Identification in Sifted Mixture Models with an Application to Item Response Models*

Michel Mouchart and Ernesto San Martín
Université catholique de Louvain.

Abstract

After introducing, at the level of model specification, three basic distinctions –the first one between structurals and incidentals, the second one between incidentals and the third one between observables and unobservables– this paper establishes that, in the presence of incidentals, a sifting condition ensures that the identification of a conditional model corresponding to a given sample size may be obtained from the identification of the corresponding individual conditional submodels. Using these results, sufficient conditions are given for the identification of the statistical model obtained after integrating out the unobservables; the problem of minimal predictive sufficency is also analysed. The second part of this paper illustrates the use of the main results in the class of IRT models; this class is specified by introducing a double sifting structure in the conditional model generating the observables conditionally on the unobservables, and an hypothesis of exchangeability in the marginal model generating the unobservables. Under such a specification, the identifiability of the statistical model in IRT models for psychomotor assesment as well as for a Rasch model are analysed.

Key words: incidentals and structurals, measurable separability, sifting sequence, strong and weak identification, completeness, sufficiency and minimal sufficiency, minimal predictive sufficiency, mixture, exchangeability, item response model

*This research was supported by the contract *Projet d'Actions de Recherche Concertées* ARC 98/03-217 of the Belgian Government. Comments by E. Scheihing are gratefully acknowledged as they have been useful for the improvement of the exposition. Without implication, a particular debt is owed to J. -M. Rolin for many discussions leading to clarifying and improving several issues of this paper.

1 Introduction: model specification and overview of the paper

The development of statistical models in the social and behavioural sciences, such as in micro-econometrics, psychometrics or sociometrics, is often oriented toward the modelling of individual data; examples include the analysis of the results of educational tests in psychology, the analysis of survey or panel data on voting behaviour in political sciences, or the analysis of household consumption data in micro-econometrics and surveys; many examples may also be found in medical applications. In those applications, the data basis typically embraces a large number of individuals and the underlying theoretical framework insists on the idea that individual are *not identical*; the statistical modelling rests eventually on the search for stochastic structures providing interpretable and managable patterns.

The object of this paper is therefore to propose a modelling strategy of individual data putting a particular interest to the following problem: in the presence of latent (or, unobservable) variables, under which assumptions the (statistical) information provided by a reference population may be recovered by the (statistical) information provided by each individual data (or, statistical unit)? This section makes precise such a problem by specifiying the basic model; the model specification consists in making three distinctions: the first one between structurals and incidentals, a motivation of which is given in subsection 1.1; the second one between incidentals, and the third one between observables and unobservables. These distinctions are formalized in subsection 1.2; subsection 1.3 gives an overview of the paper.

1.1 Incidentality and structurality: general motivations

Let us consider a sample of n individuals $L_n = \{i_1, \ldots, i_n\}$, where the i_j's are "labels" identifying each individual drawn from a given population of reference. From now on, we shall alleviate the notation into $L_n = \{1, \ldots, i, \ldots, n\}$ with the implicit understanding that i actually denotes a label identifying an individual. We want to progressively elaborate a statistical model under which data relative to these individuals will be analysed. Our modelling strategy wants to take into account several aspects, among which (i) assuming the sampling stopping non informative, the basic structure of the model should be valid for any sample size and accept as an asymptotic limit an adequate probability measure on the infinite sequences,

(ii) as the sample size n increases, the information also increases monotonically, but the model should make explicit which part of the increasing information is constant within a given population of reference; for example, when modelling voting behaviour, or consumption pattern, one crucial aspect is precisely the distinction between the strictly individual component of the behaviour (such an information increases with the sample size) and the "environmental" (or populational) component (such an information is the same for any sample size).

The above suggests that, from a statistical point of view, when modelling individual data, a first distinction should be operated between "incidental" characteristics and "structural" characteristics: the first one represents particular features of each individual (or, statistical unit) and therefore the number of incidentals increases with the number of statistical units or with the sample size (see *e.g.* Neyman and Scott, 1948 and Wald, 1948), whereas the second one represents features common to all individuals. The usual concept of "population of reference" eventually becomes operational by defining the structurals, namely by gathering in this population the individuals (or, statistical units) sharing a same value of the structurals.

This distinction suggests some change of emphasis in the modelling strategy with respect to the standard modelling. Indeed, the standard strategy starts at the individual level and thereafter combines the individual likelihood into a sample likelihood by adding suitable assumptions, typically in the form of independence. Such a procedure takes as granted, or as "obviously meaningful", the distinction between the structurals and the incidentals, and therefore what is the population of reference. This paper wants to make explicit basic hypotheses deemed to meet the following purposes: (i) each one should be elementary enough so as to be endowable with a clear contextual interpretation; (ii) taken all together these hypotheses should justify the usual likelihood forms. Making those assumptions explicit serves a dual role: to ensure a logical coherence (or mathematical rigor) *and* to ensure an adequate contextual interpretability; with respect to this last aspect, see *e.g.* Cox (1990), Wermuth and Lauritzen (1990) and Cox and Wermuth (1996). This paper is therefore concerned with a theory of model building or, in terms of hypothesis testing, with the statistical foundation of a maintained hypothesis. At this level, the issue of identification is essential.

Easing the contextual interpretability of each hypotheses calls for two requeriments: (i) each single hypothesis should be as simple as possible and eventually avoid inessential side-features; thus the hypotheses will avoid

parametric forms (in the sense of specific finite-dimensional characterization of distributions) and even specific coordinate choice of the random variables; (ii) in order to avoid redundancies among hypotheses, an important aspect of interpretability, hypothese are introduced one by one in such an order that each one be easily interpretable conditionally on the hypotheses previously introduced. By so doing, the role of each hypothesis is interpreted in the light of the previous ones.

The first requeriment motivates a σ-algebraic approach, *i.e.* an approach where all hypotheses are stated in terms of σ-fields, which will be identified by the use of script letters. Indeed, the σ-field generated by a random variable summarizes all information provided by any coordinate transformation of such a random variable. Moreover, in the presence of latent (or, unobservable) variables, the statistical model –obtained after integrating out the unobservable variables– is invariant under any coordinate change of the unobservable variables; thus, the statistical model is identified, at least, up to an arbitrary coordinate change of the unobservable variable, *i.e.* a genuinely σ-algebraic condition; for details, see Mouchart and San Martín (1999). The reader reluctant to handling σ-fields may however read the exposition by interpreting the script letters as random variables (or vectors) *provided* (s)he keeps in mind that the hypotheses are not relative to a particular choice of coordinates, *i.e.* that for each such random variable, (s)he be ready to add something like "and any bijective bimeasurable transformation of it" and that the techniques of proof are often more involved in terms of random variables than in terms of σ-fields.

1.2 A filtered structural experiment

Let us consider again a sample of n individuals $L_n = \{1, \ldots n\}$. To this sample, let us associate a sample space $(I_{[n]}, \mathcal{I}_{[n]})$, which may be viewed as the range space of the characteristics of interest defined on the sample. The *experimental model* is built from a family of probability measures on the sample space, to be called *sampling probabilities*, and we consider any such probability measure as a *structure*. The set of sampling probabilities may be represented as $(S, \mathcal{S})$, where the elements of S are called *structurals*. Two aspects of the structurals should be pointed out: firstly, they characterize a distribution on the sample space, that is, a role of *parameters*; secondly, they represent features common to all individuals belonging to a reference population implicitly defined by a probability distribution. The model is

completed by introducing a unique probability measure $\Pi_{[n]}$ on the product space $I_{[n]} \times S$, eventually defining a *structural experiment* $\mathcal{E}_{[n]}$ by

$$\mathcal{E}_{[n]} \;=\; (S \times I_{[n]}, \mathcal{S} \vee \mathcal{I}_{[n]}, \Pi_{[n]}). \tag{1.1}$$

The probability measure $\Pi_{[n]}$ is typically constructed as a markovian product of a sampling transition considered as a probability on the sample space conditional on the structurals and of a probability on the structurals, possible called a prior probability; for details, see Florens *et al.* (1990, section 1.2). At the level of model building, this prior probability is either left unspecified or constrained by very general properties such as the σ-field generated by its null sets and some independence properties. In this paper, the main role of this prior probability is to allow probabilistic reasoning and integration with respect to the underlying probability space $\mathcal{E}_{[n]}$.

The information captured by the sample space $(I_{[n]}, \mathcal{I}_{[n]})$ is "incidental" in the sense that their content depends on and monotonically increases with the sample size. More explicitly, the incidentals are formalized by means of a *filtration of incidentals* $\mathcal{I}_{[n]} \uparrow \mathcal{I}_{[\infty]}$, that is, $\mathcal{I}_{[n]} \subset \mathcal{I}_{[n+1]}\ \forall\, n \geq 1$ and $\mathcal{I}_{[\infty]} = \bigvee_{n \geq 1} \mathcal{I}_{[n]}$. When we consider an increasing sequence of sample sizes, we obtain a *filtered structural experiment* $\mathcal{E}$ given by

$$\mathcal{E} \;=\; (S \times I_{[\infty]}, \mathcal{S} \vee \mathcal{I}_{[\infty]}, \Pi, \mathcal{I}_{[n]} \uparrow \mathcal{I}_{[\infty]}), \tag{1.2}$$

provided the sequence $\Pi_{[n]}$, in (1.1), forms a projective system, the projective limit of which is Π; for details about projective systems, see *e.g.* Rao (1971).

1.2.1 Separating incidentals and structurals

Let us face a first identification problem: how do we separate incidentals from structurals? It is indeed desirable, for the interpretation, to ensure oneself that the incidentals will not incorporate common features and that these ones are completely embodied in the structurals. Mathematically, this may be expresed by requiring that the elements common to $\mathcal{S}$ and $\mathcal{I}_{[n]}$ may only be trivial, relatively to $\Pi_{[n]}$, namely:

$$\overline{\mathcal{S}} \cap \overline{\mathcal{I}_{[n]}} \;=\; \overline{\mathcal{M}_{0,n}}, \tag{1.3}$$

where $\mathcal{M}_{0,n} = \{\emptyset, S \times I_{[n]}\}$ is the trivial σ-field and the upper bar refers to the measurable $\Pi_{[n]}$-completion, that is

$$\overline{\mathcal{M}_{0,n}} \;=\; \{A \in S \times I_{[n]} \;:\; \Pi_{[n]}(A) \in \{0,1\}\},$$

$$\overline{\mathcal{M}} \;=\; \mathcal{M} \vee \overline{\mathcal{M}_{0,n}} \quad \text{for any } \mathcal{M} \subset \mathcal{S} \vee \mathcal{I}_{[n]};$$

for details concerning the measurable completion, see Florens *et al.* (1990, section 2.2.3). A condition equivalent to property (1.3) is:

> For any positive $\mathcal{S}$-measurable function s, $\mathbb{E}(s \mid \mathcal{I}_{[n]}) = s$ if and only if s is almost surely a constant,

or, equivalently,

> For any positive $\mathcal{I}_{[n]}$-measurable function i_n, $\mathbb{E}(i_n \mid \mathcal{S}) = i_n$ if and only if i_n is almost surely a constant.

Property (1.3) is called the *measurable separability* of $\mathcal{S}$ and $\mathcal{I}_{[n]}$; this property is implied by, and is much weaker than, the independence (see Florens *et al.*, 1990, Theorem 5.2.7). From a probabilistic point of view, this property allows one to avoid pathologies errors due to unwarranted uses of densities; examples may be found in Dawid (1979) and in Florens *et al.* (1993). For details on measurable separability, see *e.g.* Florens *et al.* (1990, sections 5.2 and 5.3). In this paper, however, the introduction of this condition is more motivated by interpretative reasons than by technical conditions.

Remark 1.1 Because the sequence of $\Pi_{[n]}$ generates a projective system, the sequence of the completed trivial σ-fields, $\overline{\mathcal{M}_{0,n}}$, conform a filtration, the limit of which is $\overline{\mathcal{M}_{0,\infty}}$, that is $\overline{\mathcal{M}_{0,n}} \uparrow \overline{\mathcal{M}_{0,\infty}}$, where $\overline{\mathcal{M}_{0,\infty}}$ is the completed trivial σ-field with respect to the projective limit Π as defined in (1.2).

1.2.2 Separating incidental information

The next step is to specify the role of each individual, or statistical unit, in conforming the information structure associated to the sample L_n. To this end, we associate to each individual i a size one sample space $(I_i, \mathcal{I}_i)$ such that

$$\mathcal{I}_{[n]} \;=\; \bigvee_{1 \le i \le n} \mathcal{I}_i. \tag{1.4}$$

In other words, $\mathcal{I}_{[n]}$ is now interpreted as the canonical filtration associated to the sequence $\{\mathcal{I}_i : 1 \leq i \leq n\}$. We now face a second identification problem: how do we separate the information coming from different individuals? Mathematically, this may be expressed by requiring that the elements common to two different individuals may only be structural; more precisely,

$$\overline{\mathcal{I}_i \vee \mathcal{S}} \cap \overline{\mathcal{I}_j \vee \mathcal{S}} = \overline{\mathcal{S}} \qquad \forall\, i \neq j, \tag{1.5}$$

or, equivalently,

> For any positive $(\mathcal{I}_i \vee \mathcal{S})$-measurable function s_i, $\mathbb{E}(s_i \mid \mathcal{I}_j \vee \mathcal{S}) = s_i$ if and only if s_i is a $\mathcal{S}$-measurable function $\forall\, i \neq j$.

Property (1.5) is a conditional extension of property (1.3) and is read "$\mathcal{I}_i$ and $\mathcal{I}_j$ are measurably separated conditionally on $\mathcal{S}$".

An elementary property of measurable separability (see Florens *et al.*, 1990, Theorem 5.2.5) ensures that conditions (1.3) and (1.5) jointly imply the measurable separability of $\mathcal{I}_i$ and $\mathcal{I}_j$ for any $i \neq j$, that is

$$\overline{\mathcal{I}_i} \cap \overline{\mathcal{I}_j} = \overline{\mathcal{M}_{0,n}}. \tag{1.6}$$

Therefore, after integrating out the structural $\mathcal{S}$, the *only* common information to those associated to individuals i and j $(i \neq j)$ is trivial.

1.2.3 Individual structural experiments

Given properties (1.4) and (1.6), n *individual experiments* $\{\mathcal{E}_i : i = 1, \ldots, n\}$ may be built from the global experiment $\mathcal{E}_{[n]}$ by defining a probability measure Π_i as the trace of $\Pi_{[n]}$ on $(\mathcal{S} \vee \mathcal{I}_i)$, for $i = 1, \ldots, n$, namely:

$$\mathcal{E}_i = (S \times I_i,\, \mathcal{S} \vee \mathcal{I}_i, \Pi_i). \tag{1.7}$$

Thus, the structural $\mathcal{S}$ is common to all individual experiments $\mathcal{E}_i$; hence, the complete structural experiment $\mathcal{E}_{[n]}$, given by (1.1), may be fully characterized by the n individuals structural experiments $\{\mathcal{E}_i : i = 1, \ldots, n\}$ under a mutual independence condition of the $\mathcal{I}_i$'s, namely

$$\underset{1 \leq i \leq n}{\perp\!\!\!\perp} \mathcal{I}_i \mid \mathcal{S}. \tag{1.8}$$

The problem considered in this paper is related with the characterization of the complete structural experiment $\mathcal{E}_{[n]}$ from the individual structural experiments $\mathcal{E}_i$'s, but in the presence of latent (or unobservable) variables.

