \mid X_3)] = \{a \in \mathbb{R}^{p_1} : a'X_1 = \mathbb{E}\,(a'X_1 \mid X_3) \quad a.s.\}$$

$$Ker[C(X_2, X_1 \mid X_3)] = \{a \in \mathbb{R}^{p_1} : C(X_2, a'X_1 \mid X_3) = 0 \quad a.s.\}$$

where $V(\cdot \mid \cdot)$ and $C(\cdot, \cdot \mid \cdot)$ are the conditional variance and the conditional covariance operators, respectively. Suppose that $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu(X_3), \Sigma(X_3))$. Lemma 1 gives a simple result which provides an easy key to characterize the weak as well as the strong identification concepts under a normal hypothesis; for a proof, see section 6.1.

Lemma 1 *If $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu(X_3), \Sigma(X_3))$, then $X_3 - a.s.$*

$$Ker[V(X_1 \mid X_3)] = Ker[V(X_1 \mid X_2, X_3)] \cap Ker[C(X_2, X_1 \mid X_3)]. \qquad (2)$$

The following proposition shows, under a joint normal assumption, that the weak and the p-strong identification concepts are equivalent; moreover, this proposition gives a characterization of these concepts in terms of null space and of rank condition; for a proof, see section 6.1.

Proposition 2 *If* $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu(X_3), \Sigma(X_3))$*, then the following conditions are* $X_3 - a.s.$ *equivalent:*

(i) $X_1 \ll_p X_2 \mid X_3 \quad \forall\, p \in [1, \infty]$.

(ii) $X_1 \prec X_2 \mid X_3$.

(iii) $r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)]$.

(iv) $Ker[C(X_2, X_1 \mid X_3)] = Ker[V(X_1 \mid X_3)]$.

(v) $Ker[C(X_2, X_1 \mid X_3)] \subset Ker[V(X_1 \mid X_2, X_3)]$.

Proposition 2 deserves the following three comments:

1. As pointed out in section 2, $X_1 \prec X_2 \mid X_3$ essentially depends on the distribution of $(X_2 \mid X_1, X_3)$, whereas $X_1 \ll_p X_2 \mid X_3$ essentially depends on the distribution of $(X_1 \mid X_2, X_3)$. In the particular case of a joint normal distribution, these two conditions are equivalent as they essentially rely on a rank condition of the covariance matrix $C(X_2, X_1 \mid X_3)$.

2. From a σ-algebraic point of view, in the relation $X_1 \ll_p X_2 \mid X_3$ there is no loss of generality in assuming that $V(X_1 \mid X_3)$ is nonsingular (see, in the proof of this proposition, (iii) $\Longrightarrow$ (i)); therefore, the (weak or strong) identification of X_1 by X_2 conditionally on X_3 implies that $p_1 = r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)] \leq \min\{p_1, p_2\}$ $X_3 - a.s.$, *i.e.* the identification implies a dimension restriction between X_1 and X_2, namely $p_1 \leq p_2$ $X_3 - a.s.$

3. The model $(X_2 \mid X_1)$ is sampling identified as soon as condition (iii) is satisfied (with $\sigma(X_3) = \{\emptyset, M\}$). Thus, in this joint normal case, weak, strong and sampling identification concepts are equivalent.

4 Main Results

The main results of this note may be classified into two groups: those related with the weak identifiability of the complete model generating $\mathcal{Y} \vee \mathcal{H}$, and those related with the weak identifiability of the statistical model generating $\mathcal{Y}$. These two groups are nevertheless related as the following lemma establishes; its proof follows from noting that $[\mathcal{Y}]^+ \subset [\mathcal{Y} \vee \mathcal{H}]^+$.

Lemma 2 *Let $p \in [1, \infty]$. The weak (resp. p-strong) identification of $\mathcal{A}_1 \vee \mathcal{A}_2$ by $\mathcal{Y} \vee \mathcal{H}$ is a necessary condition for the weak (resp. p-strong) identification of $\mathcal{A}_1 \vee \mathcal{A}_2$ by $\mathcal{Y}$.*

4.1 Conditions for the Identifiability of the Complete Model

The following theorem establishes relationships between the weak identifiability of the complete model and the weak identifiability of $\mathcal{A}_1$ and of $\mathcal{A}_2$ by $\mathcal{Y} \vee \mathcal{H}$; also a necessary condition for the weak identifiability of the complete model may be deduced; similar results are obtained in the p-strong case. For a proof, see section 6.2.