1.2.4 Observable-unobservable decomposition

Let us now therefore decompose the incidental components $\mathcal{I}_{[n]}$ into observable and unobservable components. As an example, let us consider the following situation: for an individual i, let y_i be, for instance, the result of a behavioural test which measures, with error, a concept not directly observable and represented by a latent variable η_i. Let $\mathcal{Y}_i = \sigma(y_i)$, $\mathcal{H}_i = \sigma(\eta_i)$, $\mathcal{Y}_1^n = \bigvee_{1 \leq i \leq n} \mathcal{Y}_i$ and similarly for $\mathcal{H}_1^n$. By so-doing, the filtration $\mathcal{I}_{[n]}$ in (1.4) is decomposed into

$$\mathcal{I}_{[n]} = \mathcal{Y}_1^n \vee \mathcal{H}_1^n \qquad \forall\, n \geq 1. \tag{1.9}$$

Structural modelling typically leads to a marginal-conditional decomposition of $\mathcal{E}_{[n]}$ with respect to the unobservable filtration $\mathcal{H}_1^n$, *i.e.* to a pair of reduced models specifying the generation of $\mathcal{H}_1^n$ and $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$, respectively. The structural parameter $\mathcal{S}$ is accordingly decomposed into $\mathcal{S} = \mathcal{S}_1 \vee \mathcal{S}_2$, where $\mathcal{S}_i$, $i = 1, 2$, are *defined* by the following properties:

$$\text{(i) } \mathcal{H}_1^n \perp\!\!\!\perp \mathcal{S} \mid \mathcal{S}_1 \quad \forall\, n \geq 1, \quad \text{(ii) } \mathcal{Y}_1^n \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_1^n \vee \mathcal{S}_2 \quad \forall\, n \geq 1. \tag{1.10}$$

Conditions (1.10) mean that $\mathcal{S}_1$ and $\mathcal{S}_2$ are defined as structural sufficient parameters for the submodels generating $\mathcal{H}_1^n$ and $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$, respectively.

Remark 1.2

1. Neither the existence nor the construction of $\mathcal{S}_i$, $i = 1, 2$, satisfying (1.10) for a given structural experiment is within the scope of this paper; for a partial answer of this topic, see *e.g.* Florens *et al.* (1990, chapter 2).

2. In case no unobservable incidentals are considered, the distinction between incidentals and structurals boils down to the usual distinction between observations and parameters.

3. Note that property (1.10.i) implies that $\overline{\mathcal{H}_i \vee \mathcal{S}_1} \cap \overline{\mathcal{S}_1 \vee \mathcal{S}_2} = \overline{\mathcal{S}_1}$ for all $i = 1, \ldots, n$, which, along with property (1.5), implies that

$$\overline{\mathcal{H}_i \vee \mathcal{S}_1} \cap \overline{\mathcal{H}_j \vee \mathcal{S}_1} = \overline{\mathcal{S}_1} \qquad \forall\, i \neq j, \tag{1.11}$$

that is, the information common to the unobservable incidentals $\mathcal{H}_i$ and $\mathcal{H}_j$ with $i \neq j$ is only the structural $\mathcal{S}_1$ (which characterizes the global process generating the unobservable incidentals $\mathcal{H}_1^n$). Furthermore, (1.3) and (1.11) jointly imply the marginal measurable separability $\overline{\mathcal{H}_i} \cap \overline{\mathcal{H}_j} = \overline{\mathcal{M}_{0,n}}$ for $i \neq j$.

1.3 Overview of the paper

Once $\mathcal{H}_1^n$ is considered as an unobservable filtration, the natural way for the specification of the complete structural experiment $\mathcal{E}_{[n]}$ is to specify the marginal process generating $\mathcal{H}_1^n$ and the conditional process generating $\mathcal{Y}_1^n$ conditionally on $\mathcal{H}_1^n$. The following problems, to be considered in this paper, arise: (1) under which assumptions the structural $\mathcal{S}_2$ (sufficient for the process generating $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$) is sufficient for each individual conditional process generating $(\mathcal{Y}_i \mid \mathcal{H}_i)$ for $i = 1, \ldots, n$? Proposition 3.1 provides an answer. (2) Under which assumptions identification conditions for models corresponding to a given sample size (that is, defined in the global conditional structural experiment given $\mathcal{H}_1^n$) may be obtained from conditions defined on a conditional model corresponding to samples of size 1 (that is, defined in a conditional structural experiment given $\mathcal{H}_i$)? Theorems 3.1 and 3.2 provide those results for strong and for weak identification, respectively; these two concepts of identifications are also reminded in section 2. (3) A third problem regards the identification of the *statistical model*, bearing on the manifest variables as well as the structural parameters only; the difficulty of this problem comes from the fact that the incidentals involve individual unobservables (along with individual observables) endowing eventually the statistical model with a structure of mixture. The identification of the statistical model is eventually based on conditions of identification of each submodel, namely the marginal one generating the unobservable incidentals and the conitional one generating the observable incidentals conditionally on the unobservable incidentals. Theorem 3.3 gives sufficient conditions for the weak identification of the statistical model. Theorem 3.4 provides the main consequence of the previous results for predictive purposes.

The second part of this paper illustrates the use of the main results to the Item Response (IRT) Models, a class of models widely used in educational testing and in psychometrics. In section 4.1, this class of models is specified in two steps: the first one displays a double sifting structure, defined in section 2, in order to define the basic framework of the conditional model generating the responses of the individuals conditionally on the non observable incidentals; the second one assumes that the non observable incidentals are exchangeable. Under this general specification, Theorem 4.1 gives sufficient conditions for the identification of the structural parameters in a general class of semiparametric IRT models; corollaries 4.1 and 4.2 apply the general theorem to the psychomotor assesment model and to the Rasch model, respectively. Section 5 gathers some concluding remarks.

The reader may like to distinguish, in this paper, three parts, namely (i) an exposition of a general framework in sections 2, 3 and 5; (ii) an application of the general framework developed in section 4, where most (but not all) of the basic facts developed in the expository part are exemplified; (iii) a formal complement, developed in section 6, gives some comments on the basic concepts along with the proofs of the results.

2 Basic concepts

This section recalls some useful definitions and introduces the accompanying notation; the reader unfamiliar with these concepts may like to consult section 6.1 where (s)he will also find some explanatory comments.

Let $(M, \mathcal{M}, P)$ be an abstract probability space; a sub-σ-field of $\mathcal{M}$ is denoted by $\mathcal{X}_i$ and its measurable completion by $\overline{\mathcal{X}_i}$. $[\mathcal{X}_i]_p^+$ denotes the set of non-negative $\mathcal{X}_i$-measurable and p-integrable functions, with $p \in [1, \infty]$, and p (resp. $+$) is dropped when denoting all non negative (resp. p-integrable) functions.

Definition 2.1 Let $p \in [1, \infty]$; $\mathcal{X}_1$ is *p-strongly identified by* $\mathcal{X}_2$ if $\forall\, f_1 \in [\mathcal{X}_1]_p$ 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* $\mathcal{X}_3$ if $\mathcal{X}_1 \vee \mathcal{X}_3$ is p-strongly identified by $\mathcal{X}_2 \vee \mathcal{X}_3$. In graphical figures and in the formal section, p-strong and conditional p-strong identification are denoted as $\mathcal{X}_1 \ll_p \mathcal{X}_2$ or $\mathcal{X}_1 \ll_p \mathcal{X}_2 \mid \mathcal{X}_3$, respectively.

Remark that, $\mathcal{X}_1$ p-strongly identified by $\mathcal{X}_2$ conditionally on $\mathcal{X}_3$ implies

$\mathcal{X}_1$ is q-strongly identified by $\mathcal{X}_2$ conditionally on $\mathcal{X}_3$ $\forall\, p \leq q \leq \infty$, since $[\mathcal{X}_1 \vee \mathcal{X}_3]_q \subset [\mathcal{X}_1 \vee \mathcal{X}_3]_p$ once $p \leq q$.

Definition 2.2 The sub-σ-field $\mathcal{X}_1$ *is weakly identified by the sub-σ-field* $\mathcal{X}_2$ if $\sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_1) \ : \ f_2 \in [\mathcal{X}_2]^+\} = \mathcal{X}_1$; and $\mathcal{X}_1$ *is weakly identified by* $\mathcal{X}_2$ *conditionally on* $\mathcal{X}_3$ if $(\mathcal{X}_1 \vee \mathcal{X}_3)$ is weakly identified by $(\mathcal{X}_2 \vee \mathcal{X}_3)$. In graphical figures and in the formal section, weak and conditional weak identification are denoted as $\mathcal{X}_1 \prec \mathcal{X}_2$ or $\mathcal{X}_1 \prec \mathcal{X}_2 \mid \mathcal{X}_3$, respectively.

Definition 2.3 Let $\{\mathcal{F}_i\}$ and $\{\mathcal{G}_i\}$ two sequences of sub-σ-fields of $\mathcal{M}$, and let $\mathcal{M}_1$ be a sub-σ-field of $\mathcal{M}$. The sequence $\{\mathcal{F}_i\}$ is said to be $\mathcal{M}_1$*-sifted by* $\{\mathcal{G}_i\}$ if the following two conditions holds:

$$\begin{aligned} &\text{(i)} \quad \underset{1 \leq i \leq n}{\perp\!\!\!\perp} \mathcal{F}_i \mid \mathcal{G}_1^n \vee \mathcal{M}_1 \quad \forall\, n \geq 1, \\ &\text{(ii)} \quad \mathcal{F}_i \perp\!\!\!\perp \mathcal{G}_1^n \mid \mathcal{G}_i \vee \mathcal{M}_1 \ \ \forall\, i \leq n, \ \forall\, n \geq 1. \end{aligned} \tag{2.1}$$

Property (2.1.i) is a condition of mutual independence, whereas property (2.1.ii) is called an *allocation* condition. When $\mathcal{M}_1$ is the trivial σ-field, one simply says that $\{\mathcal{F}_i\}$ is sifted by $\{\mathcal{G}_i\}$.

3 Main results

Four theorems constitute the main results of this papers: the first two theorems deal with the relationships between the identification of individual submodels and the identification of the corresponding global model; the third one deals with the identification of the statistical model, and the fourth one with the predictive model.

For the latent marginal process generating $\mathcal{H}_1^n$ the following simple facts should be singled out: (a) it follows from (1.10.i) that $\mathcal{H}_i \perp\!\!\!\perp \mathcal{S} \mid \mathcal{S}_1$ for $i = 1, \ldots, n$, that is, $\mathcal{S}_1$ is sufficient for each marginal individual process generating $\mathcal{H}_i$; (b) the complete latent marginal model generating $\mathcal{H}_1^n$ is *fully* characterized by the n individual latent marginal models generating $\mathcal{H}_i$ under the condition

$$\underset{1 \leq i \leq n}{\perp\!\!\!\perp} \mathcal{H}_i \mid \mathcal{S}_1 \tag{3.1}$$

(which in turn implies property (1.11)); (c) if $\mathcal{S}_1$ is weakly (resp. p-strongly, with $p \in [1, \infty]$) identified by $\mathcal{H}_i$ for at least one $i \in \{1, \ldots, n\}$ then $\mathcal{S}_1$ is weakly (resp. p-strongly) identified by $\mathcal{H}_1^n$; note that this result follows *without* the additional assumption (3.1).

3.1 Global identification in the presence of incidentals

In order to obtain similar results for the complete conditional process generating $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$, the following sifting condition is introduced:

$$\{\mathcal{Y}_i\} \text{ is } \mathcal{S}_2\text{-sifted by } \{\mathcal{H}_i\}. \tag{3.2}$$

Next proposition shows that, under the sifting condition (3.2), the sufficiency of the structural parameter $\mathcal{S}_2$ for the global model relative to a sample of size n implies the sufficiency of the same $\mathcal{S}_2$ for the submodels corresponding to any subsample of individuals in $I \subset \{1, \ldots, n\}$; its proof may be found in section 6.2.

Proposition 3.1 *Let $I \subset \{1, \ldots, n\}$ with $n \geq 1$. Then, under the sifting condition (3.2), $\mathcal{S}_2$ (resp. $\mathcal{H}_I \vee \mathcal{S}_2$) is a sufficient parameter in the model generating $(\mathcal{Y}_I \mid \mathcal{H}_I)$ (resp. $\mathcal{Y}_I$), where $\mathcal{Y}_I = \bigvee_{i \in I} \mathcal{Y}_i$, and similarly for $\mathcal{H}_I$; that is, under the sifting condition (3.2), condition (1.10.ii) implies*

$$\mathcal{Y}_I \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, I \subset \{1, \ldots, n\}. \tag{3.3}$$

The analogue of fact (a) above obtains by using Proposition 3.1. Indeed, let $n \geq 1$ and $I \subset \{1, \ldots, n\}$; then this proposition implies that, under the sifting condition (3.2), the sufficiency of $\mathcal{S}_2$ (resp. of $\mathcal{S}_2 \vee \mathcal{H}_I$) for the model generating $(\mathcal{Y}_I \mid \mathcal{H}_I)$ (resp. $\mathcal{Y}_I$) for some $I \subset \{1, \ldots, n\}$ implies the sufficiency of $\mathcal{S}_2$ (resp. of $\mathcal{S}_2 \vee \mathcal{H}_i$) for each individual submodel generating $(\mathcal{Y}_i \mid \mathcal{H}_i)$ (resp. $\mathcal{H}_i$) with $i \in I$.

The analogue of fact (b) above is straighforward under Proposition 3.1. The analogue of fact (c) above is obtained in Theorems 3.1 and 3.2 which show that, under the sifting assumption (3.2), the identification of the individual submodels is a sufficient condition for the identification of the global model; their proofs may be found in section 6.3 and 6.4.

Theorem 3.1 *Under the sifting condition (3.2), for any $p \in [1, \infty]$ and for any $I \subset \{1, \ldots, n\}$:*

(i) *If, for any* $i \in I$, $\mathcal{H}_i \vee \mathcal{S}_2$ *is p-strongly identified by* $\mathcal{Y}_i$, *then* $\mathcal{H}_I \vee \mathcal{S}_2$ *is p-strongly identified by* $\mathcal{Y}_I$.

(ii) *If, for any* $i \in I$, $\mathcal{S}_2$ *is p-strongly identified by* $\mathcal{Y}_i$ *conditionally on* $\mathcal{H}_i$, *then* $\mathcal{S}_2$ *is p-strongly identified by* $\mathcal{Y}_I$ *conditionally on* $\mathcal{H}_I$.

Theorem 3.2 *Under the sifting condition (3.2), for any* $I \subset \{1, \ldots, n\}$*:*

(i) *If, for any* $i \in I$, $\mathcal{H}_i \vee \mathcal{S}_2$ *is weakly identified by* $\mathcal{Y}_i$, *then* $\mathcal{H}_I \vee \mathcal{S}_2$ *is weakly identified by* $\mathcal{Y}_I$.

(ii) *If, for any* $i \in I$, $\mathcal{S}_2$ *is weakly identified by* $\mathcal{Y}_i$ *conditionally on* $\mathcal{H}_i$, *then* $\mathcal{S}_2$ *is weakly identified by* $\mathcal{Y}_I$ *conditionally on* $\mathcal{H}_I$.

Remark 3.1 Theorems 3.1 and 3.2 may also be interpreted as follows: if a sufficient parameter $\mathcal{S}_2$ is minimal sufficient (resp. p-complete) in each conditional experiment

$$\mathcal{E}_i^{\mathcal{H}_i} \;=\; (S \times V_1 \times V_2, \mathcal{S} \vee \mathcal{H}_i \vee \mathcal{Y}_i, \Pi_i^{\mathcal{H}_i}) \qquad i = 1, \ldots, n,$$

where $\Pi_i^{\mathcal{H}_i}$ is a conditional probability of Π_i given $\mathcal{H}_i$, then it is minimal sufficient (resp. p-complete) in the global conditional experiment defined by a conditional probability of $\Pi_{[n]}$ given $\mathcal{H}_1^n$. Similarly, if a sufficient parameter $\mathcal{S}_2 \vee \mathcal{H}_i$ is minimal sufficient (resp. p-complete) in each marginal experiment

$$\mathcal{E}_{\mathcal{Y}_i} \;=\; (S \times V_1 \times V_2, \mathcal{S} \vee \mathcal{H}_i \vee \mathcal{Y}_i, \Pi_{\mathcal{Y}_i}) \qquad i = 1, \ldots, n,$$

where $\Pi_{\mathcal{Y}_i}$ is the restriction of Π_i on $\mathcal{Y}_i$, then $\mathcal{S}_2 \vee \mathcal{H}_1^n$ is minimal sufficient (resp. p-strong) in the global marginal experiment defined by a marginal probability obtained after restricting $\Pi_{[n]}$ on $\mathcal{Y}_1^n$.

Remark 3.2 Let $p \in [1, \infty]$ and $I \subset \{1, \ldots, n\}$; since the p-strong identification implies the weak one, the identification conditions of Theorem 3.1 are related with that of Theorem 3.2. The following diagram, in the form of a directed graph, summarizes them, in which each "arrow" represents an "implication"; arrows without explicit reference to theorems follow from general properties of weak and strong identification, which can be found in Florens *et al.* (1990, chapters 4 and 5). Note that condition $\{\mathcal{H}_i \vee \mathcal{S}_2 \ll_p \mathcal{Y}_i \text{ for } i \in I\}$ is accordingly the strongest one, whereas condition $\mathcal{S}_2 \prec \mathcal{Y}_I \mid \mathcal{H}_I$ is the weakest one.

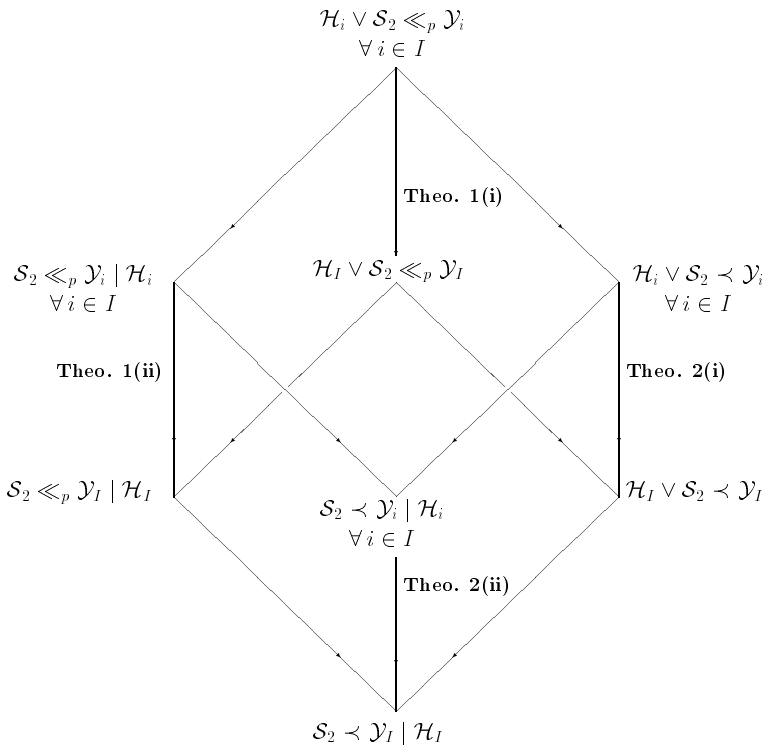

Fig. 1 Global-Individual Identification Relationships

Remark 3.3 The relationships between Theorems 3.1 and 3.2 with the results established in Landers (1972, 1974) and Landers and Rogge (1976) are of interest. Landers (1972) showed that the family of all product measures, say $\times_{i\in I} P_i$ with $P_i \in \mathcal{P}_i$ and $\mathcal{P}_i$ a family of probability measures defined on basic sets Y_i, admits a minimal sufficient σ-field if and only if each family $\mathcal{P}_i$ admits a minimal sufficient σ-field. Moreover, Landers also established that the minimal sufficient σ-field of the product family is generated by (the cylinders based on) each minimal sufficient σ-fields of the component families; see also Landers (1974) for a similar result concerning a minimal sufficient statistics. Landers and Rogge (1976) showed also that if each component of a family of probability measures is 1-complete

(resp. boundedly complete) then the product family is 1-complete (resp. boundedly complete). Theorems 3.1(i) and 3.2(i) establish similar results in the context of a sequence of conditional experiments, which in turn means that the minimal sufficient (resp. p-complete) σ-fields of each individual conditional experiment are mutually *different*. Similarly, Theorems 3.1(ii) and 3.2(ii) establish those results in the context of marginal experiments in which the minimal sufficient (resp. p-complete) σ-fields have eventually an incidentals character.

3.2 Identification of the statistical model

The model specification developed in section 1 implies that the *statistical model*, bearing on the manifest variables as well as the structurals only, is a mixture obtained after integrating out the unobservable incidentals $\mathcal{H}_I$ for $I \subset \{1, \ldots, n\}$; this mixture model provides us with $\mathbb{P}\{A \mid \mathcal{S}_1 \vee \mathcal{S}_2\}$ for $A \in \mathcal{Y}_I$. For any $I \subset \{1, \ldots, n\}$, the hypothesis

(C1) $\mathcal{S}_1$ is weakly identified by $\mathcal{H}_i$ for (at least one) $i \in I$, and $(\forall\, i \in I)$ $\mathcal{H}_i \vee \mathcal{S}_2$ is 2-strongly identified by $\mathcal{Y}_i$,

represents identification conditions for individual submodels that, under the sifting assumption (3.2), are sufficient for the identification of the statistical model. More precisely, we have:

Theorem 3.3 *Let $I \subset \{1, \ldots, n\}$. Then, under the sifting assumption (3.2), condition* **(C1)** *implies that $\mathcal{S}_1 \vee \mathcal{S}_2$ is weakly identified by $\mathcal{Y}_I$.*

For a proof, see section 6.5. It should be noticed that Theorem 3.3 does not impose any condition of independence for the model generating $\mathcal{H}_1^n$.

3.3 Minimal predictive sufficiency

Let us now consider the minimal sufficiency (or, weak identification) problem in the predictive model in case of a mutually independent statistical model, namely

$$\perp\!\!\!\perp_{1 \le i \le n} \mathcal{Y}_i \mid \mathcal{S}_1 \vee \mathcal{S}_2 \qquad \forall\, n \ge 1. \tag{3.4}$$

This condition is implied by the sifting condition (3.2) under the additional assumption (3.1); see Florens *et al.* (1990, Theorem 7.6.9). For $m, n \geq 1$, write $\mathcal{Y}_1^n$ for the "past" information and $\mathcal{Y}_{n+1}^{n+m}$ for the "future" information; condition (3.4) is equivalently rewritten as

$$\mathcal{Y}_1^n \perp\!\!\!\perp \mathcal{Y}_{n+1}^{n+m} \mid \mathcal{S}_1 \vee \mathcal{S}_2 \qquad \forall\, m, n \geq 1. \tag{3.5}$$

The following lemma establishes, for the predictive model generating $(\mathcal{Y}_{n+1}^{m+n} \mid \mathcal{Y}_1^n)$, a relationship between the minimal *predictive* sufficient statistics based on the "past" and the minimal sufficient statistics based on the "past"; its proof may be found in section 6.6.

Lemma 3.1 *For the predictive model generating* $(\mathcal{Y}_{n+1}^{m+n} \mid \mathcal{Y}_1^n)$*, condition (3.5) implies*

$$\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{Y}_{n+1}^{n+m}]^+\} \subset \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{S}_1 \vee \mathcal{S}_2]^+\}. \tag{3.6}$$

Relation 3.6 means that the part of the "past" which is actually informative on the "future" (*i.e.* the minimal *predictive* sufficient statistics based on the "past") is contained in the minimal sufficient statistic based on the "past". Nevertheless, a best possible solution, with equality rather than inclusion, may be obtained by considering the following condition:

(C2) $\mathcal{S}_1$ is 2-strongly identified by $\mathcal{H}_i$ for (at least one) $i \in I$, and $(\forall\, i \in I)$ $\mathcal{H}_i \vee \mathcal{S}_2$ is 2-strongly identified by $\mathcal{Y}_i$,

for any $I \subset \{1, \ldots, n+m\}$; such a result, established in next theorem, is actually a corollary of both Theorem 3.1(i) and Theorem 5.4.14 in Florens *et al.* (1990):

Theorem 3.4 *Under the sifting condition (3.2), the mutual independence condition (3.1) and condition* **(C2)***, the (past) sample information used for predictive purpose is the same as the (past) sample information used for the inference on the structural parameters, i.e.*

$$\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{Y}_{n+1}^{n+m}]^+\} = \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{S}_1 \vee \mathcal{S}_2]^+\}. \tag{3.7}$$

Remark 3.4 Under the sifting property (3.2), condition **(C2)** implies that $\mathcal{S}_1 \vee \mathcal{S}_2$ is 2-strongly identified by $\mathcal{Y}_I$; for a proof, see Mouchart and San Martín (1999, Theorem 3 (ii)). Furthermore, **(C2)** implies condition **(C1)**. Thus, under the sifting condition (3.2), a relationship between Theorems 3.3 and 3.4 may be depicted, for any $I \subset \{1, \ldots, n+m\}$, as follows:

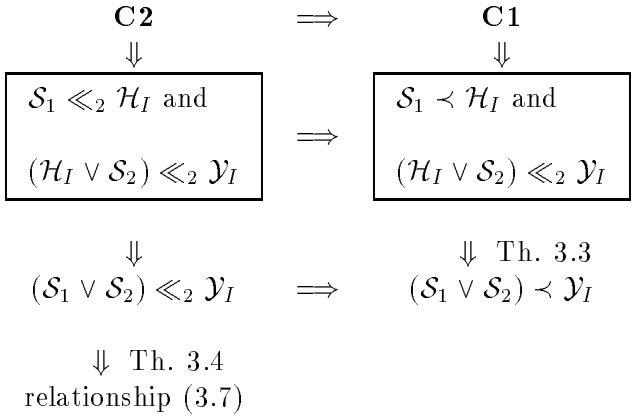

Fig. 2 Relationships between Theorems 3.3 and 3.4

4 Identification problems in item response models

This section illustrates a possible use of the previous results for the specification and the identification analysis of a general class of item response models (IRT). The first subsection specifies the general structure of this class of models by using both sifting conditions and an hypothesis of exchangeability. The second subsection gives sufficient conditions for the identification of the statistical model; subsections 4.2.2 and 4.2.3 apply previous results to the psychomotor assesment model and to the Rasch model, respectively.

4.1 A general specification of item response models

4.1.1 A usual specification

An item response model may be specified as follows: suppose that each of n individuals, labelled $i = 1, \ldots, n$, are exposed to m items (for instance,

questions to be answered, problems to be solved, etc.), labelled $j = 1, \ldots, m$. For each subject i and each item j, an answer y_{ij} is recorded: for example, in polytomous models, y_{ij} takes one of l possible values for each item j; thus a disjunctive coding takes the form $y_{ij} = (y_{ij1}, \cdots, y_{ijk}, \ldots, y_{ijl}) \in \{0,1\}^l$ with $\sum_k y_{ijk} = 1 \ \forall (i,j)$ (see *e.g.* Andersen, 1973a, p. 142; 1980, section 6.10); or, in the field of psychomotor assesment, y_{ij} represents the number of (say) sit-ups or push-ups completed in a minute by the individual i after carrying out the task j, in such a case $y_{ij} \in \mathbb{N}$ (see *e.g.* Hambleton *et al.*, 1991). This class of models assumes that the probability distribution of the answer y_{ij} depends both on some *individual parameter* η_i (representing individual i's ability in solving or carrying out problems or tasks) and on some *item parameter* s_j (representing the difficulty of problem or task j); both η_i and s_j are unobservable, whereas y_{ij} is observable. It is also assumed that "given the values of the parameters, all answers are stochastically independent" (Rasch, 1966a, p. 50; see also Lindsay *et al.*, 1991, p. 97 or Pfanzagl, 1994, p. 251). For comments and details, see *e.g.* Rasch (1966a, b), Lord and Novick (1968), Andersen (1973b, 1980), Tjur (1982), and Fischer (1995); for general reviews, see Andersen (1982) and Clogg (1992).

The usual parametric specification of this class of models is often made *individual-wise* by modelling the distribution of $(y_{ij} \mid \eta_i, s_j)$ and/or that of (η_i, s_j); the semiparametric as well as the nonparametric versions of IRT models are also built individual-wise; see *e.g.* Holland and Rosenbaum (1986) and Holland (1990, pp. 591-592). Recently, the monotonicity of the distribution of $(y_{ij} \mid \eta_i, s_j)$ has been introduced, in the specification of this class of models, as a nonparametric hypothesis; for details, see Junker (1993), Junker and Ellis (1997) and Ellis and Junker (1997). Martin-Löf (1974) and Lauritzen (1984) specify such a class of models by using the concept of *repetitive structure*.