Theorem 1 *Assume that*

$$\mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2. \tag{3}$$

(i) *If $\mathcal{A}_1 \vee \mathcal{A}_2$ is weakly identified by $\mathcal{Y} \vee \mathcal{H}$ then*

1. *$\mathcal{A}_1$ is weakly identified by $\mathcal{Y} \vee \mathcal{H}$, which in turn is equivalent to the weak identification of $\mathcal{A}_1$ by $\mathcal{H}$;*
2. *$\mathcal{A}_2$ is weakly identified by $\mathcal{Y} \vee \mathcal{H}$;*

(ii) *Let $p \in [1, \infty]$. If $\mathcal{A}_1 \vee \mathcal{A}_2$ is p-strongly identified by $\mathcal{Y} \vee \mathcal{H}$ then $\mathcal{A}_1$ is p-strongly identified by $\mathcal{H}$.*

In Theorem 1, condition (3) defines a *cut* between the marginal model generating $\mathcal{H}$ and the conditional model generating $(\mathcal{Y} \mid \mathcal{H})$; for details, see Florens *et al.* (1990), section 3.4.3; this is a (structural) condition on the (prior) distribution of $\mathcal{A}$. In a pure sampling theory approach, condition (3) may be replaced by a condition of factorization of the parameter space (or a condition of *variation-free*), namely $(a_1, a_2) \in A = A_1 \times A_2$; for details, see Barndorff-Nielsen (1978) or, recently, Barndorff-Nielsen and Koudou (1995). Thus, under this cut assumption (3), this theorem establishes that the weak (resp. p-strong) identification of the marginal model generating $\mathcal{H}$ is a necessary condition for the weak (resp. p-strong) identification of the complete model. Nevertheless, it is known that the weak identification of the conditional model is not a necessary condition for the weak identification of the complete model; for a counter-example, see *e.g.* Florens *et al.* (1979).

The following theorem establishes, without the cut condition (3), sufficient conditions for the identifiability of the complete model; these conditions are based on the identifiability of both the marginal model and the conditional model; for a proof, see section 6.2.

Theorem 2

(i) *If $\mathcal{A}_1$ is weakly identified by $\mathcal{H}$ and $\mathcal{A}_2$ is 2-strongly identified by $\mathcal{Y}$ conditionally on $\mathcal{H}$, then $\mathcal{A}_1 \vee \mathcal{A}_2$ is weakly identified by $\mathcal{Y} \vee \mathcal{H}$.*

(ii) *Let $p \in [1, \infty]$. If $\mathcal{A}_1$ is p-strongly identified by $\mathcal{H}$ and $\mathcal{A}_2$ is p-strongly identified by $\mathcal{Y}$ conditionally on $\mathcal{H}$, then $\mathcal{A}_1 \vee \mathcal{A}_2$ is p-strongly identified by $\mathcal{Y} \vee \mathcal{H}$.*

In other words, Theorem 2 establishes that if $\mathcal{A}_1$ is a minimal sufficient (resp. p-complete) parameter for the marginal experiment defined by the marginal probability measure $\Pi_{\mathcal{H}}$ and if $\mathcal{A}_2$ is a 2-complete parameter (resp. p-complete) for the conditional experiment defined by the conditional probability measure $\Pi^{\mathcal{H}}$, then $\mathcal{A}_1 \vee \mathcal{A}_2$ is a minimal sufficient (resp. p-complete) parameter for the structural experiment $\mathcal{E}$.

Remark 1 In Theorem 2 (i), the condition $\mathcal{A}_2 \ll_2 \mathcal{Y} \mid \mathcal{H}$ can not in general be replaced by the weaker condition $\mathcal{A}_2 \prec \mathcal{Y} \mid \mathcal{H}$ as the following counter-example shows: let $\mathcal{Y} = \sigma(y)$, $\mathcal{H} = \sigma(\eta)$ and $\mathcal{A}_r = \sigma(a_r)$, $r = 1, 2$, defined by the properties (1); suppose that a_r, $r = 1, 2$, η and y are discrete random variables with range spaces $a_r \in N_r$ $(r = 1, 2)$, $\eta \in N_3$ and $y \in N_4$, and assume that the marginal probabilities are strictly positive on each N_r. Then, by Proposition 1, $a_1 \prec \eta$ and $a_2 \prec y \mid \eta$ if and only if

$$\begin{aligned} &(\forall\, i_1 \neq i_1{}') \;\; (\exists\, i_3 \in N_3) \quad p_{i_3|i_1} \;\neq\; p_{i_3|i_1{}'}, \\ &(\forall\, i_3 \in N_3) \;\; (\forall\, i_2 \neq i_2{}') \;\; (\exists\, i_4 \in N_4) \quad p_{i_4|i_3\, i_2} \;\neq\; p_{i_4|i_3\,, i_2{}'}, \end{aligned} \tag{4}$$

respectively; and $(a_1, a_2) \prec (\eta, y)$ if and only if $\forall\,(i_1, i_2) \neq (i_1{}', i_2{}')\ \exists\,(i_3, i_4) \in N_3 \times N_4$ such that $p_{i_4\, i_3 | i_2\, i_1} \neq p_{i_4\, i_3 | i_2{}'\, i_1{}'}$, which, by (1), is equivalently rewritten as

$$p_{i_4 | i_3\, i_2}\; p_{i_3 | i_1} \neq p_{i_4 | i_3\, i_2{}'}\; p_{i_3 | i_1{}'}. \tag{5}$$