In the next subsections, this class of models is specified progressively; two reasons may be given in order to motivate such a strategy. On the one hand, the complexity of the hypotheses introduced in this class of models requires a careful sequencing of the hypotheses in order to allow for a contextual interpretation of each hypothesis. On the other hand, a sequential introduction of the hypotheses permits a progressive analysis of the identification problems arising in this class of models. The global model generates $\{(y_{ij}, \eta_i, s_j) : i = 1, \ldots, n, \ j = 1, \ldots, m\}$; in line with section 3, the observable incidentals are denoted as

$$\mathcal{Y}_i = \bigvee_{1 \le j \le m} \mathcal{Y}_{ij}, \qquad \text{where } \mathcal{Y}_{ij} = \sigma(y_{ij}), \tag{4.1}$$

whereas the latent incidentals and the structural are denoted as

$$\begin{aligned} \mathcal{H}_i &= \sigma(\eta_i), \\ \mathcal{S} &= \bigvee_{1 \le j \le m} \mathcal{S}_j, \qquad \text{where } \mathcal{S}_j = \sigma(s_j). \end{aligned} \tag{4.2}$$

The (global) model generating $\mathcal{Y}_1^n \vee \mathcal{H}_1^n$ is specified in two steps: the first one deals with the specification of the (global) conditional model generating $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$, and the second one deals with the (global) unobservable marginal model generating $\mathcal{H}_1^n$.

4.1.2 Sifting conditions for the conditional model

The general structure of the model generating the observables conditionally on the non observables is typically based on sifting conditions. Indeed, the $\mathcal{H}_i$'s may be considered as incidentals since each of them is a characteristic of an individual i only, whereas $\mathcal{S}$ may be considered as a structural parameter since it is assumed to be *constant* over individuals (see Lord and Novick, 1968, p. 328, or Lindsay *et al.*, 1991, p. 97), that is, all individuals answer the *same* test characterized by a difficulty parameter $\mathcal{S}$ assumed common to all of them; for some comments on this terminology, see also Andersen (1973a, p. 143) or Colonius (1977). This features actually leads to a double sifting structure: one among the individuals, another one among the items for each individual. Thus the model rests on the following conditions:

$$\begin{aligned} &\text{(i) } \{\mathcal{Y}_i\} \text{ is } \mathcal{S}\text{-sifted by } \{\mathcal{H}_i\}, \\ &\text{(ii) } (\forall\, i = 1, \ldots, n)\ \{\mathcal{Y}_{ij}\} \text{ is } \mathcal{H}_i\text{-sifted by } \{\mathcal{S}_j\}. \end{aligned} \tag{4.3}$$

This double sifting structure constitutes the basic framework within which is specified the conditional model generating $(\mathcal{Y}_1^n \mid \mathcal{H}_1^n)$; moreover, such a global conditional model may, under this double sifting structure, be specified individual-wise: in such a way, the conditional model is *fully* characterized by the respective individual processes.

Contextually, the allocation condition corresponding to property (4.3.i) (*i.e.* $\mathcal{Y}_i \perp\!\!\!\perp \mathcal{H}_1^n \mid \mathcal{S} \vee \mathcal{H}_i$) formalizes the assumption that each subject parameter characterizes one individual only, whereas the allocation condition corresponding to property (4.3.ii) (*i.e.* $\mathcal{Y}_{ij} \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_i \vee \mathcal{S}_j$ for $j = 1, \ldots, m$) means that each individual i answers the m items of the test with its same ability $\mathcal{H}_i$, and therefore this class of models excludes, for instance, a "learning process" of an individual i obtained after answering the first j items and used in order to answer the next $j+1$ item; thus, the individual i's ability $\mathcal{H}_i$ along with the difficulty parameter $\mathcal{S}_j$ characterizes the generation of the answer $\mathcal{Y}_{ij}$ to item j. Condition (4.3.ii) may also be considered either as a formal version of the axiom of local independence in a very general (or nonparametric) context –for details about this axiom, see *e.g.* Lazarsfeld (1966), Lord and Novick (1968), Novick (1979), Holland and Rosenbaum (1980), Bartholomew (1987), Sobel (1997)–, or as a nonparametric version of the "specific objectivity axiom" –for details about this axiom, see Fischer (1995, axiom V) and Irtel (1995).

4.1.3 Exchangeability of the non observable incidentals

The marginal model generating the ability incidentals $\mathcal{H}_1^\infty$ is specified by introducing an hypothesis of exchangeability. Indeed, in such a process a mutual independence condition between the η_i's is often introduced; see *e.g.* Lord and Novick (1968, chapters 22 and 23) or the references given in Lindsay *et al.* (1991). Nevertheless, an assumption of exchangeability may be preferred not only because such an assumption is weaker than the mutual independence one, but also because the exchangeability may be contextually more desirable: "Suppose that prior to estimating the abilities of examinees in a group, we are willing to make the assumption that the information regarding the ability of any one examinee is not different from that of any other examinee; i.e., the information on examinees is *exchangeable*. This assumption implies that the abilities ... may be considered as a random sample from a population" (Hambleton and Swaminathan, 1985, p. 92).

Formally, let $\eta = \{\eta_i : i \geq 1\}$ be the infinite sequence of non observable abilities defined on a basic space, say $(M, \mathcal{M})$, with values in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{(\mathbb{N})})$; thus $\mathcal{H}_i = \sigma(\eta_i) = \eta_i^{-1}(\mathcal{B})$ for $i \geq 1$. In order to make precise the assumption that η is exchageable, let π be a finite permutation of $\mathbb{N}$, that is, π is a bijection of $\mathbb{N}$ such that $(\exists\, k < \infty)\ \pi(n) = n\ \forall\, n \geq k$. The corresponding finite permutation operator π of $\mathbb{R}^{\mathbb{N}}$ is defined by

$$\pi(u) \;=\; [u_{\pi(n)}]$$

for any $u = [u_n] \in \mathbb{R}^{\mathbb{N}}$. The set of all finite permutation operators relative to a given k $(< \infty)$ is denoted by Σ_k; $\Sigma = \cup_{k\in\mathbb{N}} \Sigma_k$ is the set of all finite permutation operators; endowed with the product of composition, Σ is a group of measurable transformations on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{(\mathbb{N})})$. Let now the group Σ act by composition on the process η, *i.e.* for any $\pi \in \Sigma$,

$$\pi \circ \eta \;=\; \pi(\eta) \;=\; [\eta_{\pi(n)}].$$

Thus $\pi \circ \eta$ is the transformed stochastic process obtained by permuting, according to π, the first k coordinates of the stochastic process η, for some finite k. The *σ-field of symmetric sets* in $\mathbb{R}^{\mathbb{N}}$ is defined then as

$$\mathcal{B}_\Sigma \;=\; \{B \in \mathcal{B}^{(\mathbb{N})} \,:\, \pi^{-1}(B) \;=\; B \quad \forall\, \pi \in \Sigma\},$$

and the *symmetric events of the stochastic process* η is therefore defined as $\mathcal{H}_\Sigma = \eta^{-1}(\mathcal{B}_\Sigma) \subset \mathcal{M}$. Thus, the process η is assumed to be exchangeable if

$$\mathbb{E}(f \circ \pi \circ \eta) \;=\; \mathbb{E}(f \circ \eta) \qquad \forall\, \pi \in \Sigma \quad \forall\, f \in [\mathcal{B}^{(\mathbb{N})}]^+.$$

By the De Finetti's Theorem, it follows that η is exchangeable if and only if η is $\mathcal{H}_\Sigma$-i.i.d, that is

$$\begin{aligned} &\perp\!\!\!\perp_{1\le i\le n} \mathcal{H}_i \mid \mathcal{H}_\Sigma \qquad \forall\, n \ge 1, \\ &\mathbb{E}[g(\eta_i) \mid \mathcal{H}_\Sigma] \;=\; \mathbb{E}[g(\eta_1) \mid \mathcal{H}_\Sigma] \qquad \forall\, i \ge 1 \quad \forall\, g \in [\mathcal{B}]^+. \end{aligned} \tag{4.4}$$

For details about this theorem, see *e.g.* Hewitt and Savage (1955), Dellacherie and Meyer (1980, chapter 5, ns.47-52), Aldous (1985) or Chow and Teicher (1988, section 7.3). For a conditional version of this theorem, see Florens *et al.* (1990, Theorem 9.3.11); see also Diaconis *et al.* (1992) for a discussion concerning finite versions of the De Finetti's Theorem.

Let G be the conditional probability of $(\eta_1 \mid \mathcal{H}_\Sigma)$ and let us to assume the existence of a regular version of this conditional probability; this is a random probability measure on $(\mathbb{R}, \mathcal{B})$ defined as

$$G(B) = \mathbb{P}\{\eta_1 \in B \mid \mathcal{H}_\Sigma\}, \qquad B \in \mathcal{B}. \tag{4.5}$$

Thus, $G(B) \in [\mathcal{H}_\Sigma]^+ \ \forall\, B \in \mathcal{B}$. The following lemma establishes that the σ-field generated by $\{G(B) : B \in \mathcal{B}\}$ coincide with the σ-field $\mathcal{H}_\Sigma$ of the symmetric events of the stochastic process η; its proof may be found in section 6.7.

Lemma 4.1 *Let* $\mathcal{G} = \sigma\{G(B) : B \in \mathcal{B}\}$. *Then* $\mathcal{G} = \mathcal{H}_\Sigma$.

Using this lemma, properties (4.4) may be equivalently rewritten as

$$\underset{1 \le i \le n}{\perp\!\!\!\perp} \eta_i \mid \mathcal{G} \quad \text{and} \quad (\eta_i \mid \mathcal{G}) \sim \mathcal{G} \quad \forall\, i = 1, \ldots, n \qquad \forall\, n \ge 1, \tag{4.6}$$

and therefore the structural parameter corresponding to the marginal model generating $\{\mathcal{H}_1^n : n \ge 1\}$ is given by the random probability measure $\mathcal{G}$. Note finally that the exchangeability of unobservable abilities is a reasonable assumption once it may be assumed that the collected information about the abilities of the individuals does not depend on the order in which individuals are tested.

4.1.4 A semiparametric specification of IRT models

In the sequel, a semiparametric specification of the IRT model is developed by using both the double sifting structure (4.3) and the exchangeability of the stochastic process η. Two particular cases are considered, namely the psychomotor assesment model and the Rasch model.

For a sample of size n, let $c_i = (c_{i1}, \ldots, c_{im}) \in V^m$, with $V \subset \mathbb{R}$, denote a possible realization of y_i $(i = 1, \ldots, n)$, $\eta_1^n = (\eta_1 \cdots \eta_i \cdots \eta_n)$ and $s = (s_1 \cdots s_j \cdots s_m)$, with $\eta_i, s_j \in \mathbb{R}$; thus $\mathcal{H}_1^n = \sigma(\eta_1^n)$ and $\mathcal{S} = \sigma(s)$. The double sifting structure (4.3) provides a formal justification for the usual form of the so-called conditional likelihood:

$$\mathbb{P}(y_1 = c_1, \ldots, y_n = c_n \mid \eta_1^n, s) = \prod_{i=1}^{n} \prod_{j=1}^{m} p(y_{ij} = c_{ij} \mid \eta_i, s_j). \tag{4.7}$$

In the field of psychomotor assesment, y_{ij} is typically assumed to be distributed according to a Poisson distribution of parameter λ_{ij}, with

$$\lambda_{ij} \;=\; \exp(\eta_i - s_j) \qquad \forall\,(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,m\}; \qquad (4.8)$$

in this case, $V = \mathbb{N}$; see *e.g.* Hambleton *et al.* (1991). In the Rasch, or binary IRT, model (*i.e.* $l = 1$), the random variable is a binary coding of an erroneous answer of the individual i to item j; in this case, it is assumed that $(y_{ij} \mid \lambda_{ij}) \sim \text{Bernoulli}\,(\lambda_{ij})$, where

$$\lambda_{ij} \;=\; \frac{\exp(\eta_i - s_j)}{1 + \exp(\eta_i \,-\, s_j)} \qquad \forall\,(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,m\}; \qquad (4.9)$$

in this case, $V = \{0,1\}$; see *e.g.* Birnbaum (1968, chapter 1967). These definitions of the parameter λ_{ij} are based on an hypothesis of "separability" between the ability parameter and the item difficulty parameter and constitutes a main characteristic of this class of models; for details, see Rasch (1966a, b) or Lord and Novick (1968).

In σ-algebraic terms, (4.8) as well as (4.9) implies that $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$ and moreover either the Poisson-distribution hypothesis or the Bernoulli-distribution hypothesis implies, for each $i = 1,\ldots,n$, the sufficiency of λ_{ij}; more precisely, for $i = 1,\ldots,n$,

$$\mathcal{Y}_{ij} \;\perp\!\!\!\perp\; \mathcal{H}_i \vee \mathcal{S}_j \mid \sigma(\lambda_{ij}) \qquad \forall\, j = 1,\ldots,n. \qquad (4.10)$$

Thus, in σ-algebraic terms, the Rasch model is very similar to the psychomotor assesment model; their difference will be appreciated after analysing the weak identifiability of the corresponding semiparametric models.

It is known that the model (4.7), along with the property $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$, is not (weakly) identified. Indeed, for any constant $c \in \mathbb{R}$, it follows that $\eta_i - s_j = \eta_i^* - s_j^*$, with $\eta_i^* = \eta_i - c$ and $s_j^* = s_j - c$; this arbitrariness is removed by setting a linear restriction on the item parameters s; see *e.g.* Lazarsfeld and Henry (1968, chapter 3), Andersen (1980, p. 243), Lindsay *et al.* (1991, p. 97), Hambleton *et al.* (1991, pp. 41-42, 126-129), Jansen and van Duijn (1992, p. 405), Jansen (1994, p. 321) and Ghosh (1995, p. 165).

When estimating the item parameters s from the conditional likelihood (4.7), the maximum likelihood estimator is known to be, in general, inconsistent, when $n \to \infty$, because of the non observable incidentals $\mathcal{H}_1^n$; see *e.g.* Neyman and Scott (1948), Andersen (1973a), Haberman (1977), McCullagh

and Nelder (1989, p. 211), Agresti (1993) or Ghosh (1995). An alternative strategy is to consider a mixture model obtained after integrating out the unobservable incidentals $\mathcal{H}_1^n$; see *e.g.* Lindsay *et al.* (1991). Reminding that the identifiability of the structural parameter is a necessary condition for the existence of a consistent estimator, we shall, in the sequel, check that identifiability in a semi-parametric form of the IRT model, which implies, under both the double sifting property (4.3) and the exchangeability of the ability parameters η_i's, a statistical model of the following form:

$$\begin{aligned} \mathbb{P}(y_1 = c_1, \ldots, y_n = c_n \mid G, s) \; = \\ = \int \prod_{i=1}^{n} \prod_{j=1}^{m} p(y_{ij} \; = \; c_{ij} \mid \eta_i, s_j) \, G(d\eta_1^n). \end{aligned} \tag{4.11}$$

4.2 Identification of the structural parameter in a semiparametric IRT model

This section analyses the identification of the structural parameter $\mathcal{S} \vee \mathcal{G}$ by the observations $\mathcal{Y}_1^n$ in the statistical model (4.12) in both the psychomotor assesment case and the Bernoulli case. Since in σ-algebraic terms these two models are very similar, the identification analysis starts by concentrating the attention on those similar aspects leaving particular analytic features for separate treatement.

4.2.1 Progressive analysis of the identifiability

The identification of the structural parameter $\mathcal{S} \vee \mathcal{G}$ may be obtained in a progressive way by analysing the identifiability of the component submodels. Indeed, Theorem 3.3 ensures that, under both the sifting condition (4.3.i) and the conditional i.i.d condition (4.6), the weak identification of the structural parameter $\mathcal{S} \vee \mathcal{G}$ by the observations $\mathcal{Y}_1^n$ may be obtained in two steps, namely

(i) that $\mathcal{G}$ be weakly identified by $\mathcal{H}_i$ for (at least one) $i \in \{1, \ldots, n\}$; and

(ii) that $\mathcal{S} \vee \mathcal{H}_i$ be 2-strongly identified by $\mathcal{Y}_i$ for each $i = 1, \ldots, n$.

Each of these steps may be obtained as follows:

• For step (i): by hypothesis, η_i's are $\mathcal{G}$-identically distributed; it follows that $\mathcal{G}$ is weakly identified by $\mathcal{H}_i$ since for each $i = 1, \ldots, n$ we have, from (4.5) and Lemma 4.1, that

$$G(B) = \mathbb{P}\{\eta_i \in B \mid G\} \quad \forall\, B \in \mathcal{B}, \quad \forall\, \eta_i \in [\mathcal{H}_i]^+$$

(for a similar argument, see *e.g.* Florens *et al.*, 1992, proof of Proposition 3.6).

Remark 4.1 This result, along with Lemma 4.1, shows that the σ-field $\mathcal{H}_\Sigma$ of symmetric events of the stochastic process η is a minimal sufficient one for the process generating $\mathcal{H}_i$.

• For step (ii): let $i \in \{1, \ldots, n\}$ and $p \in [1, \infty]$; by definition (4.1) of $\mathcal{Y}_i$ and (4.2) of $\mathcal{S}$, $\mathcal{S} \vee \mathcal{H}_i$ p-strongly identified by $\mathcal{Y}_i$ is equivalently rewritten as

$$\left(\bigvee_{1 \le j \le m} \mathcal{S}_j \right) \vee \mathcal{H}_i \text{ is } p\text{-strongly identified by } \left(\bigvee_{1 \le j \le m} \mathcal{Y}_{ij} \right). \qquad (4.12)$$

Taking into account that $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$, the following lemma establishes that, if $\sigma(\lambda_{ij})$ is p-strongly identified by $\mathcal{Y}_{ij}$ for each $j = 1, \ldots, m$, then, under both the sifting condition (4.3.ii) and the following identification condition **(IC)** on the item parameters s, condition (4.12) follows:

(IC) The vector s of item parameters lies Π-a.s. in a known $(m-1)$-dimensional subspace B of $\mathbb{R}^m$ such that B does not contain the main diagonal vector $i_m = (1, \ldots, 1)' \in \mathbb{R}^m$.

Lemma 4.2 *Assume the sifting condition (4.3.ii) and let $p \in [1, \infty]$. If $\sigma(\lambda_{ij})$ is p-strongly identified by $\mathcal{Y}_{ij}$ for any $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$, then, under condition* **(IC)**, *$\mathcal{S} \vee \mathcal{H}_i$ is p-strongly identified by $\mathcal{Y}_i$.*

For a proof of Lemma 4.2, see section 6.8.

Consequently, the p-strong identification of the submodel generating $(\mathcal{Y}_i \mid \mathcal{H}_i)$ is reduced to check that $\sigma(\lambda_{ij})$ is p-strongly identified by $\mathcal{Y}_{ij}$

$\forall\, j \in \{1, \ldots, m\}$, that is, that λ_{ij} is a parameter p-complete for the process generating $(y_{ij} \mid \lambda_{ij})$, where $(y_{ij} \mid \lambda_{ij}) \sim F^{\lambda_{ij}}$ with $F^{\lambda_{ij}}$ a known probability distribution depending on the sufficient parameter λ_{ij} defined by (4.10); in the psychomotor assesment case, $F^{\lambda_{ij}}$ corresponds to a Poisson distribution, whereas in the Rasch model case $F^{\lambda_{ij}}$ corresponds to a Bernoulli distribution. Before cheking the p-strong identification in these cases, we can establish the following general theorem dealing with the strong identification of the of the conditional model (4.7), and with the weak identification of the statistical model (4.12):

Theorem 4.1 *Assume the double sifting structure (4.3). Let $y_{ij} \in V$ for $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, m\}$. If $\{(y_{ij}, \lambda_{ij})\}$ satisfies the following conditions:*

(i) $(y_{ij} \mid \lambda_{ij}) \sim F^{\lambda_{ij}}$, *where* $\forall\, (i, j)\ \sigma(\lambda_{ij})$ *is a sufficient parameter p-strongly identified by $\mathcal{Y}_{ij}$ ($p \in [1, \infty]$) such that $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$;*

(ii) $s \in \mathbb{R}^m$ *satisfies condition* **(IC)**,

then

1. *$\mathcal{S} \vee \mathcal{H}_1^n$ is p-strongly identified by $\mathcal{Y}_1^n$ ($p \in [1, \infty]$), and therefore $\mathcal{S}$ is weakly identified by $\mathcal{Y}_1^n$.*

2. *If furthermore, the marginal process generating η is asumed to be exchangeable, $\mathcal{S} \vee \mathcal{G}$ is weakly identified by $\mathcal{Y}_1^n$ for any $n \geq 1$.*

Remark 4.2 This theorem deserve the following comments:

1. The identification results established in this theorem crucially depend on the condition (i); in the following subsections, such a condition is verified in the Poisson case as well as in the Bernoulli case.

2. Condition **(IC)** excludes the case of *constant difficulty* among items of a given test, namely $s_j = s_1$ for $j = 1, \ldots, m$. Furthermore, the identification restriction **(IC)** has a "structural character" since the σ-field $\mathcal{S}$ is a structural one. In other words, the a.s. linear restriction may be expressed as $a's = 0$, where $a \in B^{\perp}$ is the same for every individual.

3. If in the latent model generating η_1^n particular hypotheses, such as exchangeability, are not introduced, assertion (1) still ensures that the structural parameter $\mathcal{S}$ is weakly identified by the observations.

4. If the sufficient parameter λ_{ij} is a minimal one, then $\sigma(\lambda_{ij})$ is weakly identified by $\mathcal{Y}_{ij}$. Furthermore, under condition **(IC)**, the sifting property (4.3.ii) implies that $\{\mathcal{Y}_{ij}\}$ is sifted by $\{\sigma(\lambda_{ij})\}$ (see proof of Lemma 4.2). Consequently, by using Theorem 3.2(i), it follows that $\mathcal{S} \vee \mathcal{H}_i$ is weakly identified by $\mathcal{Y}_i$, which in turn implies, under the sifting property (4.3.i), that $\mathcal{S} \vee \mathcal{H}_1^n$ is weakly identified by $\mathcal{Y}_1^n$. Note finally that this last identification relationship is implied by, and is much weaker than, assertion (1) of Theorem 4.1, and implies also that the ability parameters are weakly identified by the observations conditionally on the item parameters s.

4.2.2 Identification of the structural parameters in the psychomotor assesment

In the field of psychomotor assesment, y_{ij} is typically assumed to be distributed according to a Poisson distribution of parameter λ_{ij}, where λ_{ij} is defined by (4.8). In this case, the following lemma provides the necessary qualification for obtaining condition (i) in Theorem 4.1; for a proof, see section 6.9.

Lemma 4.3 *Let $(y \mid \lambda) \sim$ Poisson(λ), where $\lambda \in \mathbb{R}_+$ and $y \in \mathbb{N}$. Then for any (prior) probability measure μ_λ defined on $(\mathbb{R}_+, \mathcal{B})$ and for any $p \in [1, \infty]$, the parameter λ is p-strongly identified by the observation y.*

Furthermore, if μ_λ is assumed to be a σ-finite measure only, this result is valid under an additional regularity condition on μ_λ; for details, see Remark 6.3 in section 6.9. Therefore, in the context of pyschomotor assesment, Theorem 4.1 implies the following corollary:

Corollary 4.1 *Assume the double sifting structure (4.3). Let $y_{ij} \in \mathbb{N}$ be the number of sit-ups or push-ups completed in an interval of time by the individual i after carrying out the task j. If $(y_{ij} \mid \lambda_{ij}) \sim$ Poisson (λ_{ij}), with $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$, where s satisfies condition* **(IC)** *and the stochastic process η is exchangeable, then $\mathcal{S} \vee \mathcal{G}$ is weakly identified by $\mathcal{Y}_1^n$ for any $n \geq 1$.*

Remark 4.3 Conclusion of Corollary 4.1 is valid for any prior probability measure $\mu_{\lambda_{ij}}$ on $(\mathbb{R}_+, \mathcal{B})$. Thus, the condition **(IC)** is the unique identification restriction imposed on $\mathcal{S}$ for the weak identification of the structural parameter $\mathcal{S} \vee \mathcal{G}$ by the statistical model. In other words, in the psychomotor assesment case, the identification restriction used in order to avoid overparametrization of the conditional likelihood (4.7) is sufficient to ensure the weak identification of the statistical model (4.12).

4.2.3 Identification of the structural parameter in the Rasch model

In a binary item response model, the random variable y_{ij} is assumed to be distributed according to a Bernoulli of parameter λ_{ij}, where λ_{ij} is defined by (4.9). In this case, the following Lemma provides the necessary qualification for checking condition (i) in Theorem 4.1; for a proof, see subsection 6.10.

Lemma 4.4 *Let* $(y \mid \lambda) \sim$ *Bernoulli*(λ)*, where* $\lambda \in [0,1]$*,* $y \in \{0,1\}$ *and* $(\lambda \mid a,b) \sim$ *Beta*(a,b) *with* $a,b \geq 1$*. Then, for any* $p \in [1,\infty]$*, the parameter* λ *is* p*-strongly identified by the observation* y*.*

With respect to the polytomous case, the arguments are similar to that used in the binary case, the difference being that the prior probability distribution is a Dirichlet one; details may be found in section 6.10. Thus, in the context of the Rasch model, Theorem 4.1 implies the following corollary:

Corollary 4.2 *Assume the double sifting structure (4.3). If* $(y_{ij} \mid \lambda_{ij}) \sim$ *Bernoulli* (λ_{ij})*, with* $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$*, where* s *satisfies condition* **(IC)***, the stochastic process* η *is exchangeable, and* $(\lambda_{ij} \mid a,b) \sim$ *Beta*(a,b) *with* $a,b \geq 1$*, then* $\mathcal{S} \vee \mathcal{G}$ *is weakly identified by* $\mathcal{Y}_1^n$ *for any* $n \geq 1$*.*

Remark 4.4 This corollary deserves the following comments:

1. The identification restriction **(IC)** is again the useful (general) condition introduced to avoid overparametrization of the corresponding conditional likelihood (4.7). Corollary 4.2 adds an additional identification restriction, namely condition (ii), for the identification of the statistical model (4.12).

2. In both the psychomotor assesment case and the Bernoulli case, the parameter λ_{ij} satisfies that $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$. Nevertheless, Corollaries 4.1 and 4.2 illustrate different techniques of proof, the first one

makes use of a general (prior) probability on λ_{ij}, whereas the second one makes use of a particular (beta or Dirichlet) specification.

5 Concluding remarks

A first issue considered in this paper addresses the general problem of developing a modelling strategy for the statistical analysis of individual data. The general framework relies on the contention that if one wants the basic statistical model, or maintained hypothesis in the parlance of hypothesis testing, to be elaborated with regards to field knowledge, then it should be decomposed into an ordered sequence of hypothesis simple enough to be endowed with as clear as possible a field interpretation. It is accordingly suggested that a first issue deals with the distinction between characteristics the dimension of which increases with the sample size and the characteristics representing common features shared by all individuals conforming a reference population. This is the root of the formalization of the two basic concepts of incidentals and structurals.

A second step of the suggested modelling strategy consists in decomposing the incidental information $\mathcal{I}_1^n$ given by a sample of size n into n pieces of information, $\mathcal{I}_i$, each of them being characteristic of one and only one individual. This individualization is pushed one step further by decomposing the individual $\mathcal{I}_i$ into two component $\mathcal{Y}_i$ and $\mathcal{H}_i$ with an allocation property in the form of a sifting condition; this sifting condition is the basic assumption which *implies* and gives substance to such a decomposition. The sifting condition is a basic assumption in the asymptotic theory of conditional models (for details, see *e.g.* Florens *et al.*, 1990, chapter 7); this paper exploits its *substantive meaning* in the specification of individual data; on the concept of "substantive hypothesis", see *e.g.* Cox (1990). More specifically, the sifting property is used in this paper to formalize a concept of individual as logically distinguishable from a collective and to provide results linking the identification for a sample of given size with the identification of individual submodels with incidental parameters.

A third step recognizes that once incidentals are introduced to formalize individual heterogeneity, part of them will be unobservable. The statistical model, dealing with observable incidentals only, is eventually built by integrating out the non observable incidentals and takes the form of a mixture model. This integration typically raises further identification problems.

The identification problem constitutes a second issue developed in this paper. Indeed, since a global conditional model is, under a sifting structure, fully characterized by the corresponding individual conditional subprocesses, a first question to be considered is the following: is it true that the identification of a conditional model corresponding to a given sample size may be reduced, under a sifting structure, to the identification of the corresponding individual conditional submodels? Theorems 3.1 and 3.2 give a positive answer at the level of a (general) conditional structural experiment defined from $\mathcal{E}_{[n]}$. Note that *without* the sifting property these results can not be established.

As a next question, this paper considers the identification of a statistical model built from a structural model embodying unobservable heterogeneity. Identifiability for an n-sample is obtained in Theorem 3.3 by using the identifiability of individual processes corresponding to the component submodels, namely the conditional submodel and the marginal one; Theorem 3.3 does not require additional assumptions on the marginal model generating the unobservable incidentals but a supplementary independence assumption in the marginal model ensures the statistical model a simple behaviour, namely a mutual independence property, and a suitable predictive weak identification, obtained in Theorem 3.4.

Whereas section 3 deals with a general issue, the use of the main results is illustrated, in section 4, in the framework of the large class of IRT models. Thanks to a systematic recourse to a σ-algebraic argument, it becomes apparent that a a semiparametric class of models is naturally dealt with. Most of the properties dealt with do not depend on parametric specification: the results obtained in Theorem 4.1 are actually relevant at a first stage of modelling, but Corollaries 4.1 and 4.2 also show that, thanks to such a theoretical framework, analytical arguments are easily concentrated on clearly located aspects of the specification. By so-doing, it is hoped that the contextual meaning of each hypothesis is better evaluated and the proof of results specific to a given model are simplified by allowing the use of general intermediary results.

6 Formal complements

6.1 Some comments on the basic concepts of identification

(1) 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), and is linked to that of identifiation of mixtures as *e.g.* Teicher (1961), Barndorff-Nielsen (1965) and Chandra (1977); for details, see Florens *et al.* (1990, chapter 5). When $\mathcal{X}_1$ represents a parameter and $\mathcal{X}_2$ an observation, $\mathcal{X}_1$ p-strongly identified by $\mathcal{X}_2$ means that $\mathcal{X}_1$ is a *parameter p-complete* for the process generating $(\mathcal{X}_2 \mid \mathcal{X}_1)$; in this case, Definition 2.1 essentially depends on the posterior distribution of $(\mathcal{X}_1 \mid \mathcal{X}_2)$.