It is clear that the condition (4) does not imply condition (5). In general terms, this situation may be explained as follows: property (1.ii) implies that

$$\sigma\{\mathbb{E}\,(f \mid \mathcal{A}_1 \vee \mathcal{A}_2)\ :\ f \in [\mathcal{Y} \vee \mathcal{H}]^+\} \subset \sigma\{\mathbb{E}\,(f \mid \mathcal{A}_1 \vee \mathcal{A}_2)\ :\ f \in [\mathcal{T}]^+\}, \tag{6}$$

where $\mathcal{T} = \sigma\{\mathbb{E}\,(f \mid \mathcal{A}_2 \vee \mathcal{H})\ :\ f \in [\mathcal{Y} \vee \mathcal{H}]^+\}$. Thus, if $\mathcal{A}_2 \prec \mathcal{Y} \mid \mathcal{H}$ then $\mathcal{T}$ becomes equal to $\mathcal{A}_2 \vee \mathcal{H}$ and therefore (6) becomes

$$\sigma\{\mathbb{E}(f \mid \mathcal{A}_1 \vee \mathcal{A}_2) : f \in [\mathcal{Y} \vee \mathcal{H}]^+\} \subset \sigma\{\mathbb{E}(f \mid \mathcal{A}_1 \vee \mathcal{A}_2) : f \in [\mathcal{A}_2 \vee \mathcal{H}]^+\}. \tag{7}$$

If furthermore $\mathcal{A}_1 \prec \mathcal{H}$, then (7) implies that $\sigma\{\mathbb{E}\ (f \mid \mathcal{A}_1 \vee \mathcal{A}_2)\ :\ f \in [\mathcal{Y} \vee \mathcal{H}]^+\} \subset \mathcal{A}_1 \vee \mathcal{A}_2$ only. Condition $\mathcal{A}_2 \ll_2 \mathcal{Y} \mid \mathcal{H}$ ensures precisely an equality rather than an inclusion in relation (7) obtaining eventually the identifiability of the complete model generating $\mathcal{Y} \vee \mathcal{H}$.

Theorems 1 and 2 imply the following corollary:

Corollary 1 *Let $\mathcal{A}_1$ be a sufficient parameter for the model generating $\mathcal{H}$ and $\mathcal{A}_2$ be a 2-complete parameter for the conditional model generating $(\mathcal{Y} \mid \mathcal{H})$. Assume that $\mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2$. Then $\mathcal{A}_1 \vee \mathcal{A}_2$ is a minimal sufficient parameter for the complete model generating $\mathcal{Y} \vee \mathcal{H}$ if and only if $\mathcal{A}_1$ is a minimal sufficient parameter for the marginal model generating $\mathcal{H}$.*

4.2 Conditions for the Identifiability of the Statistical Model

Let us consider the weak identification of the statistical model $\mathcal{E}_\mathcal{Y}$, and note the following: (i) Lemma 2 and Theorem 1 give, under the cut condition (3), a necessary condition for the weak identification of the statistical model, namely the weak identification of the marginal model generating $\mathcal{H}$; (ii) although both the weak identifiability of the marginal model and the 2-strong identifiability of the conditional model have been used in Theorem 2, the weak identifiability of the statistical model has still not been obtained. The following theorem gives, among other, sufficient conditions for the identifiability of the statistical model by replacing the condition $\mathcal{A}_2 \ll_2 \mathcal{Y} \mid \mathcal{H}$ by a stronger one; for a proof, see section 6.2.

Theorem 3

(i) *If $\mathcal{A}_1$ is weakly identified by $\mathcal{H}$ and $\mathcal{A}_2 \vee \mathcal{H}$ is 2-strongly identified by $\mathcal{Y}$, then*

1. *$\mathcal{A}_1$ is weakly identified by $\mathcal{Y}$;*
2. *$\mathcal{A}_2$ is weakly identified by $\mathcal{Y}$;*
3. *$\mathcal{A}_1 \vee \mathcal{A}_2$ is weakly identified by $\mathcal{Y}$.*

(ii) *Let $p \in [1, \infty]$. If $\mathcal{A}_1$ is p-strongly identified by $\mathcal{H}$ and $\mathcal{A}_2 \vee \mathcal{H}$ is p-strongly identified by $\mathcal{Y}$ then $\mathcal{A}_1 \vee \mathcal{A}_2$ is p-strongly identified by $\mathcal{Y}$.*

Theorem 3 deserves the following three comments:

1. Assertion (i-1) (resp. (i-2)) states that $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) is a minimal sufficient parameter in the marginal statistical model obtained after integrating out $\mathcal{A}_2$ (resp. $\mathcal{A}_1$), whereas assertion (i-3) states that $\mathcal{A}_1 \vee \mathcal{A}_2$ is a minimal sufficient parameter in the statistical model $\mathcal{E}_\mathcal{Y}$. Note that assertion (i-2) is a straightforward consequence of the hypothesis of 2-strong identification of $\mathcal{A}_2 \vee \mathcal{H}$ by $\mathcal{Y}$.

2. This theorem gives sufficient conditions for the identifiability of a mixture model, namely that $\mathcal{A}_1$ be a minimal sufficient parameter for the process generating $\mathcal{H}$ and that $\mathcal{A}_2 \vee \mathcal{H}$ be a 2-complete parameter for the process generating $\mathcal{Y}$. In such a case, $\mathcal{H}$ might represent a non observable incidental, $(\mathcal{Y} \mid \mathcal{H} \vee \mathcal{A}_2)$ a sampling model and $(\mathcal{H} \mid \mathcal{A}_1)$ a component of a hierarchical prior distribution defined on $\mathcal{A}_1 \vee \mathcal{A}_2 \vee \mathcal{H}$. Note that the condition $\mathcal{A}_2 \vee \mathcal{H} \ll_2 \mathcal{Y}$ implies the condition $\mathcal{A}_2 \ll_2 \mathcal{Y} \mid \mathcal{H}$ (which is used in Theorem 2 (i)). Thus, relationships between Theorems 2 (i) and 3 (i) may be summarized as follows: under the condition $\mathcal{A}_1 \prec \mathcal{H}$, it follows that:

$$\begin{array}{ccc}
\mathcal{A}_2 \vee \mathcal{H} \ll_2 \mathcal{Y} & \Longrightarrow & \mathcal{A}_2 \ll_2 \mathcal{Y} \mid \mathcal{H} \\
\Downarrow \text{ th. 3} & & \Downarrow \text{ th. 2} \\
\mathcal{A}_1 \vee \mathcal{A}_2 \prec \mathcal{Y} & \underset{\text{lemma 2}}{\Longrightarrow} & \mathcal{A}_1 \vee \mathcal{A}_2 \prec \mathcal{Y} \vee \mathcal{H}.
\end{array}$$

3. Condition $\mathcal{A}_2 \vee \mathcal{H} \ll_2 \mathcal{Y}$ typically imposes a restriction between the number of latent variables and the number of observable variables; in both the discrete case and the normal case, this restriction may be computed explicitly (see section 3.2).

5 Concluding Remarks

The basic model that motivated this note is a model of partial observability that, in terms of density, may be written as $p(y \mid a) = \int p(y \mid \eta, a_2)\, p(\eta \mid a_1)\, d\eta$ with $a = (a_1, a_2)$ and η being not observable. Note first that any coordinate change of η does not affect the statistical model: this is aptly noticed *e.g.* by Dempster *et al.* (1977) or Bartholomew (1993), and motivates the σ-algebraic approach used in this note; indeed, this invariance property means that, because η is not observable, a mixture model is identified, at best, up to an arbitrary coordinate change of η, *i.e.* a genuinely σ-algebraic condition.

This note has made use of two (standard) concepts of identification: the weak one is standard in the theory of parametric identification and the strong one is standard in the theory of complete statistics. This note has shown, by means of a simple example, that these concepts are not equivalent and their combination give sufficient conditions for the identification of the statistical model. We guess that our theorems might give sufficient conditions, as long, at least, as the hypothesis of strong identification, of one of the component submodels, may not be replaced by a condition of weak identification.

When reading our theorems, the reader should keep in mind that $\mathcal{H} = \sigma(\eta)$ may stands for purely structural parameters, for purely incidental parameters (equivalently, latent variables) or for a combination of both. In particular, the authors' interest is motivated by a more extensive project bearing on the modelling of individual data in the presence of latent variables, as frequently occurs in LISREL type models, Item Response Models (see Mouchart and San Martín, 1999), etc.

6 Proofs

6.1 Proofs of Propositions

Proof of Proposition 1 By definition, $X_1 \prec X_2 \mid X_3$ (resp. $\ll_p$) is equivalent to $(X_1, X_3) \prec (X_2, X_3)$ (resp. $\ll_p$); therefore, assertion (i) (resp. (ii)) need to be proved for $\sigma(X_3) = \{\emptyset, M\}$ only. Let $N_r^* = \{i \in N_r : P(X_r = i) > 0\}$ for $r = 1, 2$ and let the $|N_1^*| \times |N_2|$ (resp. $|N_1| \times |N_2^*|$) matrix $\mathbf{P}_*$ (resp. $\mathbf{P}^*$) defined as $\mathbf{P}_* = [\ (P[X_1 = i, X_2 = j])_{ij}\ ]$ for $(i,j) \in N_1^* \times N_2$ (resp. $\mathbf{P}^* = [\ (P[X_1 = i, X_2 = j])_{ij}\ ]$ for $(i,j) \in N_1 \times N_2^*$). In this discrete case, any function of X_2 (resp. of X_1) may be written as a linear function, namely $f(X_2) = f'X_2$ (resp. $g(X_1) = g'X_1$) for an arbitrary coding of X_2 (resp. of X_1), and where $f \in \mathbb{R}^{|N_2|}$ (resp. $g \in \mathbb{R}^{|N_1|}$).
For (i), let the $|N_1^*| \times |N_2|$ matrix $\mathbf{P}_{2|1}$ defined as