(2) The *weak identification* is defined by means of the σ-field $\sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_1) : f_2 \in [\mathcal{X}_2]^+\} \subset \mathcal{X}_1$; this is the smallest sub-σ-field of $\mathcal{X}_1$ which, on the one hand, makes measurable all expectations of functions of $\mathcal{X}_2$ conditionally on $\mathcal{X}_1$, and, on the other hand, makes $\mathcal{X}_1$ and $\mathcal{X}_2$ conditionally independents. Also it contains all null sets of $\mathcal{X}_1$, *i.e.* $\mathcal{X}_1 \cap \overline{\mathcal{M}_0} \subset \sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_1) : f_2 \in [\mathcal{X}_2]^+\}$; for more details, see Florens *et al.* (1990, section 4.3). Thus, when $\mathcal{X}_1$ represents a parameter and $\mathcal{X}_2$ an observation, to say that $\mathcal{X}_1$ is weakly identified by $\mathcal{X}_2$ means that any (positive) measurable function of the parameters may be represented as a function of countably many sampling expectations of statistics. Because $\sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_1) : f_2 \in [\mathcal{X}_2]^+\}$ represents the minimal sufficient parameter, the weak identification problem appears as a problem of *minimal sufficiency* on the parameter space; this formal analogy has already been remarked by *e.g.* Barankin (1961), Kadane (1974) or Picci (1977).

(3) From a statistical point of view the concept of identification of a statistical model, namely the injectivity of the mapping from the parameter space to the space of sampling distributions (for details, see *e.g.* Rothenberg, 1971 or Bunke and Bunke, 1974), corresponds to the weak identification of the σ-field generated by the parameter by the σ-field generated by the observation. More precisely, if the parameter space is a Blackwell σ-field and the sampling space is separable, then, for any prior probability measure defined on the parameter space, the concept of sampling identification implies that of weak identification; for details and proofs, see Florens *et al.* (1990, section 4.6.2), and for details concerning Blackwell σ-fields, see *e.g.* Blackwell (1956) and Dellacherie and Meyer (1975, chapters 1 and 3). In particular, if both the parameter space and the sampling space are Borel σ-fields of

a finite dimensional Euclidean space, it follows that sampling identification implies weak identification for any prior probability.

(4) In general, if $\mathcal{X}_1$ is p-strongly identified by $\mathcal{X}_2$ conditionally on $\mathcal{X}_3$ (with $p > 0$), then $\mathcal{X}_1$ is weakly identified by $\mathcal{X}_2$ conditionally on $\mathcal{X}_3$; see Florens *et al.* (1990, Theorem 5.4.12). The weak identification does not imply the strong one; for an example in the discrete case, see Mouchart and San Martín (1999, Proposition 1). Nevertheless, these two concepts are equivalent if $(X_1, X_2 \mid X_3)$ is normally distributed, with $\mathcal{X}_i = \sigma(X_i)$; for a proof, see Mouchart and San Martín (1999, Proposition 2).

6.2 Proof of Proposition 3.1

On the one hand, the sifting condition (3.2) implies the following *transitivity* condition (for a proof, see Florens *et al.*, 1990, Theorem 7.6.9):

$$\mathcal{Y}_I \perp\!\!\!\perp \mathcal{H}_{I^c} \mid \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, I \subset \{1, \ldots, n\}, \tag{6.1}$$

where $I^c = \{1, \ldots, n\} \setminus I$. On the other hand, property (1.10.ii) implies

$$\mathcal{Y}_I \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_{I^c} \vee \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, I \subset \{1, \ldots, n\}. \tag{6.2}$$

The proposition follows from conditions (6.1) and (6.2).

□

6.3 Proof of Theorem 3.1

The proof of Theorem 3.1(i) follows by using the following lemma, which states conditions under which the p-strong identification may be increased in both sides (with $p > 0$).

Lemma 6.1 *Let $(M, \mathcal{M}, P)$ be a probability space, and let $\mathcal{X}_i$, $i = 1, \ldots, 5$, be five sub-σ-fields of $\mathcal{M}$. The following p-strong identification conditions (with $p \in [1, \infty]$)*

$$\text{(i) } \mathcal{X}_1 \ll_p \mathcal{X}_2 \mid \mathcal{X}_5, \quad \text{(ii) } \mathcal{X}_3 \ll_p \mathcal{X}_4 \mid \mathcal{X}_5$$

jointly imply that $\mathcal{X}_1 \vee \mathcal{X}_3 \ll_p \mathcal{X}_2 \vee \mathcal{X}_4 \mid \mathcal{X}_5$ under the following two conditions:

$$\text{(iii) } \mathcal{X}_2 \perp\!\!\!\perp \mathcal{X}_3 \mid \mathcal{X}_1 \vee \mathcal{X}_5, \quad \text{(iv) } \mathcal{X}_4 \perp\!\!\!\perp \mathcal{X}_1 \vee \mathcal{X}_2 \mid \mathcal{X}_3 \vee \mathcal{X}_5.$$

Proof. Let $p \in [1, \infty]$. Since in general $\mathcal{X}_1 \ll_p \mathcal{X}_2 \mid \mathcal{X}_5$ is equivalent to $\mathcal{X}_1 \vee \mathcal{X}_5 \ll_p \mathcal{X}_2 \vee \mathcal{X}_5$, it is sufficient to prove this lemma for $\mathcal{X}_5 = \{\emptyset, M\}$. Firstly, $\mathcal{X}_1 \ll_p \mathcal{X}_2$ along with $\mathcal{X}_2 \perp\!\!\!\perp \mathcal{X}_3 \mid \mathcal{X}_1$ implies (v) $\mathcal{X}_1 \vee \mathcal{X}_3 \ll_p \mathcal{X}_2 \vee \mathcal{X}_3$ (for a proof, see Theorem 5.4.5 in Florens *et al.*, 1990). Similarly, $\mathcal{X}_3 \ll_p \mathcal{X}_4$ along with $\mathcal{X}_2 \perp\!\!\!\perp \mathcal{X}_4 \mid \mathcal{X}_3$ (a property implied by condition (iv) of Lemma 6.1) implies (vi) $\mathcal{X}_2 \vee \mathcal{X}_3 \ll_p \mathcal{X}_2 \vee \mathcal{X}_4$. Since $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{X}_4 \mid \mathcal{X}_2 \vee \mathcal{X}_3$ (a property implied by condition (iv) of Lemma 6.1), the p-strong conditions (v) and (vi) jointly imply $\mathcal{X}_1 \vee \mathcal{X}_3 \ll_p \mathcal{X}_2 \vee \mathcal{X}_4$ by using the transivity property of the strong identification; see Theorem 5.4.10 in Florens *et al.* (1990).

□

Remark 6.1 Note that conditions (iii) and (iv) of Lemma 6.1 are implied by the stronger condition $\mathcal{X}_1 \vee \mathcal{X}_2 \perp\!\!\!\perp \mathcal{X}_3 \vee \mathcal{X}_4 \mid \mathcal{X}_5$; moreover, under this condition, properties (i) and (ii) of Lemma 6.1 are both equivalent to $\mathcal{X}_1 \vee \mathcal{X}_3 \ll_p \mathcal{X}_2 \vee \mathcal{X}_4 \mid \mathcal{X}_5$.

Proof of Theorem 3.1(i). Let $I \subset \{1, \ldots, n\}$ and $j \in \{1, \ldots, n\} \setminus I$. Let $p \in [1, \infty]$; then, by using Lemma 6.1, both $\mathcal{H}_I \vee \mathcal{S}_2 \ll_p \mathcal{Y}_I$ and $\mathcal{H}_j \vee \mathcal{S}_2 \ll_p \mathcal{Y}_j$ imply $\mathcal{H}_{I \cup \{j\}} \vee \mathcal{S}_2 \ll_p \mathcal{Y}_{I \cup \{j\}}$ under the following two conditions:

$$\mathcal{Y}_I \perp\!\!\!\perp \mathcal{H}_j \mid \mathcal{H}_I \vee \mathcal{S}_2 \text{ and } \mathcal{Y}_j \perp\!\!\!\perp \mathcal{H}_I \vee \mathcal{Y}_I \mid \mathcal{H}_j \vee \mathcal{S}_2 \quad \forall\, j \in \{1, \ldots, n\} \setminus I. \tag{6.3}$$

But these conditions are implied by the sifting one. Indeed, the sifting condition (3.2) implies both the transitivity condition (6.1) and

$$\mathcal{Y}_I \perp\!\!\!\perp \mathcal{Y}_{I^c} \mid \mathcal{H}_1^n \vee \mathcal{S}_2 \quad \forall\, I \subset \{1, \ldots, n\},$$

which in turn are jointly equivalent to $\mathcal{Y}_I \perp\!\!\!\perp \mathcal{Y}_{I^c} \vee \mathcal{H}_{I^c} \mid \mathcal{H}_I \vee \mathcal{S}_2 \ \forall\, I \subset \{1, \ldots, n\}$; from this last condition (6.3) follows; the proof is concluded by induction on j.

□

The proof of Theorem 3.1(ii) follows by using the same argument as for part (i).

6.4 Proof of Theorem 3.2

The proof of Theorem 3.2(i) is based on the following two characterizations of the conditional independence:

1. By definition, $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{X}_2 \mid \mathcal{X}_3$ is equivalent to $\mathbb{E}\{f_1 f_2 \mid \mathcal{X}_3\} = \mathbb{E}\{f_1 \mid \mathcal{X}_3\}\,\mathbb{E}\{f_2 \mid \mathcal{X}_3\}$ for all $f_i \in [\mathcal{X}_i]^+$, $i = 1, 2$; that is,

$$\sigma\{\mathbb{E}(f_{12} \mid \mathcal{X}_3) \,:\, f_{12} \in [\mathcal{X}_1 \vee \mathcal{X}_2]^+\} = \tag{6.4}$$

$$= \sigma\{\mathbb{E}(f_1 \mid \mathcal{X}_3) \,:\, f_1 \in [\mathcal{X}_1]^+\} \vee \sigma\{\mathbb{E}(f_2 \mid \mathcal{X}_3) \,:\, f_2 \in [\mathcal{X}_2]^+\}.$$

2. By definition, $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{X}_2 \mid \mathcal{X}_3$ is equivalent to $\mathbb{E}\{f_1 \mid \mathcal{X}_2 \vee \mathcal{X}_3\} = \mathbb{E}\{f_1 \mid \mathcal{X}_3\}$ for all $f_1 \in [\mathcal{X}_1]^+$; that is,

$$\sigma\{\mathbb{E}(f_1 \mid \mathcal{X}_3) \,:\, f_1 \in [\mathcal{X}_1]^+\} \subset \tag{6.5}$$

$$\sigma\{\mathbb{E}(f_1 \mid \mathcal{X}_2 \vee \mathcal{X}_3) \,:\, f_1 \in [\mathcal{X}_1]^+\} \subset \overline{\sigma\{\mathbb{E}(f_1 \mid \mathcal{X}_3) \,:\, f_1 \in [\mathcal{X}_1]^+\}};$$

for more details, see Florens *et al.* (1990, section 4.3).

The the following lemma is also used in the proof of Theorem 3.2:

Lemma 6.2 *Under the allocation condition corresponding to the sifting condition (3.2), namely $\mathcal{Y}_i \perp\!\!\!\perp \mathcal{H}_1^n \mid \mathcal{H}_i \vee \mathcal{S}_2$ for $i = 1, \ldots, n$, it follows that*

$$\mathop{\perp\!\!\!\perp}_{1 \le i \le n} \mathcal{Y}_i \mid \mathcal{H}_1^n \vee \mathcal{S}_2 \iff \mathop{\perp\!\!\!\perp}_{i \in I} \mathcal{Y}_i \mid \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, I \subset \{1, \ldots, n\}.$$

Proof. Let $I \subset \{1, \ldots, n\}$; on the one hand, condition $\mathop{\perp\!\!\!\perp}_{1 \le i \le n} \mathcal{Y}_i \mid \mathcal{H}_1^n \vee \mathcal{S}_2$ implies

$$\mathcal{Y}_J \perp\!\!\!\perp \mathcal{Y}_{\{1,\ldots,n\} \setminus J} \mid \mathcal{H}_{\{1,\ldots,n\} \setminus I} \vee \mathcal{H}_I \vee \mathcal{S}_2 \quad \forall\, J \subset I; \tag{6.6}$$

on the other hand, the transitivity condition (6.1) implies

$$\mathcal{Y}_J \perp\!\!\!\perp \mathcal{H}_{\{1,\ldots,n\} \setminus I} \mid \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, J \subset I. \tag{6.7}$$

Thus, conditions (6.6) and (6.7) jointly imply that $\mathcal{Y}_J \perp\!\!\!\perp \mathcal{Y}_{\{1,\ldots,n\} \setminus J} \mid \mathcal{H}_I \vee \mathcal{S}_2$, which in turn implies

$$\mathcal{Y}_J \perp\!\!\!\perp \mathcal{Y}_{I\setminus J} \mid \mathcal{H}_I \vee \mathcal{S}_2 \qquad \forall\, J \subset I,$$

i.e. $\mathop{\perp\!\!\!\perp}_{i\in I} \mathcal{Y}_i \mid \mathcal{H}_I \vee \mathcal{S}_2$. The converse implication follows by taking $I = \{1,\ldots,n\}$.

□

Proof of Theorem 3.2(i). Let $I \subset \{1,\ldots,n\}$; by Lemma 6.2 it follows that the sifting property (3.2) implies $\mathop{\perp\!\!\!\perp}_{i\in I} \mathcal{Y}_i \mid \mathcal{H}_I \vee \mathcal{S}_2$, which in turn implies, by using inductively property (6.4), that

$$\sigma\{\mathbb{E}(f \mid \mathcal{H}_I \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_I]^+\} = \tag{6.8}$$

$$= \bigvee_{i\in I} \sigma\{\mathbb{E}(f \mid \mathcal{H}_I \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\}.$$

Nevertheless, by using property (6.5), conditions $\mathcal{Y}_i \perp\!\!\!\perp \mathcal{H}_I \mid \mathcal{H}_i \vee \mathcal{S}_2$ for each $i \in I$ (implied by the transitivity one (6.1)) imply

$$\sigma\{\mathbb{E}(f \mid \mathcal{H}_i \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\} \subset \tag{6.9}$$

$$\subset\ \sigma\{\mathbb{E}(f \mid \mathcal{H}_I \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\} \subset \overline{\sigma\{\mathbb{E}(f \mid \mathcal{H}_i \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\}}$$

for each $i \in I$; the measurable completions are taken with respect to the probability measure defined on $\mathcal{S} \vee \bigvee_{i\in I}(\mathcal{Y}_i \vee \mathcal{H}_i)$. Therefore (6.8) becomes

$$\bigvee_{i\in I} \sigma\{\mathbb{E}(f \mid \mathcal{H}_i \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\} \subset \tag{6.10}$$

$$\subset \bigvee_{i\in I} \sigma\{\mathbb{E}(f \mid \mathcal{H}_I \vee \mathcal{S}_2) \,:\, f \in [\mathcal{Y}_i]^+\}.$$

Consequently, if $\mathcal{H}_i \vee \mathcal{S}_2 \prec \mathcal{Y}_i$ for each $i \in I$, it follows that $\mathcal{H}_I \vee \mathcal{S}_2 \prec \mathcal{Y}_I$.

□

Noting that, for any $I \subset \{1,\ldots,n\}$,

$$\sigma\{\mathbb{E}(f \mid \mathcal{S}_2 \vee \mathcal{H}_I) : f \in [\mathcal{Y}_I \vee \mathcal{H}_I]^+\} = \mathcal{H}_I \vee \sigma\{\mathbb{E}(f \mid \mathcal{S}_2 \vee \mathcal{H}_I) : f \in [\mathcal{Y}_I]^+\},$$

the proof of Theorem 3.2(ii) follows by the same argument as for part (i).

6.5 Proof of Theorem 3.3

On the one hand, conditions $\mathcal{S}_2 \vee \mathcal{H}_i \ll_2 \mathcal{Y}_i$ for any $i \in I$ imply, under the sifting assumption (3.2), that $\mathcal{S}_2 \vee \mathcal{H}_I \ll_2 \mathcal{Y}_I$ by Theorem 3.1. On the other hand, $\mathcal{S}_1 \prec \mathcal{H}_i$ for (at least one) $i \in I$ imply that $\mathcal{S}_1 \prec \mathcal{H}_I$. But the sifting property (3.2) implies, by Proposition 3.1, that $\mathcal{Y}_I \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_I$; the Theorem follows by using Theorem 3 in Mouchart and San Martín (1999).

□

Remark 6.2 An alternative proof to Theorem 3.3 may be given: Let $i \in I$; on the one hand, condition (1.10.i) implies $\mathcal{H}_i \perp\!\!\!\perp \mathcal{S} \mid \mathcal{S}_1$ for any $i \in I$; on the other hand, the sifting condition (3.2) implies that $\mathcal{Y}_i \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_i \vee \mathcal{S}_2$ for any $i \in I$. Therefore, $\mathcal{S}_1 \prec \mathcal{H}_i$ for (at least one) $i \in I$ and $\mathcal{S}_2 \vee \mathcal{H}_i \ll_2 \mathcal{Y}_i$ for any $i \in I$ jointly implies that $\mathcal{S}_1 \vee \mathcal{S}_2 \prec \mathcal{Y}_i$ for (at least one) $i \in I$ by using Theorem 3 in Mouchart and San Martín (1999). From this last condition it follows that $\mathcal{S}_1 \vee \mathcal{S}_2 \prec \mathcal{Y}_I$.

6.6 Proof of Lemma 3.1

Let $\mathcal{S} = \mathcal{S}_1 \vee \mathcal{S}_2$; since $\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) : f \in [\mathcal{S}]^+\} \subset \mathcal{Y}_1^n$ and

$$\mathcal{Y}_1^n \perp\!\!\!\perp \mathcal{S} \mid \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) : f \in [\mathcal{S}]^+\},$$

condition (3.5) implies

$$\mathcal{Y}_{n+1}^{n+m} \vee \mathcal{S} \perp\!\!\!\perp \mathcal{Y}_1^n \mid \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) : f \in [\mathcal{S}]^+\}; \tag{6.11}$$

therefore, by using the characterization of the conditional independence in terms of measurability (see Theorem 2.2.6 in Florens *et al.*, 1990), (6.11) is equivalent to

$$\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) : f \in [\mathcal{Y}_{n+1}^{n+m} \vee \mathcal{S}]^+\} \subset \overline{\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) : f \in [\mathcal{S}]^+\}},$$

which in turn implies

$$\begin{aligned}\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{Y}_{n+1}^{n+m}]^+\} &\subset \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{Y}_{n+1}^{n+m} \vee \mathcal{S}]^+\} \\ &\subset \overline{\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{S}]^+\}} \cap \mathcal{Y}_1^n \\ &= \sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{S}]^+\},\end{aligned}$$

since $\sigma\{\mathbb{E}(f \mid \mathcal{Y}_1^n) \,:\, f \in [\mathcal{S}]^+\}$ contains all the null sets of $\mathcal{Y}_1^n$.

□

6.7 Proof of Lemma 4.1

By definition of $G(B)$, it follows that $\mathcal{G} \subset \mathcal{H}_\Sigma$. Conversely, since η is $\mathcal{H}_\Sigma$-i.i.d., it follows, for $\eta_i \in [\mathcal{H}_i]^+$ and $B_i \in \mathcal{B}$, $i = 1, \ldots, n$, that

$$\begin{aligned}\mathbb{P}\left\{\bigcap_{1 \le i \le n} \{\eta_i \in B_i\} \mid \mathcal{H}_\Sigma\right\} &= \prod_{i=1}^n \mathbb{P}\{\eta_i \in B_i \mid \mathcal{H}_\Sigma\} \\ &= \prod_{i=1}^n G(B_i) \in [\mathcal{G}]^+.\end{aligned} \tag{6.12}$$

By using the Montone Class Theorem, (6.12) implies that

$$\sigma\{\mathbb{E}(f \mid \mathcal{H}_\Sigma) \,:\, f \in [\mathcal{H}_1^\infty]^+\} \subset \mathcal{G}; \tag{6.13}$$

but $\mathcal{H}_\Sigma \subset \mathcal{H}_1^\infty$; therefore, (6.13) implies that $\mathcal{H}_\Sigma \subset \mathcal{G}$.

□

6.8 Proof of Lemma 4.2

Let us represent the identification restriction **(IC)** in the form $a's = 0$ Π-a.s. where $a \in \mathbb{R}^m$ is a basis of the 1-dimensional orthogonal complement of the linear subspace B. For any $i \in \{1, \ldots, n\}$, the linear restriction, along with the property $\sigma(\lambda_{ij}) = \sigma(\eta_i - s_j)$, may be rewritten as

$$\begin{pmatrix} f(\lambda_{i1}) \\ \vdots \\ f(\lambda_{im}) \\ 0 \end{pmatrix} = \begin{pmatrix} i_m & -I_m \\ 0 & a' \end{pmatrix} \begin{pmatrix} \eta_i \\ s \end{pmatrix},$$

where $i'_m = (1, \ldots, 1) \in \mathbb{R}^m$ and f is a bijective transformation applied on the λ_{ij}'s. The rank of the above matrix is equal to $m+1$ provided that $i_m{}'a \neq 0$; in such a case,

$$\overline{\mathcal{S} \vee \mathcal{H}_i} = \bigvee_{1 \leq j \leq m} \overline{\sigma\{f(\lambda_{ij})\}} = \bigvee_{1 \leq j \leq m} \overline{\sigma(\lambda_{ij})}.$$

Consequently, it follows, on the one hand, that

$$\perp\!\!\!\perp_{1 \leq j \leq m} \mathcal{Y}_{ij} \mid \mathcal{H}_i \vee \mathcal{S} \iff \perp\!\!\!\perp_{1 \leq j \leq m} \mathcal{Y}_{ij} \mid \overline{\mathcal{H}_i \vee \mathcal{S}} \iff \perp\!\!\!\perp_{1 \leq j \leq m} \mathcal{Y}_{ij} \mid \bigvee_{1 \leq j \leq m} \overline{\sigma(\lambda_{ij})};$$

on the other hand, $\mathcal{Y}_{ij} \perp\!\!\!\perp \mathcal{S} \mid \mathcal{H}_i \vee \mathcal{S}_j$ and condition (4.10) are jointly equivalent to

$$\begin{aligned} \mathcal{Y}_{ij} \perp\!\!\!\perp \mathcal{H}_i \vee \mathcal{S} \mid \sigma(\lambda_{ij}) &\iff \overline{\mathcal{Y}_{ij}} \perp\!\!\!\perp \overline{\mathcal{H}_i \vee \mathcal{S}} \mid \sigma(\lambda_{ij}) \\ &\iff \overline{\mathcal{Y}_{ij}} \perp\!\!\!\perp \bigvee_{1 \leq j \leq m} \overline{\sigma(\lambda_{ij})} \mid \sigma(\lambda_{ij}) \end{aligned}$$

which in turn implies that

$$\mathcal{Y}_{ij} \perp\!\!\!\perp \bigvee_{1 \leq j \leq m} \overline{\sigma(\lambda_{ij})} \mid \overline{\sigma(\lambda_{ij})} \qquad \forall\, j = 1, \ldots, n.$$

Therefore, the sifting condition (4.3.ii) implies the sifting property $\{\mathcal{Y}_{ij}\}$ is sifted by $\overline{\{\sigma(\lambda_{ij})\}}$; the proposition follows by Theorem 3.1 and by noting that in general $\mathcal{X}_1 \ll_p \mathcal{X}_2$ is equivalent to $\overline{\mathcal{X}_1} \ll_p \mathcal{X}_2$.

□

6.9 Proof of Lemma 4.3

Let $(y \mid \lambda) \sim \text{Poisson}(\lambda)$, with $y \in \mathbb{N}$, and let μ_λ be a probability measure on $(\mathbb{R}_+, \mathcal{B})$. By definition, $\sigma(\lambda) \ll_p \mathcal{Y}$ if

$$\forall\, h \in L^p(\mathbb{R}_+, \mu_\lambda) \qquad \int_0^\infty h(t)\, d\mu_{\lambda|y}(t) \;=\; 0 \implies h \;=\; 0 \;\; \mu_\lambda - \text{a.s.} \quad (6.14)$$

where $\mu_{\lambda|y}$ is a posterior probability measure on $\mathbb{R}_+$. The nullity of the integral in equation (6.14) is equivalent to

$$\int_0^\infty h(t)\, \frac{t^y}{y!}\, e^{-t}\, d\mu_\lambda(t) \;=\; 0 \quad \forall\, y \in \mathbb{N}. \qquad (6.15)$$

1. Let $\mathcal{L}_0$ be the set defined as

$$\mathcal{L}_0 \;=\; \left\{ f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \mid f(t) \;=\; e^{-t} \sum_{y=0}^{N} a_y \frac{t^y}{y!} \;:\; y,\, N \in \mathbb{N},\;\; a_y \in \mathbb{R}_+ \right\};$$

then (6.15) implies that

$$\int_0^\infty h(t)\, g(t)\, d\mu_\lambda(t) \;=\; 0 \qquad \forall\, g \in \mathcal{L}_0. \qquad (6.16)$$

Let $\mathcal{L}$ be the following set:

$$\mathcal{L} = \left\{ f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \mid f(t) \;=\; e^{-kt}\, p(t) \;:\; k \in \mathbb{N}_0,\; p(t) \geq 0 \text{ a polynomial} \right\},$$

where $\mathbb{N}_0 = \mathbb{N} \backslash \{0\}$. By definition, $\mathcal{L}_0 \subset \mathcal{L}$; moreover, since

$$e^{-kt} \;=\; \lim_{N \to \infty} \sum_{l=0}^{N} (-1)^l\, (k-1)^l\, \frac{t^l}{l!}\, e^{-t} \qquad \forall\, k \in \mathbb{N}_0,$$

it follows that $\mathcal{L}_0$ is dense in $\mathcal{L}$ with the sup norm $\| \cdot \|_\infty$.

2. Let $C_0(\mathbb{R}_+)$ be the set of continuous positive functions on $\mathbb{R}_+$ that vanish at infinity. Then $\mathcal{L} \subset C_0(\mathbb{R}_+)$ is dense in $C_0(\mathbb{R}_+)$ with $\| \cdot \|_\infty$. Indeed, $\mathcal{L}$ is a subalgebra of $C_0(\mathbb{R}_+)$ that separates the points of $\mathbb{R}_+$. Let

$\overline{\mathbb{R}_+}$ be the one point compactification of $\mathbb{R}_+$; then $C_0(\mathbb{R}_+)$ is identified with $\{f \in C(\overline{\mathbb{R}_+}) : f(\infty) = 0\}$ (see Conway, 1985, p. 69); so $\mathcal{L}$ becomes a subalgebra of $C(\overline{\mathbb{R}_+})$ and, consequently, by the Stone-Weierstrass Theorem (see *e.g.* Conway, section V.8), $\mathcal{L}$ is dense in $C(\overline{\mathbb{R}_+})$ (or, in $C_0(\mathbb{R}_+)$) with $\|\cdot\|_\infty$.

3. Since μ_λ is a probability measure on $(\mathbb{R}_+, \mathcal{B})$, it follows that $L^p(\mathbb{R}_+, \mu_\lambda) \subset L^1(\mathbb{R}_+, \mu_\lambda)\ \forall p \in [1, \infty]$, and therefore if $\lambda \ll_1 y$ then $\lambda \ll_p y$ for any $p \in [1, \infty]$. Consequently, it is sufficient to show that $\lambda \ll_1 y$. Indeed, for $p = 1$, let $g \in \mathcal{L}$; then, by step **2**, there exists a sequence $\{g_n\} \subset \mathcal{L}_0$ such that $\|g_n - g\|_\infty \longrightarrow 0$ as $n \to \infty$. Therefore, it follows that

$$\int_0^\infty h(t)\,[g_n(t) - g(t)]\,d\mu_\lambda(t) \;\leq \tag{6.17}$$

$$\leq \left[\int_0^\infty |h(t)|\,d\mu_\lambda(t)\right] \|g_n - g\|_\infty \longrightarrow 0 \quad \text{as } n \to \infty$$

because $h \in L^1(\mathbb{R}_+, \mu_\lambda)$; consequently, (6.16) is valid when $g \in \mathcal{L}_0$ is replaced by any $g \in \mathcal{L}$. Similarly, since $\mathcal{L}$ is dense in $C_0(\mathbb{R}_+)$, it follows that

$$\int_0^\infty h(t)\,g(t)\,d\mu_\lambda(t) = 0 \qquad \forall\, g \in C_0(\mathbb{R}_+). \tag{6.18}$$

But $C_c(\mathbb{R}_+) \subset C_0(\mathbb{R}_+)$, where $C_c(\mathbb{R}_+)$ is the set of continuous positive functions on $\mathbb{R}_+$ with a compact support; therefore, (6.18) is valid for any $g \in C_c(\mathbb{R}_+)$. In such a case, (6.18) implies that $h = 0$ μ_λ-a.s. since $h \in L^1(\mathbb{R}_+, \mu_\lambda)$ (see Brezis, 1983, Lemma IV. 2).

□

Remark 6.3 This proof deserves the following comments:

1. As the Laguerre Polynomials are complete in $L^2(\mathbb{R}_+, \mu_\lambda)$ (see *e.g.* Sansone, 1959, chapter IV), a Gamma distribution, taken as a prior probability measure, implies that $\lambda \ll_2 y$. Consequently, Lemma 4.3 constitutes a generalization of that result to L^p-spaces for any prior probability measure and for any $p \in [1, \infty]$.