$$\mathbf{P}_{2|1} = [\ (P[X_2 = j \mid X_1 = i])_{ij}\ ] = [\ p'_{2|i}\ ] \quad \text{for } (i,j) \in N_1^* \times N_2,$$

where $p'_{2|i}$ is the i-th row of $\mathbf{P}_{2|1}$. Then, by viewing $\mathbf{P}_{2|1}{}'$ as describing a function of X_1, namely $\{X_1 = i\} \longmapsto p_{2|i}$ for $i \in N_1^*$, valued in the $(|N_2|-1)$-simplex, and by recalling that $\sigma\{\mathbb{E}(f \mid X_1) : f \in [\sigma(X_2)]^+\}$ contains all the null sets of $\sigma(X_1)$, it follows that

$$\begin{aligned} &\sigma\{\mathbb{E}(f \mid X_1) : f \in [\sigma(X_2)]^+\} = \\ &\qquad \sigma\{f'\mathbf{P}_{2|1}{}' : f \in \mathbb{R}^{|N_2|}\} \vee \sigma\{\{X_1 = i\} : i \in N_1^{*^c}\} \\ &\qquad = \sigma\{\mathbf{P}_{2|1}{}'\} \vee \sigma\{\{X_1 = i\} : i \in N_1^{*^c}\}. \end{aligned}$$

Therefore $X_1 \prec X_2$ follows if and only if $\sigma\{\mathbf{P}_{2|1}{}'\} = \sigma\{\{X_1 = i\} : i \in N_1^*\}$; this last equality boils down to the injectivity of the mapping described by $\mathbf{P}_{2|1}{}'$, that is, $i \neq i'$ implies $p_{2|i} \neq p_{2|i'}$ for $i, i' \in N_1^*$, or equivalently,

$(\forall\, i, i' \in N_1^*)$ $(\exists\, j \in N_2)$ such that $p_{j|i} \neq p_{j|i'}$, which is equivalent to say that any two rows of $\mathbf{P}_*$ are linearly independent.
For (ii), let the $|N_1| \times |N_2^*|$ matrix $\mathbf{P}_{1|2}$ defined as

$$\mathbf{P}_{1|2} = [\,(P[X_1 = i \mid X_2 = j])_{ij}\,] \quad \text{for } (i,j) \in N_1 \times N_2^*.$$

Then, for $j \in N_2^*$, $\mathbb{E}[g(X_1) \mid X_2 = j] = g'\mathbf{P}_{1|2}\, e_j$ where e_j is the j-th column of $I_{|N_2^*|}$. Then $X_1 \ll_p X_2$ if and only if the following implication follows: $g'\mathbf{P}_{1|2} = 0 \implies g = 0$; that is, $\mathbf{P}'_{1|2}$ is an injective linear transformation, or, equivalently, that $\mathbf{P}^{*\prime}$ is an injective linear transformation.

□

Lemma 1 and Proposition 2 (i), (iii), (iv) and (v) are given without proof in Florens *et al.* (1993); we give here the proof in order to illustrate the technical aspects involved in the concepts of weak as well as of strong identification.
Proof of Lemma 1 In general,

$$V(a'X_1 \mid X_3) = \mathbb{E}\,[V(a'X_1 \mid X_2, X_3) \mid X_3] + V\,[\mathbb{E}\,(a'X_1 \mid X_2, X_3) \mid X_3]\,.$$

The nullity of each member of this equality corresponds to a pertaining to the respective null space of (2).

□

Proof of Proposition 2 In relation to the notation, A^+ denotes the Moore-Penrose inverse of the matrix A. Let $(M, \mathcal{M}, P)$ be an abstract probability space; the completed σ-field $\overline{\mathcal{X}}$ of $\mathcal{X} \subset \mathcal{M}$ is defined as $\overline{\mathcal{X}} = \mathcal{X} \vee \{A \in \mathcal{M} : P^2(A) = P(A)\}$; for details, see Florens *et al.* (1990), section 2.2.3. Finally, $\mathcal{B}$ denotes the Borel σ-field and, for $p \geq 1$, $[\mathcal{B}]_p$ denotes the set of p-integrable $\mathcal{B}$-measurable functions. As in proof of Proposition 1, this proposition need to be proved for $\sigma(X_3) = \{\emptyset, M\}$ only.
(iii) $\iff$ (iv) is a consequence of the rank theorem (see Halmos, 1974, Theorem 1, section 50).
(iv) $\iff$ (v) is a consequence of Lemma 1.
(ii) $\iff$ (iii) From the hypothesis it follows that $(X_2 \mid X_1) \sim \mathcal{N}_{p_2}(\mathbb{E}\,(X_2 \mid X_1), \Phi)$, where the conditional covariance matrix $\Phi = V(X_2 \mid X_1)$ does not depends on X_1, and $\mathbb{E}(X_2 \mid X_1) = g + R_{21}X_1$, with $R_{21} = C(X_2, X_1)\, V(X_1)^+$ is a $p_2 \times p_1$ matrix. Therefore $\sigma\{\mathbb{E}\,(f \mid \mathcal{X}_2) : f \in [\mathcal{X}_1]^+\} = \mathcal{X}_1$ if and only if $r(R_{21}) = r[V(X_1)]$ *a.s.* From Marsaglia (1964), it follows that $C(X_2, X_1) =$

$R_{21}\,V(X_1)$, which implies that $r[C(X_2, X_1)] \leq r(R_{21})$; since the other inequality is straightforward, it follows that $r[V(X_1)] = r[C(X_2, X_1)]$ *a.s.*
(i) $\Longrightarrow$ (iv) Let $d \in Ker[C(X_2, X_1)]$; then, under normality, it follows that $\mathbb{E}\,[f(d'X_1) \mid X_2] = \mathbb{E}\,[f(d'X_1)]\;\forall\, f \in [\mathcal{B}]_p$; this is equivalent to

$$\mathbb{E}\,\{f(d'X_1) - \mathbb{E}\,[f(d'X_1)] \mid X_2\} = 0 \quad a.s. \quad \forall\, f \in [\mathcal{B}]_p. \tag{8}$$

If $\mathcal{X}_1 \ll_p \mathcal{X}_2$, the equality (8) implies that $f(d'X_1) = \mathbb{E}\,[f(d'X_1)]$ *a.s.*, which is equivalent to $f(d'X_1) \in \overline{\{\emptyset, M\}}\;\forall\, f \in [\mathcal{B}]_p$, *i.e.* $d'X_1 \in \overline{\{\emptyset, M\}}$, or, equivalently, that $V(d'X_1) = 0$. Therefore, $Ker[C(X_2, X_1)] \subset Ker[V(X_1)]$. The other inclusion is a consequence of Lemma 1.
(iii) $\Longrightarrow$ (i) From a σ-algebraic point of view, there is no loss of generality in assuming that $V(X_1 \mid X_2)$ is non singular. Indeed, let $X_1^* = A_1{}'X_1 \in \mathbb{R}^{q_1}$, where $A_1 = (a_{ij})$ is a $p_1 \times q_1$ matrix with $r(A_1) = q_1$ and possibily $a_{ij} \in [\overline{\{\emptyset, M\}}]$, such that $\Phi = V(X_1^* \mid X_2)$ is non singular and $\mathcal{X}_1 \subset \overline{\mathcal{X}_1^*}$; then $\overline{\mathcal{X}_1^*} = \overline{\mathcal{X}_1}$. Therefore, $\mathcal{X}_1 \ll_1 \mathcal{X}_2 \Longleftrightarrow \overline{\mathcal{X}_1} \ll_1 \mathcal{X}_2 \Longleftrightarrow \overline{\mathcal{X}_1^*} \ll_1 \mathcal{X}_2 \Longleftrightarrow \mathcal{X}_1^* \ll_1 \mathcal{X}_2$. Moreover, from the hypothesis of normality, it follows that $(X_1^* \mid X_2) \sim \mathcal{N}_{q_1}(\mathbb{E}\,(X_1^* \mid X_2), \Phi)$ where $\mathbb{E}\,(X_1^* \mid X_2) = g + R_{12}^* X_2$ and $R_{12}^* = C(X_1^*, X_2)V(X_2)^+$. Let us recall that in the canonical exponential family of the form $p(x \mid \theta) = a(x)b(\theta)\exp[x'd(\theta)]$, the sufficient statistic x is L_1-complete as soon as the interior of the range of d is not empty (see *e.g.* Barndorff-Nielsen, 1978, Lemma 8.2). Applying this result to $\mathcal{X}_1^* \ll_1 \mathcal{X}_2$, with $\mathcal{X}_1^*$ (respectively, $\mathcal{X}_2$) in the role of the statistic (respectively, of the parameter), let us define

$$a(X_1^*) = (2\pi)^{-q_1/2}|\Phi|^{-1/2}\exp\{X_1^{*\prime}\Phi^{-1}\,[g - \tfrac{1}{2}X_1^*]\,\},$$

$$b(X_2) = \exp\{-\tfrac{1}{2}\,\{X_2{}'R_{12}^*{}'\Phi^{-1}R_{12}^*X_2 + g'\Phi^{-1}\,[g + 2R_{12}^*X_2]\,\}\,\},$$

$$c(X_1^*) = X_1^*,$$

$$d(X_2) = \Phi^{-1}R_{12}^*X_2.$$

Therefore, $\mathcal{X}_1^* \ll_1 \mathcal{X}_2$ if $r(R_{12}^*) = q_1 = r[V(X_1^* \mid X_2)] \leq r[V(X_1^*)] \leq q_1$, *i.e.* if $r(R_{12}^*) = r[V(X_1^*)]$. From Marsaglia (1964) and by construction of A_1, this equality is equivalent to $r[C(X_1, X_2)] = r[V(X_1)]$, which is equivalent to condition (iii). Thus, condition (iii) implies that $\mathcal{X}_1 \ll_1 \mathcal{X}_2$, which in turn implies $\mathcal{X}_1 \ll_p \mathcal{X}_2\;\forall\, p \geq 2$, that is, condition (i).

$\square$

6.2 Proofs of Theorems

Proofs of Theorems 1, 2 and 3 are based in the following general results established in Florens *et al.* (1990), to be referred in the sequel as EBS. The assertion "$\mathcal{M}_1$ is weakly (resp. p-strongly) identified by $\mathcal{M}_2$" is denoted as "$\mathcal{M}_1 \prec (\ll_p) \mathcal{M}_2$".

Theorem 4 *Let $p \in [1, \infty]$. If $\mathcal{M}_2 \perp\!\!\!\perp \mathcal{M}_4 \mid \mathcal{M}_1 \vee \mathcal{M}_3$ then*

(i) *(EBS: Theorems 4.5.3, resp. 5.4.5) $\mathcal{M}_1 \prec (\ll_p) \mathcal{M}_2 \mid \mathcal{M}_3$ implies $\mathcal{M}_1 \prec (\ll_p) \mathcal{M}_2 \mid \mathcal{M}_3 \vee \mathcal{M}_4$.*

(ii) *(EBS: Theorems 4.5.4, resp. 5.4.6) $\mathcal{M}_2 \prec (\ll_p) \mathcal{M}_1 \vee \mathcal{M}_4 \mid \mathcal{M}_3$ implies $\mathcal{M}_2 \prec (\ll_p) \mathcal{M}_1 \mid \mathcal{M}_3$.*

(iii) *(EBS: Theorem 5.4.10) If $\mathcal{M}_4 \ll_p \mathcal{M}_1 \mid \mathcal{M}_3$ and $\mathcal{M}_1 \ll_p \mathcal{M}_2 \mid \mathcal{M}_3$ then $\mathcal{M}_4 \ll_p \mathcal{M}_2 \mid \mathcal{M}_3$.*

(iv) *(EBS: Theorem 5.4.14 (ii)) with $\mathcal{M}_3 = \{\emptyset, M\}$, if $\mathcal{M}_1 \ll_p \mathcal{M}_4$ then, for all $p \in (1, \infty)$,*

$$\sigma\{\mathbb{E}\,(f \mid \mathcal{M}_2) \,:\, f \in [\mathcal{M}_4]^+\} \;=\; \sigma\{\mathbb{E}\,(f \mid \mathcal{M}_2) \,:\, f \in [\mathcal{M}_1]^+\}.$$

Proof of Theorem 1 Firstly note that the cut assumption (3) along with property (1) implies the following three conditions:

$$\text{(i) } \mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2 \mid \mathcal{H}, \quad \text{(ii) } \mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2 \mid \mathcal{Y} \vee \mathcal{H}, \quad \text{(iii) } \mathcal{Y} \perp\!\!\!\perp \mathcal{A}_1 \mid \mathcal{H}. \tag{9}$$

The following arguments are established by using Theorem 4 (ii):
For (i-1), $\mathcal{A}_1 \vee \mathcal{A}_2 \prec \mathcal{Y} \vee \mathcal{H}$ implies $\mathcal{A}_1 \prec \mathcal{Y} \vee \mathcal{H} \vee \mathcal{A}_2$, wich in turn implies $\mathcal{A}_1 \prec \mathcal{H} \vee \mathcal{A}_2$ since property (1.ii) implies $\mathcal{Y} \perp\!\!\!\perp \mathcal{A}_1 \mid \mathcal{H} \vee \mathcal{A}_2$. But $\mathcal{A}_1 \prec \mathcal{H} \vee \mathcal{A}_2$ implies $\mathcal{A}_1 \prec \mathcal{H}$ under condition (9.i), which is equivalent to $\mathcal{A}_1 \prec \mathcal{Y} \vee \mathcal{H}$. Indeed, $\mathcal{A}_1 \prec \mathcal{H}$ implies $\mathcal{A}_1 \prec \mathcal{Y} \vee \mathcal{H}$ since $[\mathcal{H}]^+ \subset [\mathcal{Y} \vee \mathcal{H}]^+$; conversely, $\mathcal{A}_1 \prec \mathcal{Y} \vee \mathcal{H}$ implies $\mathcal{A}_1 \prec \mathcal{H}$ under condition (9.iii).
For (i-2), $\mathcal{A}_1 \vee \mathcal{A}_2 \prec \mathcal{Y} \vee \mathcal{H}$ implies $\mathcal{A}_2 \prec \mathcal{Y} \vee \mathcal{H} \vee \mathcal{A}_1$, which in turn implies $\mathcal{A}_2 \prec \mathcal{Y} \vee \mathcal{H}$ under condition (9.ii).

For (ii), $\mathcal{A}_1 \vee \mathcal{A}_2 \ll_p \mathcal{Y} \vee \mathcal{H}$ implies $\mathcal{A}_1 \ll_p \mathcal{Y} \vee \mathcal{H}$, which in turn implies $\mathcal{A}_1 \ll_p \mathcal{H}$ under condition (9.iii).

□

Proof of Theorem 2

For (i), $\mathcal{A}_2 \vee \mathcal{H} \ll_2 \mathcal{Y} \vee \mathcal{H}$ along with $\mathcal{A}_1 \vee \mathcal{A}_2 \perp\!\!\!\perp \mathcal{Y} \vee \mathcal{H} \mid \mathcal{A}_2 \vee \mathcal{H}$ (property (1.ii)) implies

$$\sigma\{\mathbb{E}(f \mid \mathcal{A}_1 \vee \mathcal{A}_2) : f \in [\mathcal{Y} \vee \mathcal{H}]^+\} = \sigma\{\mathbb{E}(f \mid \mathcal{A}_1 \vee \mathcal{A}_2) : f \in [\mathcal{A}_2 \vee \mathcal{H}]^+\} \quad (10)$$

by Theorem 4 (iv). But $\mathcal{A}_1 \prec \mathcal{H}$ along with $\mathcal{A}_2 \perp\!\!\!\perp \mathcal{H} \mid \mathcal{A}_1$ (property (1.i)) implies $\mathcal{A}_1 \vee \mathcal{A}_2 \prec \mathcal{A}_2 \vee \mathcal{H}$ by Theorem 4 (i); therefore $\sigma\{\mathbb{E}\,(f \mid \mathcal{A}_1 \vee \mathcal{A}_2) : f \in [\mathcal{A}_2 \vee \mathcal{H}]^+\} = \mathcal{A}_1 \vee \mathcal{A}_2$. The conclusion follows by using relationship (10).

For (ii), $\mathcal{A}_1 \ll_p \mathcal{H}$ along with $\mathcal{A}_2 \perp\!\!\!\perp \mathcal{H} \mid \mathcal{A}_1$ (property (1.i)) implies $\mathcal{A}_1 \vee \mathcal{A}_2 \ll_p \mathcal{A}_2 \vee \mathcal{H}$; this last condition along with $\mathcal{A}_2 \vee \mathcal{H} \ll_p \mathcal{Y} \vee \mathcal{H}$ implies $\mathcal{A}_1 \vee \mathcal{A}_2 \ll_p \mathcal{Y} \vee \mathcal{H}$ under (1.ii) by using Theorem 4 (iii).

□

Proof of Theorem 3 The proof of (i-1) and (i-3) are similar to that of Theorem 2 (i), whereas the proof of (ii) is similar to that of Theorem 2 (ii).

□

Acknowledgements

Without implication, a particular debt is owed to J. -M. Rolin for many discussions leading to clarifying several issues and improving several results.

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

Barndorff-Nielsen, O. and Koudou, A. E. (1995). Cuts in Natural Exponential Families, *Theory Probab. Appl.* **40**, 220-229.

Bartholomew, D. J. (1993). Estimating Relationships Between Latent Variables, *Sankhyā A* **55**, 409-419.

Bunke, H. and Bunke, O. (1974). Identifiability and Estimability, *Math. Operationsforsch. Statist.* **5**, 223-233.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. R. Statist. Soc. B* **39**, 1-38.

Florens, J. -P., Mouchart, M. and Richard, J. -F. (1979). Specification and Inference in Linear Models, Discussion Paper 7943, CORE, Université catholique de Louvain.

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

Florens, J. -P., Mouchart, M. and Rolin, J. -M. (1993). Noncausality and Marginalization of Markov Processes, *Econ. Theory.* **9**, 241-262.

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

Halmos, P. (1974). *Finite-Dimensional Vector Space.* Springer, New York.

Koopmans, T. C. and Reiersøl, O. (1950), The Identification of Strutural Characteristics, *Ann. Math. Statist.* **21**, 165-181.

Lehmann, E. L. and Scheffé, H. (1955). Completeness, similar regions and unbiased tests, Part II, *Sankhyā* **15**, 219-236.

Marsaglia, G. (1964). Conditional Means and Covariance of Normal Variable with Singular Covariance Matrix, *J. Amer. Statis. Ass.* **59**, 1203-1204.

Mouchart, M. and San Martín, E. (1999). Identification in Sifted Mixture Models, forthcoming Discussion Paper, Institut de Statistique, Université catholique de Louvain.