2. If μ_λ is any σ-finite measure on $(\mathbb{R}_+, \mathcal{B})$, Lemma 4.3 is still valid for $p = 1$, provided that the corresponding predictive measure is σ-finite

since in such a case the posterior measure is actually a probability measure; for details, see Mouchart (1976) and Florens *et al.* (1990, chapter 1).

3. For $p \in (1, \infty)$, if μ_λ is a σ-finite measure on $(\mathbb{R}_+, \mathcal{B})$, Lemma 4.3 follows by adding an additional assumption ensuring that $h \in L^p(\mathbb{R}_+, \mu_\lambda)$ implies $h \in L^1(\mathbb{R}_+, \mu_\lambda)$, provided that the corresponding predictive measure is σ-finite; for example, $e^t \in L^1(\mathbb{R}_+, \mu_\lambda)$ (this additional condition implies restrictions on the σ-finite measure μ_λ). Indeed, under such an assumption, (6.17) may be used, and therefore (6.18) is valid for any $g \in C_c(\mathbb{R}_+)$. The proof concludes by using firstly the fact that $C_c(\mathbb{R}_+)$ is dense in $L^q(\mathbb{R}_+, \mu_\lambda)$ for any $q \in [1, \infty)$ and secondly by using a duality argument in L^p-spaces.

6.10 Proof of Lemma 4.4

Let $(y \mid \lambda) \sim \text{Bernoulli}(\lambda)$ and $(\lambda \mid a, b) \sim \text{Beta}(a, b)$, with $a, b \geq 1$, be the prior distribution of λ defined on $([0,1], \mathcal{B}_{[0,1]})$, denoted as μ_λ; the restriction $a, b \geq 1$ is introduced in order to ensure the boundedness of the beta density in $[0, 1]$. By definition, $\lambda \ll_p y$ for any $p \in [1, \infty]$ if and only if $\forall\, h \in L^p([0,1], \mu_\lambda)$

$$\int_0^1 h(t)\, t^{y+a-1}\, (1-t)^{b-y}\, dt \;=\; 0 \quad \forall\, y \in \{0,1\}, \tag{6.19}$$

then $h = 0$ $\mu_\lambda-$a.s. Indeed, let $p \in [1, \infty)$ and $\mathcal{L}$ be the set defined as

$$\mathcal{L} \;=\; \{f \,:\, [0,1] \longrightarrow [0,1] \,|\, f(t) \;=\; t^{y+a-1}\, (1-t)^{b-y} \;:\; y \in \{0,1\},\;\; a, b \geq 1\};$$

then $\mathcal{L}$ is a subalgebra of $C[0,1]$ that separates the points in $[0,1]$, with $1 \in \mathcal{L}$. Therefore, by the Stone-Weierstrass Theorem, $\mathcal{L}$ is dense in $C[0,1]$ with $\|\cdot\|_\infty$; consequently, (6.19) implies that

$$\forall\, h \in L^p([0,1], \mu_\lambda) \qquad \int_0^1 h(t)\, g(t)\, dt \;=\; 0 \quad \forall\, g \in C[0,1]. \tag{6.20}$$

Noting that $C[0,1] \subset L^q([0,1], \mu_\lambda)$, and reminding that $C[0,1]$ is dense in $L^q([0,1], \mu_\lambda)$ with $\|\cdot\|_{L^q}$ for all $q \in [1, \infty)$, it follows that (6.20) is valid for

any $g \in L^q([0,1],\mu_\lambda)$. Taking q such that $1/p + 1/q = 1$, $L^q([0,1],\mu_\lambda)$ is the toplogical dual of $L^p([0,1],\mu_\lambda)$; therefore, condition (6.20) implies that $h = 0$ μ_λ-a.s. Finally, since μ_λ is a probability measure defined on $([0,1],\mathcal{B}_{[0,1]})$, it follows that $\lambda \ll_1 y$ implies $\lambda \ll_\infty y$.

□

6.10.1 Some remarks about the identifiability in the polytomous case

As pointed out in section 4.1, the polytomous item response case is specified as follows: each individual i answers in one of l categories to each of m questions; that is, $y_{ij} = (y_{ij1},\ldots,y_{ijl}) \in \{0,1\}^l$ with $\sum_{k=1}^{l} y_{ijk} = 1$; in such a case, each category k (with $k = 1,\ldots,l$) of the question j is characterized by a difficulty parameter s_{jk}; in σ-algebraic terms this means that $\mathcal{S}_j = \bigvee_{1\leq k\leq l} \mathcal{S}_{kj}$, where $\mathcal{S}_{jk} = \sigma(s_{kj})$. Therefore, under the double sifting structure (4.3), $p(y_{ij} \mid \eta_i, s_j)$ is given by

$$p(y_{ij} \mid \eta_i, s_j) = \prod_{k=1}^{l} \lambda_{ijk}^{y_{ijk}} \tag{6.21}$$

where

$$\lambda_{ijk} = \frac{\exp(\eta_i - s_{jk})}{\sum_{k=1}^{l} \exp(\eta_i - s_{jk})} \qquad \forall\, k = 1,\ldots,l.$$

Following the arguments developed in section 4.2, the weak identification of the structural parameter by the observations is reduced to verify that $\sigma(\lambda_{ij}) = \bigvee_{1\leq k\leq l} \sigma(\lambda_{ijk})$ is 2-strongly identified by $\mathcal{Y}_{ij}$. In order to prove it, a Dirichlet distribution may be considered as the prior distribution of $(\lambda_{ij1},\ldots,\lambda_{ijl})$ on $([0,1]^l,\mathcal{B})$, that is

$$p_{l-1}(\lambda_{ij1},\ldots,\lambda_{ij,l-1}) = \frac{\Gamma(\sum_{k=1}^{l} \alpha_i)}{\prod_{k=1}^{l} \Gamma(\alpha_i)} \prod_{k=1}^{l} \lambda_{ijk}^{\alpha_k - 1}$$

where $\sum_{k=1}^{l} \lambda_{ijk} = 1$, $\lambda_{ijk} \geq 0$ and $\alpha_k \geq 1$ for all $k = 1,\ldots l$ (this last condition is a restriction in order to ensure the boundedness of the Dirichlet density in $[0,1]^l$); for details about the Dirichlet distribution, see *e.g.* Fang *et*

al. (1990, section 1.4). The proof follows by considering the same arguments to that developed in the dichotomous case.

□

References

Agresti, A. (1993). Computing Conditional Maximum Likelihood Estimates for Generalized Rasch Models Using Simple Loglinear Models with Diagonals Parameters. *Scand. J. Statist.* **20**, 63-71.

Aldous, D. J. (1985). *Exchangeability and Related Topics* (Lecture Notes in Mathematics, **1117**). Springer-Verlag, New York.

Andersen, E. B. (1973a). *Conditional Inference and Models for Measuring.* Mentalhygiejnisk Forlag, Copenhagen.

Andersen, E. B. (1973b). Conditional Inference for Multiple-Choice Questionnaires. *Brit. Jour. Math. Stat. Psych.* **26**, 31-44.

Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* North-Holland, Amsterdam.

Andersen, E. B. (1982). Latent Structure Analysis: A Survey. *Scand. J. Statist.* **9**, 1-12.

Barankin, E. W. (1961). Sufficient Parameters: Solution of the Minimal Dimensionality Problem. *Ann. Inst. Statist. Math.* **12**, 91-118.

Barndorff-Nielsen, O. (1965). Identifiability of Mixtures of Exponential Families. *J. Math. Anal. Appl.* **12**, 115-121.

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis.* Charles Griffin, London.

Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring any Examinee's Ability. In *Statistical Theories of Mental Test Scores* (by F. M. Lord and M. R. Novick), Addison-Wesley, Massachusetts, chapters 17-20.

Blackwell, D. (1956). On a Class of Probability Spaces. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probabilities* **2**, 1-6.

Brezis, H. (1983). *Analyse fonctionnelle. Théorie et applications.* Masson, Paris.

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

Chandra, S. (1977). On the Mixture of Probability Distributions. *Scand. J. Statist.* **4**, 105-112.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory. Independence, Interchangeability, Martingales*, Second Edition. Springer-Verlag, New York.

Clogg, C. C. (1992). The Impact of the Sociological Methodology on Statistical Methodology (with Discussion). *Statist. Sci.* **7**, 183-207.

Colonius, H. (1977). On Keats' Generalization on the Rasch Model. *Psychometrika* **42**, 443-445.

Conway, J. B. (1985). *A Course in Functional Analysis*. Springer-Verlag, New York.

Cox, D. R. (1990). Role of Models in Statistical Analysis. *Statist. Sci.* **5**, 169-174.

Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies. Models, analysis and interpretation*. Chapman and Hall, London.

Dawid, A. P. (1979). Some Misleading Arguments Involving Conditional Independence. *J. R. Statist. Soc., Series B* **41**, 249-252.

Dellacherie, C. and Meyer, P. -A. (1975). *Probabilités et potentiel*. Hermann, Paris.

Dellacherie, C. and Meyer, P. -A. (1980). *Probabilités et potentiel. Théorie des martingales*. Hermann, Paris.

Diaconis, P. W., Eaton, M. L. and Lauritzen, S. L. (1992). Finite de Finetti Theorems in Linear Models and Multivariate Analysis. *Scand. J. Statist.* **19**, 289-315.

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

Fang, K. -T., Kotz, S. and Ng, K. -W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

Fischer, G. H. (1995). Some Neglected Problems in IRT. *Psychometrika* **60**, 459-487.

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. (1992). Bayesian Analysis of Mixtures: Some Results on Exact Estimability and Identification. In *Bayesian Statistics 4* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), Oxford University Press, Oxford, pp. 127-145.

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

Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statist. Probab. Lett.* **23**, 165-170.

Haberman, S. J. (1977). Maximum Likelihood Estimates in Exponential Response Models. *Ann. Statist.* **5**, 815-841.

Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory. Principles and Applications*. Kluwer and Nijhoff, Boston.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, Newbury Park.

Hewitt, E. and Savage, L. J. (1955). Symmetric Measures on Cartesian Product. *Trans. Amer. Math. Soc.* **80**, 470-501.

Holland, P. W. (1990). On the Sampling Theory Foundations of Item Response Theory Models. *Psychometrika* **55**, 557-601.

Holland, P. W. and Rosenbaum, P. R. (1980). Conditional Association and Unidimensionality in Monotone Latent Variable Models. *Ann. Statist.* **14**, 1523-1543.

Irtel, H. (1995). An Extension of the Concept of Specific Objectivity. *Psychometrika* **60**, 115-118.

Jansen, M. G. H. (1994). Parameters of the Latent Distribution in Rasch's Poisson Counts Model. In *Contribution to Mathematical Psychology, Psychometrics, and Methodology* (eds. G. H. Fischer and D. Laming), Springer-Verlag, New York, pp. 319-326.

Jansen, M. G. H. and van Duijn, A. J. (1992). Extensions of Rasch's Multiplicative Poisson Model. *Psychometrika* **57**, 405-414.

Junker, B. W. (1993). Conditional Association, Essential Independence and Montone Unidimensional Item Response Models. *Ann. Statist.* **21**, 1359-1378.

Junker, B. W. and Ellis, J. L. (1997). A Characterization of Monotone Unidimensional Latent Variable Models. *Ann. Statist.* **25**, 1327-1343.

Kadane, J. B. (1974). The Role of Identification in Bayesian Theory. In *Studies in Bayesian Econometrics and Statistics* (eds. S. E. Fienberg and A. Zellner), North-Holland, Amsterdam, pp. 175-191.

Landers, D. (1972). Sufficient and Minimal Sufficient σ-Fields. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **23**, 197-207.

Landers, D. (1974). Minimal Sufficient Statistics for Families of Product Measures. *Ann. Statist.* **2**, 1335-1339.

Landers, D. and Rogge, L. (1976). A Note on Completeness. *Scand. J. Statist.* **3**, 139.

Lauritzen, S. L. (1984). Extreme Point Models in Statistics (with Discussion). *Scand. J. Statist.* **11**, 65-91.

Lazarsfeld, P. F. (1966). Latent Structure Analysis and Test Theory. In *Readings of Mathematical Social Science* (eds. P. F. Lazarsfeld and N. W. Henry), Science Research Associates, Chicago, pp. 78-88.

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

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

Lindsay, B., Clifford, C. C. and Greco, J. (1991). Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis. *J. Amer. Statist. Assoc.* **86**, 96-107.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theory of Mental Test Scores.* Addison-Wesley, Massachusetts.

Martin-Löf, P. (1974). Repetitive Structure. In *Proceedings of Conference of Foundational Questions in Statistical Inference* (eds. O. Barndorff-Nielsen, P. Blaesild and G. Schou), Aarhus, pp. 271-294.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second Edition. Chapman and Hall, London.

Mouchart, M. (1976). A Note on Bayes Theorem. *Statistica* **36**, 349-357.

Mouchart, M. and San Martín, E. (1999). Identification Problems in Mixture Models: Some General Results. Discussion Paper 9909, Institut de statistique, Université catholique de Louvain; *submitted.*

Neyman, J. and Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**, 1-32.

Novick, M. R. (1979). Comments on "Conditional Independence in Statistical Theory". *J. R. Statist. Soc., Series B*, **41**, 27.

Pfanzagl, J. (1994). On Item Parameter Estimation in Certain Latent Trait Models. In *Contribution to Mathematical Psychology, Psychometrics, and Methodology* (eds. G. H. Fischer and D. Laming), Springer-Verlag, New York, pp. 249-263.

Picci, G. (1977). Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem. *SIAM J. Appl. Math.* **33**, 383-398.

Rao, M. M. (1971). Projective Limits of Probability Spaces. *J. Multivariate Anal.* **1**, 28-57.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Test.* Danmarks Paedogogishe Institut, Copenhagen.

Rasch, G. (1966a). An Item Analysis which takes Individual Differences into Account. *Brit. Jour. Math. Stat. Psych.* **19**, 49-57.

Rasch, G. (1966b). An Individualistic Approach to Item Analysis. In *Readings of Mathematical Social Science* (eds. P. F. Lazarsfeld and N. W. Henry), Science Research Associates, Chiago, pp. 89-108.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica* **39**, 577-591.

Sansone, G. (1959). *Orthogonal Functions.* Interscience Publishers, New York.

Sobel, M. E. (1997). Measurement, Causation and Local Independence in Latent Variable Models. In *Latent Variable Modeling and Applications to Causality* (ed. M. Berkane), Springer, New York, pp. 11-28.

Teicher, H. (1961). Identifiability of Mixtures. *Ann. Math. Statist.* **32**, 244-248.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scand. J. Statist.* **9**, 23-30.

Wald, A. (1948). Estimation of a Parameter when the Number of Unknown Parameters Increases Indefinitely with the Number of Observations. *Ann. Math. Statist.* **19**, 220-227.

Wermuth, N. and Lauritzen, S. L. (1990). On Substantive Research Hypotheses, Conditional Independence Graphs and Graphical Chain Models. *J. R. Statist. Soc., Series B* **52**, 1-50.

Institut de statistique, Université catholique de Louvain, 20 Voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium.